Sentiment Classification of Food Influencer Content Reviews using Support Vector Machine Model through CRISP-DM Framework

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
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id
Submitted: 24/02/2024; Accepted: 31/03/2024; Published: 31/03/2024

Abstract—The research problem revolves around the challenges in effectively marketing culinary tourism aligned with tourist preferences in Indonesia, necessitating a substantial exploration of consumer sentiments related to culinary diversity through the lens of food influencer content. Food influencers are crucial in stimulating tourists' interest in gastronomy through culinary tourism in Indonesia. This research reveals challenges in culinary tourism marketing aligned with tourist preferences, necessitating substantial exploration of consumer sentiments related to culinary diversity through food influencer content. The sentiment classification method employed is the Cross-Industry Standard Process for Data Mining (CRISP-DM) using the Support Vector Machine (SVM) algorithm and the SMOTE operator. The data source is derived from a video with the ID PMIIIYa_bzV8, containing 114,422 comments. This study collects and processes 30,000 comments, resulting in 9,323 data points. The findings highlight the vital performance metrics of SVM models, both with and without SMOTE, showcasing high accuracy, precision, recall, and F-measure values. Specifically, SVM without SMOTE achieves 95.28% accuracy, while SVM with SMOTE achieves 98.67%. Despite some limitations in discerning positive and negative sentiments, indicated by moderate Area Under the Curve (AUC) values (0.608 to 0.658), the overall efficacy of SVM in sentiment analysis for food influencer content is apparent. Drawing from a dataset of 30,000 comments, these insights contribute to advancing sentiment analysis methodologies and offer practical implications for understanding consumer perceptions and behaviors in digital media and influencer marketing. Additionally, the prominence of frequent words such as “bang” (1322), “nonton” (1064), “makan” (921), “yang” (801), “puasa” (711), “tahun” (484), “ngiler” (448), “lagi” (384), “tanboy” (311), and “enak” (315), as extracted from RapidMiner analysis, underscores the significance of language patterns in the realm of food influencer content.

Keywords: Food; Influencer; Sentiment; Classification; SVM

1. INTRODUCTION

The influential status of food influencers has dramatically influenced the prominence of the culinary and gastronomic tourism [1]. These individuals significantly impact the expansion of culinary tourism as influential opinion leaders in the digital era [2]. The wide distribution of this material on various social media platforms promotes the interchange of varied culinary encounters, cultivating a heightened interest in and appreciation for regional and international gastronomy [3]. Their comprehensive compilation of materials enables readers to understand culinary traditions by employing visually engaging graphics and conducting thorough gastronomic investigations [4]. Food influencers have undoubtedly positioned themselves as intercultural bridge builders, connecting modern culinary trends with traditional culinary methods [5]. Nonetheless, many question whether their influence is superficial [6]. Their substantial impact on the culinary dialogue highlights the importance of food influencers in promoting the development of the culinary tourism industry and facilitating a more profound affinity between people and the vast culinary traditions of the world [7].

To underscore the possibilities of culinary and gastronomic tourism, food influencers must produce engaging content that aligns with the preferences of the broad spectrum of viewers [8]. The primary objective is to provide content that showcases the variety of gastronomic encounters and caters to the broad spectrum of preferences of the intended audience [9]. Food influencers can arouse curiosity and capture the public’s interest in diverse culinary choices by strategically adapting their content to align with well-liked cuisine and cultural inclinations [10]. Enhancing the accessibility and appeal of the material by incorporating visually appealing and relatable elements promotes a more extensive audience engagement [11]. Culinary tourism and exploring gastronomic delights are finally propelled forward by deliberate adaptations to audience preferences, notwithstanding certain apprehensions that such customization would compromise authenticity [12]. In conclusion, food influencers attuned to consumer preferences play a pivotal role in developing the potential of culinary and gastronomic tourism through their meticulous content creation [13].

The public’s perception of content attributes, particularly concerning the food and beverage business, presents a substantial opportunity to stimulate interest in food and drink tourism [14]. This potential can be primarily exploited by systematically analyzing public sentiment using the Cross-Industry Standard Process for Data Mining (CRISP-DM), which integrates Support Vector Machines (SVM) for advanced sentiment categorization [15]. Utilizing sophisticated data mining techniques, it is possible to discern and interpret the embedded sentiments in public discourse concerning culinary content [16]. This approach assists entities in the tourism industry in anticipating and addressing any gaps in culinary offerings and identifying prevailing preferences and expectations that influence public tastes to facilitate strategic interventions [17]. By integrating a CRISP-DM framework alongside SVM, sentiment analysis attains a higher methodological rigor, guaranteeing a...
more comprehensive comprehension of public sentiments [18]. However, there is a degree of skepticism surrounding the application of machine learning algorithms and sentiment analysis tools in this domain due to the potential oversimplification of human emotions [19]. Lastly, employing sophisticated analytical tools to match products with the evolving preferences of the target market to gauge public sentiment presents a strategic approach to enhancing the culinary tourism [20].

This study aims to clarify tourists' perspectives by examining public sentiment analysis about material generated by food influencers. The primary emphasis is deciphering the intricate strata of visitor viewpoints, as manifested through sentiments recognizable in the dialogue about material generated by food influencers [21]. Through a comprehensive analysis of public attitudes expressed in social media interactions, reviews, and online discussions of culinary experiences, this research reveals significant revelations regarding the determinants that impact the perceptions and choices of tourists [22]. It is crucial to investigate these attitudes to illuminate the intricate relationship between the material produced by food influencers and how it shapes the choices, expectations, and overall pleasure of tourists [23]. With the understanding that sentiment analysis may contain some degree of subjectivity, the purpose of this study is to fully understand how the public perceives material created by food influencers impacts the culinary tourism experience of tourists [24]. In summary, the examination of public sentiment provides a valuable framework for understanding and analyzing the complex interplay between tourists and the information shared by food influencers for a culinary adventure.

The pressing nature of this investigation arises from the essential requirement to fully grasp culinary tourism's dynamic terrain in the digitalization era. The primary rationale for this is the increasing impact that food influencers have on the preferences and actions of tourists, in conjunction with the growing importance of public opinion in determining gastronomic encounters [25]. Given the dynamic nature of the tourism sector, which is significantly impacted by technological progress and shifting consumer preferences, it is critical to comprehend the complex relationship between culinary influencers' content and tourists' perceptions [26]. By comprehensively understanding the fundamental mechanisms that control this correlation, individuals and organizations interested in the tourist and hospitality industries can devise well-informed approaches to exploit nascent prospects and mitigate potential obstacles [27]. Neglecting this knowledge deficit could lead to foregone prospects for destination promotion, advancement of products, and general improvement of the visitor experience [28]. In summary, this research is critical due to its potential to yield practical insights that can guide strategic interventions designed to maximize the collaboration between culinary tourism and the content produced by influencers in the food industry.

This study's theoretical and practical ramifications could enhance scholarly dialogue and culinary tourism industry methodologies. This study makes a theoretical contribution by elucidating the intricate relationship between culinary influencer content and visitor perceptions. As a result, the digital marketing, tourism, and hospitality fields gain further knowledge. Incorporating sophisticated approaches such as machine learning and sentiment analysis into examining public sentiment enhances the study's methodological rigor [29]. It may facilitate future developments in research methodology about consumer behavior analysis [30].

The practical implications of the research findings are that they can provide destination marketers, tourist stakeholders, and enterprises engaged in culinary offerings with valuable information for making strategic decisions [31]. By analyzing how tourists interact with and react to content generated by food influencers, professionals can customize marketing approaches, create culinary products specifically designed to appeal to the target audience, and ultimately improve the allure of a destination [32]. Hence, the knowledge obtained from this study has the capacity to establish a connection between theoretical concepts and real-world implementations, thereby making a significant contribution to the advancement of industry standards and academic comprehension in the dynamic realm of culinary tourism [33]. As stated in conclusion, the research's practical applicability and theoretical enrichment highlight its dual significance in furthering knowledge and promoting constructive advancements [34].

One potential constraint of this research is its dependence on digital data sources and sentiment analysis methodologies, which might not provide an exhaustive representation of the entirety of tourist feelings or preferences. Although sentiment analysis provides valuable insights into online discourse, it might fail to consider the influence of offline encounters and subtle emotions that shape travelers' views of culinary experiences. In addition, the study's emphasis on food influencer material may fail to consider cultural, social, and environmental aspects that can significantly impact tourist behavior in culinary tourism.

Notwithstanding these constraints, this study provides opportunities for more investigation to thoroughly examine the convergence of digital media, tourism, and cuisine. Further research may include mixed-method approaches, such as observational studies and qualitative interviews, to provide a more comprehensive understanding of the motives, decision-making processes, and experiences of tourists engaged in culinary tourism. Furthermore, researching the efficacy of various forms of influencer-generated content, the significance of genuineness in gastronomic encounters, and the consequences of emerging technologies on the conduct of tourists may provide scholars and professionals in the discipline with invaluable knowledge. In summary, this study enhances our comprehension of the correlation between the content produced by food influencers and the perceptions of tourists. However, there are still significant prospects for additional research to investigate and rectify the constraints identified in this study.
2. RESEARCH METHODOLOGY

2.1 Research Gap Analysis using Vosviewer for Food Influencer Topics

The research gap analysis utilizing Vosviewer for food influencer topics reveals a notable absence of comprehensive studies examining the nuanced dynamics and impact of food influencers in culinary tourism. While existing research offers valuable insights into various aspects of influencer marketing and culinary tourism separately, there remains a significant gap in integrating these domains to elucidate food influencers' specific role and influence on tourist perceptions, behaviors, and destination choices [35], [36], [36]–[41]. Despite the proliferation of food influencer content on digital platforms, scholarly inquiry into its implications for culinary tourism remains relatively limited, presenting an opportunity for further investigation and knowledge advancement [42]. By leveraging Vosviewer analysis, this research identifies thematic clusters, key research topics, and potential areas for exploration within the intersection of food influencers and culinary tourism, thereby contributing to a more comprehensive understanding of this evolving phenomenon. In conclusion, utilizing Vosviewer for research gap analysis underscores the need for concerted efforts to bridge the existing gaps in scholarship and propel the discourse on food influencer topics in the context of culinary tourism.

Figure 1. Gap Analysis of Food Influencer Topics

Based on the research gap identification results concerning food influencers, it is evident that popular topics related to food encompass food tourism, food security, food waste, food safety, influencers, and others. This identification highlights the multifaceted nature of food-related discourse and the diverse range of issues that intersect with the phenomenon of food influencers [42], [42]–[46], [46], [47]. The prominence of these topics underscores the interconnectedness between culinary experiences, societal concerns, and digital media influences, signaling the need for comprehensive research efforts to elucidate the complex dynamics at play [48], [48], [49]. By acknowledging these prevalent themes, researchers can orient their inquiries to address pressing challenges and capitalize on emerging opportunities within the dynamic landscape of food influencer discourse [50], [51]. Thus, identifying these popular topics is a valuable foundation for directing future research endeavors toward a more holistic understanding of the intersection between food influencers and broader societal issues.

This underscores the importance of research focusing on food influencers, particularly the Sentiment Classification of Food Influencer Content Reviews Using a Support Vector Machine. As evidenced by the identified popular topics related to food, a pressing need exists to delve deeper into understanding the sentiments expressed in food influencer content. Utilizing advanced methodologies like Support Vector Machine for sentiment classification can provide valuable insights into audiences’ nuanced perceptions and reactions toward culinary experiences showcased by food influencers. Researchers can unravel underlying patterns, preferences, and trends by systematically analyzing and categorizing sentiments embedded in reviews and discussions surrounding food influencer content, informing strategic interventions and decision-making processes in culinary tourism and digital marketing. Thus, the pursuit of research in this area holds significant promise in advancing knowledge and addressing the complexities inherent in the influence of food influencers on consumer behavior and perceptions.

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

This research employs the methodology of the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM provides a systematic and structured approach to data mining projects, encompassing distinct phases such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. By adhering to the CRISP-DM framework, this study ensures methodological rigor and comprehensiveness in analyzing data about food influencer content and its impact on culinary tourism. Using CRISP-DM facilitates a
disciplined approach to navigating the complexities inherent in data-driven research, thereby enhancing the reliability and validity of findings. Adopting CRISP-DM as the methodological framework underscores the commitment to conducting a rigorous and systematic inquiry into the phenomenon under investigation.

Figure 2. Implementation of CRISP-DM in Sentiment Classification

CRISP-DM is a relevant approach to sentiment classification studies based on machine learning. The comprehensive and structured framework provided by CRISP-DM aligns well with the iterative nature of sentiment classification tasks, which involve multiple stages of data preprocessing, feature engineering, model selection, and evaluation. By adhering to the CRISP-DM methodology, researchers can systematically navigate through the various phases of sentiment classification projects, ensuring methodological rigor and facilitating the effective management of complexities in analyzing textual data. Additionally, CRISP-DM encourages collaboration between domain experts, data scientists, and other stakeholders, fostering interdisciplinary insights and enhancing the overall quality of sentiment classification research. Thus, adopting CRISP-DM in sentiment classification studies represents a reasonable approach to achieving robust and reliable outcomes in machine learning-based sentiment analysis.

2.2.1 Business Understanding

In the business understanding phase, the data context pertains to the influence of food influencers on consumer behavior. This study collected 114,427 review data from video content (PMhLY buoy &t=1s) and processed 30,000 review data to analyze toxicity, sentiment, and social network dynamics. By focusing on these aspects, the research aims to understand the impact of food influencer content on consumer perceptions and interactions. The meticulous data collection and analysis underscore the commitment to rigorous inquiry and contribute to advancing knowledge in culinary tourism and digital marketing.

Figure 3. Post Per Day Statistic (Communalytic)

Based on the frequency of daily posts, it can be inferred that the public's intention and attention towards food influencers are notably high. The substantial volume of posts produced by food influencers reflects a robust level of engagement and interest from their audience. This heightened level of interaction signifies a strong affinity for the content shared by food influencers, suggesting that they wield considerable influence over consumer preferences and behaviors in gastronomy. Consequently, the frequency of posts serves as a tangible indicator of the significant impact that food influencers exert on shaping trends, fostering culinary exploration, and driving engagement within the digital landscape. Thus, the frequency of posts per day not only underscores the pervasive influence of food influencers but also highlights the profound connection between content creators and their audience, contributing to the broader discourse on the role of digital influencers in contemporary consumer culture.
2.2.2 Data Understanding

During the data understanding phase, data cleaning processes were conducted using the RapidMiner application, which included tokenization, case transformation, stopwords removal, and duplicate removal. Subsequently, the cleaned review data underwent extraction using the Extract Sentiment operator to obtain sentiment scores. These preprocessing steps are crucial for ensuring the quality and reliability of the data analysis process, as they help to eliminate noise and inconsistencies that could distort the results. By employing systematic data-cleaning techniques, this research maintains methodological rigor. It enhances the validity of the findings, thus contributing to a more robust understanding of the influence of food influencers on consumer behavior in culinary tourism.

Figure 4. Frequently Used Words in 30,000 Reviews Data (Communalytic)

Based on an analysis of 30,000 comments, it is observed that specific keywords exhibit varying frequencies, providing insights into the prevalent themes within the dataset. Notably, the term “nonton” occurs 136 times, suggesting a substantial engagement with content related to viewing or watching. Similarly, occurrences of “bang” (91 times) and ”Bang” (70 times) may reflect diverse contexts, including informal addresses or cultural expressions within the discourse. Furthermore, the recurrence of the term “makan” (82 times) implies a recurring theme related to food or dining experiences, indicative of the dataset's focus on culinary content. The prominence of “tahun” (60 times) suggests discussions about temporal aspects potentially associated with reflections on past experiences or the passage of time. This keyword frequency analysis provides a preliminary understanding of the predominant topics discussed in the dataset, laying the groundwork for further in-depth exploration and contextual interpretation of the comments.

2.2.3 Modeling

In the modeling phase, the Support Vector Machine (SVM) algorithm was employed and evaluated based on accuracy, precision, recall, F-measure, and area under the curve (AUC) values. SVM is chosen for its ability to handle complex datasets and its effectiveness in classifying sentiment in text data. By utilizing these evaluation metrics, the performance of the SVM algorithm in sentiment classification can be comprehensively assessed, providing insights into its efficacy in capturing the nuances of sentiment expressed in food influencer content reviews. This rigorous evaluation process ensures the reliability and robustness of the sentiment analysis results, thereby enhancing the credibility of the findings and contributing to a deeper understanding of the influence of food influencers on consumer behavior in culinary tourism.

The Support Vector Machine (SVM) is an algorithm for sentiment categorization that utilizes supervised learning. Support Vector Machines (SVM) execute classification decisions by creating a hyperplane in a high-dimensional space that optimally segregates data points into distinct classes. To classify sentiment, the SVM algorithm attempts to identify the best hyperplane that divides positive and negative sentiment occurrences in the feature space. The mathematical representation of the decision function of SVM is as follows:

\[ f(x) = \text{sign}(\sum_{i=1}^{N} a_i y_i K(x,x_i) + b) \]  

Where:
- \( f(x) \) represents the decision function that predicts the class label of the input instance \( x \).
- \( a_i \) is the Lagrange multiplier associated with the \( i \)th training instance.
- \( y_i \) is the class label of the \( i \)th training instance (either +1 for positive sentiment or -1 for negative sentiment).
- \( K(x,x_i) \) is the kernel function that computes the similarity between the input instance \( x \) and the support vectors \( x_i \).
- \( b \) is the bias term.

The choice of kernel function, such as linear, polynomial, or radial basis function (RBF) kernel, determines the mapping of the input data into a higher-dimensional space where the data points can be effectively separated.
The SVM algorithm aims to optimize the values of \( (a_i) \) and \((b)\) to maximize the margin between the support vectors of different classes while minimizing classification errors. Once the optimal hyperplane is determined, the sign of the decision function \( f(x) \) determines the predicted class label of the input instance \((x)\). In sentiment classification, SVM learns to classify text data by representing documents as feature vectors (e.g., using bag-of-words or TF-IDF representations) and then applying the SVM algorithm to learn the decision boundary between positive and negative sentiment instances.

### 2.2.4 Evaluation

In the evaluation phase, the algorithm's performance is assessed by comparing the results obtained with and without the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is employed to address the issue of class imbalance in the dataset, particularly in scenarios where one class significantly outweighs the other. By generating synthetic samples for the minority class, SMOTE aims to balance the distribution of classes and improve the algorithm's ability to classify sentiments accurately. Through this comparative analysis, the impact of SMOTE on the algorithm's performance can be gauged, providing valuable insights into its efficacy in handling imbalanced data in sentiment classification tasks. This evaluation process is essential for ensuring the reliability and effectiveness of the sentiment analysis model, thereby enhancing the credibility of the research findings and contributing to advancements in computational linguistics and sentiment analysis. Particularly for machine learning tasks, SMOTE (Synthetic Minority Over-sampling Technique) is a technique used to address class imbalance in datasets. It produces synthetic samples of the minority class to achieve class distribution parity. To generate synthetic samples, minority class instances and their nearest neighbors are identified, and new instances are generated along the line segments linking these neighbors. The following expression represents the formula utilized in SMOTE to generate synthetic samples:

Let \((x_{\text{minority}})\) be an instance from the minority class and \((x_{nn})\) be one of its \((k)\) nearest neighbors. A synthetic instance \((x_{\text{new}})\) is generated as a linear combination of \((x_{\text{minority}})\) and \((x_{nn})\) as follows:

\[
x_{\text{new}} = x_{\text{minority}} + \lambda \times (x_{nn} - x_{\text{minority}})
\]

Where:

- \(x_{\text{new}}\) is the synthetic instance generated.
- \(\lambda\) is a random value in the range \([0, 1]\) that determines the position of the new instance along the line segment between \(x_{\text{minority}}\) and \(x_{nn}\). It controls the amount of minority class oversampling.
- \(x_{\text{minority}}\) and \(x_{nn}\) are feature vectors representing instances from the minority class and its nearest neighbor, respectively.

By iterating this procedure for every occurrence in the minority class, a balanced dataset is produced in which every class has an equivalent number of cases, by producing synthetic samples that closely mimic instances of the minority class, SMOTE efficiently addresses the issue of class imbalance and enhances the performance of machine learning models trained on unbalanced datasets.

### 2.2.5 Deployment

In the deployment phase, valuable insights can be gleaned regarding the distribution of positive and negative sentiment classes within the review data of food influencer content. This information is essential for informing decision-making processes and strategizing interventions to enhance the effectiveness of culinary tourism marketing campaigns and digital content creation strategies. By understanding the proportion of positive and negative sentiment classes, stakeholders in the tourism and hospitality sectors can tailor their approaches to align with the prevailing sentiments of the audience, thereby maximizing engagement and optimizing the impact of their initiatives. Consequently, the deployment phase is pivotal in translating research findings into actionable strategies that resonate with consumers' sentiments and preferences in culinary tourism.

---

**Figure 5.** Toxicity Analysis (Communalytic)
Based on the results of toxicity analysis, it is discerned that the dataset exhibits varying degrees of toxicity across different categories. The toxicity scores indicate the likelihood of toxic behavior or language in the comments, with values ranging from 0.01198 to 0.08552 across different toxicity categories. Particularly noteworthy is the relatively high probability of toxicity in categories such as Profanity (0.07372) and Identity Attack (0.01517), suggesting the prevalence of offensive language and attacks on personal identity within the dataset. However, it is essential to note that the probabilities of toxicity across all categories are below 1, indicating that while toxic behavior may be present, it is not pervasive throughout the dataset. This toxicity analysis provides valuable insights into the nature of interactions within the comments, informing subsequent steps in data preprocessing and sentiment analysis to ensure the reliability and validity of the findings. Based on the results of sentiment analysis and toxicity analysis, the recommended deployment stages include implementing a sentiment classification model for real-time comment analysis, integrating a toxicity filter to screen out potentially harmful content, establishing a user feedback mechanism for continuous improvement, and implementing monitoring and evaluation processes to assess system performance. By integrating these stages into the deployment strategy, online platforms can create a safer and more positive user experience while gaining valuable insights into user sentiment and behavior.

3. RESULT AND DISCUSSION

Food influencers are of paramount importance in increasing enthusiasm for culinary tourism. Food influencers can engross audiences and motivate them to delve into various gastronomic encounters by utilizing meticulously produced content and compelling storytelling. By highlighting distinctive cuisines, regional specialties, and obscure establishments, food influencers aid in conserving and commemorating culinary heritage [52]. Furthermore, their endorsements and evaluations are reliable knowledge repositories for tourists searching for genuine and indelible gastronomic encounters [53]. Food influencers greatly enhance culinary tourism by stimulating curiosity, facilitating cultural interchange, and admiration for various gastronomic customs.

Prior to performing sentiment analysis, it is critical to ascertain the social network affiliations of individuals who offer feedback on content associated with culinary influencers. This preliminary stage is a fundamental component of the analysis procedure, enabling researchers to situate the remarks inside social media platforms or virtual communities [54]. Through the analysis of users’ social networks, researchers acquire valuable knowledge regarding the intricacies of user interactions, the demographics of the audience, and the dominant norms and behaviors in the digital [55]. Furthermore, by gaining knowledge of the social network, researchers are able to customize the sentiment analysis methodology to consider platform-specific subtleties and user attributes [56]. This improves the precision and pertinence of the study results. Hence, identifying the social network is essential to performing a thorough sentiment analysis on remarks about culinary influencers’ content.

Figure 6. Social Network and Topic Analysis (Communalytic)

Based on the results of identifying social network patterns using the reply-to method, it is observed that there are a total of 25,896 actor nodes and 395 edges based on the indegree centrality metric. This indicates a substantial level of engagement and connectivity within the social network, reflecting the active participation of users in discussions related to food influencer content. Furthermore, the subsequent implementation of topic analysis serves as a valuable methodological approach to uncovering latent themes and patterns within the dataset. By identifying and grouping semantically similar social media posts, topic analysis enables researchers to discern prevalent topics and trends within the discourse surrounding food influencer content. Consequently, integrating
both social network analysis and topic analysis offers a comprehensive understanding of the dynamics and content of discussions within the online community, thereby enriching scholarly inquiry in digital media and influencer marketing.

This research analyzes a dataset derived from the video content of Food Influencer Tanboy Kun (PMhLY_buV8), obtained from the YouTube platform, comprising 30,000 data points. Subsequently, an evaluation of 9,323 data points was conducted using the RapidMiner application. The utilization of Tanboy Kun's video content as the dataset source offers a rich and diverse selection of material for analysis, given the influencer's popularity and extensive reach within the food-centric digital media landscape. Additionally, employing RapidMiner for evaluation ensures a systematic and efficient assessment of the dataset's characteristics and suitability for further analysis. Overall, this approach facilitates rigorous and methodologically sound exploration of the sentiments and trends prevalent within food influencer content, thereby contributing valuable insights to digital media research.

**Table 1. Comparative Analysis of SVM using SMOTE and without SMOTE**

<table>
<thead>
<tr>
<th>SVM using SMOTE</th>
<th>SVM without SMOTE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PerformanceVector:</strong></td>
<td><strong>PerformanceVector:</strong></td>
</tr>
<tr>
<td>accuracy: 95.28% +/- 0.73% (micro average: 95.28%)</td>
<td>accuracy: 98.67% +/- 0.18% (micro average: 98.67%)</td>
</tr>
<tr>
<td><strong>ConfusionMatrix:</strong></td>
<td><strong>ConfusionMatrix:</strong></td>
</tr>
<tr>
<td>True: Negative</td>
<td>True: Negative</td>
</tr>
<tr>
<td>Negative: 3337</td>
<td>Negative: 10</td>
</tr>
<tr>
<td>Positive: 227</td>
<td>Positive: 80</td>
</tr>
<tr>
<td>6191</td>
<td>6429</td>
</tr>
<tr>
<td>AUC (optimistic): 0.977 +/- 0.006 (micro average: 0.977) (positive class: Positive)</td>
<td>AUC (optimistic): 0.658 +/- 0.083 (micro average: 0.658) (positive class: Positive)</td>
</tr>
<tr>
<td>AUC: 0.977 +/- 0.006 (micro average: 0.977) (positive class: Positive)</td>
<td>AUC: 0.633 +/- 0.079 (micro average: 0.633) (positive class: Positive)</td>
</tr>
<tr>
<td>AUC (pessimistic): 0.977 +/- 0.006 (micro average: 0.977) (positive class: Positive)</td>
<td>AUC (pessimistic): 0.608 +/- 0.081 (micro average: 0.608) (positive class: Positive)</td>
</tr>
<tr>
<td>precision: 96.47% +/- 0.72% (micro average: 96.46%) (positive class: Positive)</td>
<td>precision: 98.77% +/- 0.12% (micro average: 98.77%) (positive class: Positive)</td>
</tr>
<tr>
<td><strong>ConfusionMatrix:</strong></td>
<td><strong>ConfusionMatrix:</strong></td>
</tr>
<tr>
<td>True: Negative</td>
<td>True: Negative</td>
</tr>
<tr>
<td>Negative: 3337</td>
<td>Negative: 10</td>
</tr>
<tr>
<td>Positive: 227</td>
<td>Positive: 80</td>
</tr>
<tr>
<td>6191</td>
<td>6429</td>
</tr>
<tr>
<td>recall: 96.19% +/- 0.68% (micro average: 96.19%) (positive class: Positive)</td>
<td>recall: 99.89% +/- 0.15% (micro average: 99.89%) (positive class: Positive)</td>
</tr>
<tr>
<td><strong>ConfusionMatrix:</strong></td>
<td><strong>ConfusionMatrix:</strong></td>
</tr>
<tr>
<td>True: Negative</td>
<td>True: Negative</td>
</tr>
<tr>
<td>Positive: 227</td>
<td>Positive: 80</td>
</tr>
<tr>
<td>6191</td>
<td>6429</td>
</tr>
</tbody>
</table>
In evaluating the SVM without the SMOTE, the PerformanceVector demonstrates notably high accuracy, precision, recall, and F-measure values. Specifically, the SVM model achieves an accuracy of 98.67%, precision of 98.77%, recall of 99.89%, and F-measure of 99.33%. Moreover, analysis of the confusion matrix reveals minimal misclassification occurrences, with only ten instances of the negative class misclassified as positive and seven instances of the positive class misclassified as negative. However, the Area Under the Curve (AUC) values suggest moderate discrimination performance in distinguishing between positive and negative classes, with AUC values ranging from 0.608 to 0.658. Conversely, the evaluation of SVM without SMOTE yields promising performance metrics, including an accuracy of 95.28%, precision of 96.47%, recall of 96.19%, and F-measure of 96.33%.

Furthermore, the AUC values indicate strong discrimination capability between positive and negative classes, which stand at 0.977. Notably, minimal misclassification is evident in the confusion matrix, with a limited number of instances incorrectly classified as negative (245) and positive (227) classes. Overall, these findings emphasize the efficacy of SVM without SMOTE in accurately classifying sentiments within the dataset, underscoring its potential for sentiment analysis tasks in food influencer content.

In examining the impact of sentiment classification, toxicity analysis, and social network patterns on the future of food influencers and culinary tourism, this study provides comprehensive insights into the evolving landscape of digital gastronomy. The primary focus lies in evaluating the effectiveness of sentiment classification models, toxicity assessments, and social network dynamics in shaping consumer perceptions within the burgeoning realm of food influencers. The analysis of sentiments aids in understanding the emotional tone of audience interactions, while toxicity assessments shed light on the potential challenges and risks associated with influencer content. Simultaneously, exploring social network patterns reveals the interconnectedness of users and influencers, offering valuable information for strategizing marketing efforts. The integration of these facets contributes to a nuanced understanding of the potential implications for the future trajectory of food influencers and their impact on the flourishing domain of culinary tourism.

This research makes a theoretical contribution by deepening our comprehension of the complex interplay among sentiment classification, toxicity analysis, and social network patterns of culinary tourism and food influencers. This research establishes a conceptual framework that facilitates the understanding of how audience perceptions and behaviors are influenced by the sentiments conveyed in influencer material. Through an in-depth examination of toxicity analysis, the study detects possible obstacles and hazards linked to the content, providing valuable perspectives on the ever-changing realm of digital interactions. Moreover, by establishing a theoretical framework for comprehending the interdependent interactions between users and influencers, social network pattern analysis provides a foundation for future research on the network effects in digital gastronomy. Collectively, these theoretical advancements establish a framework for a more comprehensive comprehension of how food influencers impact consumer attitudes and actions. This lays the groundwork for additional academic investigations into the intersections of digital media and gastronomic tourism.

Concerning its practical implications, this study offers relevant insights to stakeholders in culinary tourism, digital marketing, and influencer management. The results of sentiment classification provide content creators and influencers with actionable advice on how to effectively customize their messages to correspond with the intended sentiments of their audience. The toxicity analysis provides practical utility by facilitating identifying and mitigating potential problems linked to offensive or harmful content by influencers and platforms. Furthermore, examining social network patterns has practical applications for marketing strategies, aiding stakeholders in comprehending how network architectures impact content distribution. The research findings offer practical suggestions about culinary tourism for destination marketers and tourism boards. These suggestions involve establishing strategic collaborations with influencers and utilizing their social media platforms to promote gourmet experiences. The knowledge acquired from this research might provide valuable input for formulating protocols for content generation and influencer partnerships, hence promoting ethical and constructive interactions in digital gastronomy. In conclusion, the practical implications of this study provide industry professionals with the ability to effectively utilize digital platforms to promote culinary tourism and traverse the ever-changing dynamics of influencer marketing.

This research makes a substantial scholarly contribution to digital media studies, specifically culinary tourism and the power of food influencers. An examination of sentiment categorization, toxicity analysis, and...
social network patterns in this field contributes to the extant body of knowledge by enhancing our theoretical comprehension of the dynamic nature of digital gastronomy. Theoretical significance is derived from constructing a conceptual framework that amalgamates sentiment analysis, toxicity assessment, and social network analysis. This framework provides a nuanced viewpoint on the intricate interconnections among influencers, content, and audience perceptions. This approach serves as a cornerstone for further scholarly investigations, motivating researchers to delve into the intricate dynamics of emotional undertones, possible hazards, and network architectures across many categories of digital information. Furthermore, this research holds academic importance as it contributes to methodological developments, specifically in applying machine learning algorithms to the classification of sentiment and toxicity analysis. The research contributes to the methodological toolset accessible to scholars in the field by demonstrating the practicality of advanced computational approaches in analyzing digital media content by utilizing cutting-edge techniques like Support Vector Machines. This study contributes to the academic conversation surrounding digital media, influencer marketing, and tourism by presenting a thorough framework and novel approaches to comprehend the complex interconnections within the digital gastronomy domain.

4. CONCLUSION

In conclusion, this research has provided valuable insights into the effectiveness of sentiment analysis techniques, particularly in the context of food influencer content. The study has demonstrated the potential of SVM algorithms in accurately classifying sentiments within the dataset through rigorous data collection, preprocessing, modeling, and evaluation processes. The findings indicate that SVM models, both with and without SMOTE, exhibit vital performance metrics, including high accuracy, precision, recall, and F-measure values. Specifically, SVM without SMOTE achieves an accuracy of 95.28%, while SVM with SMOTE achieves 98.67%. Despite some limitations in distinguishing between positive and negative sentiments, as evidenced by moderate Area Under the Curve (AUC) values ranging from 0.608 to 0.658, the overall efficacy of SVM in sentiment analysis tasks for food influencer content is evident. Based on a dataset of 30,000 comments, these findings contribute to advancing sentiment analysis methodologies and provide practical implications for understanding consumer perceptions and behaviors in digital media and influencer marketing. Moreover, frequent words such as "bang" (1322), "nonton" (1064), "makan" (921), "puasa" (711), "tahuan" (484), "ngiler" (448), "lagi" (384), "tanboy" (311), and "enak" (315) extracted from RapidMiner analysis underscore the significance of language patterns in the realm of food influencer content.

ACKNOWLEDGMENT

Acknowledgment to the Tourism Department, Faculty of Business Administration and Communication, LPPM, Atma Jaya Catholic University of Indonesia, is extended with utmost gratitude for invaluable support and resources throughout this research endeavor.

REFERENCES


Copyright © 2024 Yerik Afrianto Singgalen, Page 526
This Journal is licensed under a Creative Commons Attribution 4.0 International License


