G-Tech : Jurnal Teknologi Terapan
Volume 7, No. 3, Juli 2023, hal. 967-976
E-ISSN: 2623-064x | P-ISSN: 2580-8737

Unmasking the Sentiments of Labuan Bajo: An Instagram-based Analysis for Tourism Insights through VADER Sentiment Analysis

Johan Setiawan[1]✉, Vegeterrikin Gousander [2], Iwan Prasetiawan[3]

[1, 2, 3] Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 23-05-2023
Direvisi : 30-05-2023
Diterima : 05-06-2023

ABSTRAK

Labuan Bajo merupakan tujuan utama wisata yang terkenal, menarik banyak wisatawan asing dan lokal yang dengan antusias membagikan pengalaman mereka melalui foto dan video di Instagram. Penelitian ini bertujuan untuk mengeksplorasi umpan balik dari wisatawan lokal dan asing yang telah mengunjungi Labuan Bajo, serta mengidentifikasi tujuan wisata paling populer di daerah tersebut. Data penelitian dikumpulkan dari Instagram dengan menggunakan tagar "labuanbajo". Metode analisis sentimen berbasis leksikon VADER digunakan untuk mengukur polaritas sentimen. Temuan eksperimental menunjukkan tingkat akurasi yang impresif, yaitu 72%, dengan metode sentimen leksikon VADER. Hasil penelitian mengungkapkan bahwa tujuan wisata Labuan Bajo cenderung membangkitkan sentimen positif, dengan 58,55% dari 3.351 data yang dikumpulkan dikategorikan sebagai positif, mencapai total 1.962 kali. Destinasi populer yang sering dikunjungi wisatawan antara lain Pulau Komodo, Pulau Padar, Pink Beach, Pulau Kelor, Pulau Rinca, Pulau Kanawa, dan Desa Waerebo.

Kata Kunci:

Instagram, Analisis Sentimen, Text Mining, VADER sentimen, Labuan Bajo

***Keywords* :**

Instagram, Sentiment Analysis, Text Mining, VADER Sentiment, Labuan Bajo

ABSTRACT

Labuan Bajo is a renowned premier tourist destination that attracts numerous foreign and local tourists who enthusiastically share their experiences through photos and videos on Instagram. This research aims to explore the feedback received from both local and foreign tourists who have visited Labuan Bajo, as well as identify the most popular tourist destinations in the region. Data for this study was collected from Instagram using the hashtag "labuanbajo". The VADER lexicon-based sentiment analysis method was employed to measure sentiment polarity. Experimental findings revealed an impressive accuracy rate of 72% using the VADER sentiment lexicon method. The research results indicate that Labuan Bajo's tourist destinations predominantly evoke positive sentiment, with 58.55% of the 3,351 collected data points being labeled as positive, totaling 1,962 instances. Popular destinations frequented by tourists include Komodo Island, Padar Island, Pink Beach, Kelor Island, Rinca Island, Kanawa Island, and Waerebo Village.

***Corresponding Author* :**
Johan Setiawan
Information Systems, Faculty of Engineering and Informatics,Universitas Multimedia Nusantara.
Scientia Garden Jl. Boulevard Gading Serpong, Kab. Tangerang-Banten, 15810.
Email: johan@umn.ac.id

INTRODUCTION

According to the survey on internet penetration and user behavior in Indonesia in 2018, approximately 172 million out of the total population of 264 million were connected to the internet, accounting for approximately 64.8%. The average internet usage in Indonesia for visiting social media platforms was 18.9%, with Facebook, Instagram, and Youtube being the most frequently visited social media platforms (APJII, 2018). The advent of the internet has facilitated the sharing of media, such as information and experiences, by travelers through electronic devices like social media platforms and travel review websites. Travelers typically search for online reviews of destinations before making a definitive decision to visit a particular vacation spot. In general, the availability of the internet has led to the growth of travel blogs, which provide relevant and informative content. Within the tourism sector, social media has experienced significant growth and has great potential as a promotional tool, allowing for the display of captivating and unique photos, such as tourist attractions (Rosanensi et al., 2018; Windasari & Eridani, 2017).

Instagram has emerged as a platform for sharing activities by uploading images or videos, or both, accompanied by text in the form of captions, hashtags, mentions, and emojis. These elements allow users to describe their activities and share them with others, whether publicly, privately, or with close friends (Rosanensi et al., 2018). However, the abundance of information circulating on the internet poses challenges in filtering the necessary information, and it can be time-consuming to read through all available reviews online. If the number of reviews read is limited, doubts or biases may arise (Windasari & Eridani, 2017). The goal of sentiment analysis is to enable computers to recognize the level of positive or negative emotions conveyed in text documents (H. Himawan, W. Kaswidjanti, 2018). In January 2020, the President of the Republic of Indonesia, Mr. Jokowi, began to focus on developing tourism in Labuan Bajo (Presiden, n.d.), making it a priority tourist destination that attracts both foreign and local visitors due to its exotic and appealing landscapes.

The study conducted by (H. Himawan, W. Kaswidjanti, 2018) focuses on sentiment analysis on social media as a recommendation for favorite souvenirs, using the Support Vector Machine algorithm. The analysis yielded an accuracy, precision, and recall rate of 86%, 93.20%, and 100% respectively. The test results on the developed system indicate that the lexicon-based method provides better accuracy and precision compared to the Support Vector Machine method, achieving an accuracy of 87.78% and a precision of 94.23%. However, in terms of recall, the Support Vector Machine method outperforms the lexicon-based method with a recall rate of 100%. Other study conducted by (Phoan & Setiawan, 2022) focus on sentiment analysis of comments on sexual harrasment in colleages on four popular social media, such as Twitter, Instagram, Medium and Line Today using Support Vector Machine (SVM) algorithm. The findings SVM has an accuracy of 55.14% on the dataset collected.

In a different study, conducted by (Sinaga, 2017), sentiment analysis was carried out for ranking the popularity of destinations in Bali using social media platforms such as Facebook, Instagram, and forums. The Naïve Bayes algorithm was employed for this analysis. The validity testing of the algorithm's accuracy yielded two categories of testing: one with 100 phrases, resulting in an accuracy rate of 65.65%, and the other with 5000 phrases, resulting in an accuracy rate of 82.67%. Thus, the conclusion drawn was that the more phrases are used as the core of the algorithm, the more accurate the sentiment analysis presented. Previously,(Fuchs et al., 2013) conducted sentiment analysis research by extracting relevant knowledge from User Generated Content using a combination of dictionary-based and machine learning approaches, specifically utilizing the Support Vector Machine. The dictionary- based approach achieved a good classification performance of 71.28%.

This research is distinct from previous studies as it focuses on the object of research, Labuan Bajo, and employs the dictionary-based VADER (Valence Aware Dictionary for sEntiment Reasoning) algorithm along with the social media platform Instagram. Sentiment analysis is conducted on tourists who have visited Labuan Bajo and expressed their experiences on the social media platform Instagram. The study also aims to identify popular locations visited by tourists through the analysis of Instagram hashtags. The research methodology employs lexicon-based sentiment analysis using the VADER Sentiment model and measures the performance accuracy of the lexicon-based method (Gilbert & E., 2014).

This study utilized the VADER lexicon rule-based sentiment analysis method, with data collected from Instagram. Text captions and hashtags provide valuable feedback and information, allowing us to uncover sentiments expressed by users, such as satisfaction, boredom, disappointment, or sadness. Additionally, popular destinations can be identified through patterns of hashtag usage on the Instagram social media platform. Furthermore, the performance of the VADER sentiment analysis method in predicting the sentiment of text captions from Instagram will be evaluated.

RESEARCH METHODS

The research process was adapted from the studies (Fuchs et al., 2013) and (H. Himawan, W. Kaswidjanti, 2018), with certain adjustments made to align with the objectives of this study.

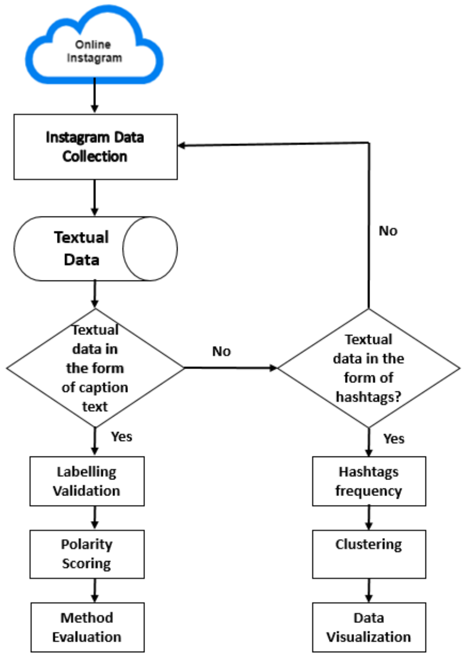

Figure 1. Research Process

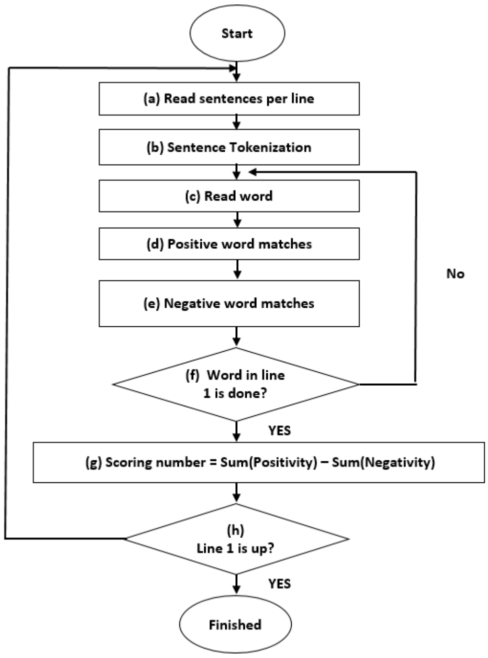

Figure 2. Vader Sentiment Scoring Process

Figure 1 shows the sentiment analysis process consists of two main streams: one for managing text captions (left side) and another for analyzing text hashtags (right side). Data collection is performed using the Instagram-scraper package. The data collected consists of Instagram feeds, including timestamps, text captions, hashtags, likes, comments, and more. The data extraction process from Instagram involves retrieving posts using the keyword "#labuanbajo," which generally identifies content related to places in Labuan Bajo. The resulting data consists of Instagram feeds in JSON format, from which text extraction is conducted to make the obtained data more organized and structured, while removing irrelevant data (spam). The labeling validation stage involves assigning actual sentiment labels using Meaningcloud tools and validating the labels.

Once the labeling validation stage is completed, the text pre-processing stage follows, involving the removal of punctuation, symbols, and special characters, language translation, case folding, and elimination of stopwords. Subsequently, the polarity score stage is conducted by assigning scores using a lexicon-based method and employing the VADER sentiment library, which yields sentiment values indicating whether the sentiment score leans towards positive, negative, or neutral. The score testing stage involves comparing the sentiment examination results obtained from the polarity score (referred to as predicted sentiment labels) with the manually examined sentiment results (actual sentiment labels) obtained from the labeling validation stage. Afterward, the method evaluation stage calculates the performance to determine the reliability level of the lexicon-based sentiment analysis method using the VADER sentiment model.

Figure 2 shows the steps involved in the sentiment scoring process. The manual sentiment assessment process involves determining whether the sentiment of a caption is positive, negative, or neutral. Subsequently, the accuracy is calculated using the following equation:

$$Accuracy = \frac{Number\ of\ correctly\ predicted\ sentiments}{(Total\ number\ of\ sentiments)}$$

This accuracy calculation measures the proportion of correctly predicted sentiments out of the total sentiments assessed manually. It provides an objective evaluation of the sentiment analysis performance and helps gauge the reliability of the polarity scoring approach. By conducting this accuracy evaluation, it is possible to assess the effectiveness of the polarity scoring method in predicting sentiment. However, it is important to note that the scoring approach solely based on word matches may have limitations in capturing the nuanced context and meaning of the text captions. Therefore, further analysis and refinement may be necessary to improve the accuracy and robustness of the sentiment analysis system.

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \text{ x } 100\% \quad (1)$$

In method evaluation, the performance of the lexicon-based sentiment classification method will be evaluated to obtain accuracy, precision, recall, and F1 Score values. The objective is to assess the effectiveness of the developed classification approach and determine its quality. The evaluation will be based on the confusion matrix table, which serves as a reference for assessing accuracy, precision, and recall.

The confusion matrix provides valuable information about True Positive, False Negative, False Positive, and True Negative, which are the outcomes produced by the sentiment analysis system. These metrics are essential for evaluating the performance of the sentiment classification method.

$$Precision = \frac{\text{TP}}{\text{(TP+FP)}} \text{ x } 100\% \quad (2)$$

$$Recall = \frac{\text{TP}}{\text{(TP+FN)}} \text{ x } 100\% \quad (3)$$

$$F1 = \frac{\text{(2 x Precision x Recall)}}{\text{(Precision+Recall)}} \text{ x } 100\% \quad (4)$$

Precision, as defined by Equation (2), measures the proportion of correctly classified positive instances out of all instances predicted as positive. Recall, as calculated by Equation (3), quantifies the proportion of positive instances that were correctly identified by the classification method. F1 Score, calculated using Equation (4), is a composite metric that considers both precision and recall, providing a balanced evaluation of the sentiment classification performance.

The evaluation metrics derived from the confusion matrix and the calculations of precision, recall, and F1 Score allow for a comprehensive assessment of the classification method's accuracy and effectiveness in accurately classifying sentiments. These metrics provide insights into the model's ability to correctly identify positive, negative, and neutral sentiments within the analyzed text data. By conducting this method evaluation, it is possible to gain a deeper understanding of the strengths and limitations of the lexicon-based sentiment classification approach. Furthermore, it enables comparisons with other sentiment analysis methods and facilitates the refinement and improvement of the classification system for more accurate sentiment classification results.

As for the sentiment analysis process for hashtags, data collection involves collecting hashtags that consist of combinations of hashtags used by users to accompany caption texts. The selection of hashtags used by Instagram users can provide insights into the effectiveness and relevance of the hashtags used to indicate the content of their posted photos or videos. Research has indicated a 66% correlation between the use of hashtags and the uploaded content (Fedushko et al., 2019; Giannoulakis & Tsapatsoulis, 2016). Tokopedia has compiled a list of recommended tourist destinations in Labuan Bajo, encompassing 21 beautiful locations such as Gili Laba, Rinca Island, Komodo Island, Kanawa Island, Padar Island, Wae Rebo Traditional Village, Lingko Rice Fields, Kelor Island, Kelimutu Lake, Ranko Cave, Cunca Lawang, Bena Traditional Village, Manta Point, Batu Cermin Cave, Pink Beach, Bidadari Island, Seraya Island, Kalong Island, Melo Village, Pede Beach, and Sano Nggoang Lake (Tokopedia, 2019).

The collected hashtag data undergoes a text pre-processing stage. This stage is necessary because the obtained set of hashtags still contains special characters, thus requiring the removal of hashtag symbols, converting all hashtag text to lowercase (case folding), and separating combined hashtags that previously contained more than one combination of hashtags within them. Following that, calculations are performed on frequently used hashtag words by users. The subsequent stage involves clustering and searching for patterns among the hashtag words that are commonly used, followed by summarizing and grouping them based on similar meanings, such as popular destination categories people visit. Once the clustering process successfully identifies the patterns, the subsequent stage visualizes the obtained hashtag words from the clustering process using a treemap chart to enhance comprehensibility.

By conducting the mentioned steps, valuable insights can be gained regarding sentiment analysis of hashtags, tourist destinations, and the correlation between user-generated content and hashtags. This information can enhance tourism marketing, improve content creation, and personalize experiences in Labuan Bajo. Additionally, meticulous data preparation ensures the inclusion of relevant and reliable data, thereby enhancing the accuracy and validity of findings related to Labuan Bajo as a tourist destination.

RESULTS AND DISCUSSION

Data preparation involved collecting data from 30,000 Instagram posts, followed by filtering and managing the data to focus on the period from December 1st to December 31st, 2019. Out of the collected data, 12,232 entries were identified as irrelevant or spam. During the data collection process, the keyword hashtag "labuanbajo" was used, leading to a significant amount of spam or irrelevant data related to Labuan Bajo as a destination. As a result, only 3,351 valid data entries remained for further analysis.

Irrelevant or spam data removed to ensure accurate labeling and efficient sentiment scoring using lexicon methods, thus avoiding unnecessary processing time. The labeling validation stage involves manually assigning sentiment labels to Instagram caption data. In method evaluation, the performance of the lexicon-based sentiment classification method evaluated to obtain accuracy, precision, recall, and F1 Score values. The evaluation will be based on the confusion matrix table, which serves as a reference for assessing accuracy, precision, and recall. The confusion matrix provides valuable information about True Positive, False Negative, False Positive, and True Negative, which are the outcomes produced by the sentiment analysis system.

The evaluation metrics derived from the confusion matrix and the calculations of precision, recall, and F1 Score allow for a comprehensive assessment of the classification method's accuracy and effectiveness in accurately classifying sentiments. These metrics provide insights into the model's ability to correctly identify positive, negative, and neutral sentiments within the analyzed text data. By conducting this method evaluation, it is possible to gain a deeper understanding of the strengths and limitations of the lexicon-based sentiment classification approach. Furthermore, it enables comparisons with other sentiment analysis methods and facilitates the refinement and improvement of the classification system for more accurate sentiment classification results.

Considering Table 1, the process of transforming raw data into valid data for sentiment analysis reveals that approximately 70% of the entries were irrelevant or spam. If this process were to be carried forward into the sentiment analysis phase, it would likely result in inaccurate information or poor data quality. The significant presence of irrelevant and spam data underscores the importance of thorough data preparation and filtering. Failing to address these issues at an early stage could undermine the accuracy and reliability of the subsequent sentiment analysis. It is crucial to ensure that only relevant and reliable data are used to generate meaningful insights and maintain the overall quality of the obtained information.

Table 1. Sentiment Data Preparation

The amount of caption text data	Data Type
12.232 Caption Text	Raw data
3.351 Caption Text	Valid Data

Source (Author, 2020)

The results of polarity scoring, based on a dataset of 3,351 text captions, revealed that there were 1,962 entries classified as positive sentiment, 230 entries classified as negative sentiment, and 1,159 entries classified as neutral sentiment.

Table 2. VADER Sentiment Prediction Label

Numbers	Sentiments	Percentage
1962	positive	58.55%
230	negative	6.86%
1159	neutral	34.59%
3351	Total	100.00%

Source (Author, 2020)

As depicted in Table 2, the sentiment analysis of tourist destinations in Labuan Bajo indicates a predominantly positive response, with 58.55% of the sentiments being positive. Conversely, negative sentiments accounted for only 6.86% of the total responses.These findings demonstrate the overall favorable sentiment towards the Labuan Bajo tourist destinations. The high percentage of positive sentiments suggests that the majority of individuals have expressed satisfaction, appreciation, or enthusiasm in their captions related to these destinations. On the other hand, the relatively low percentage of negative sentiments indicates a positive perception of the attractions, indicating a generally positive experience among visitors. To evaluate the accuracy of sentiment analysis for Labuan Bajo tourist destinations, a sample of 1,500 data entries was extracted in sequential order, corresponding to the number of labeled entries with actual sentiment labels. This testing phase aimed to assess the accuracy level by comparing the predicted sentiment labels with the actual sentiment labels.

Table 3. Sample Sentiment Score Test 1500 Data

Pos.true	Neu.true	Neg.true	Total Caption	Accuracy results
568	499	15	1500	0.721

Source (Author, 2020)

The results, as shown in Table 3, indicate an accuracy rate of 72%. This level of accuracy signifies a favorable performance of the sentiment analysis model, considering the predominant presence of positive sentiment labels compared to neutral labels. The obtained accuracy rate of 72% reflects the model's ability to correctly classify the sentiment of the sampled data entries. This result demonstrates that the sentiment analysis model is effective in capturing and categorizing the sentiments expressed in the captions related to Labuan Bajo tourist destinations. The higher dominance of positive sentiment labels suggests a generally positive sentiment among the visitors and reflects the overall positive perception of Labuan Bajo as a tourist destination.

The evaluation of the sentiment analysis model's performance was conducted using a test sample of 1,500 sequentially extracted caption data entries. Among these entries, 568 were labeled as positive sentiment.

Table 4. Evaluation of the Sentiment VADER Method

Accuracy	Precision	Recall	F-Measure
72.13%	97.43%	98.10%	97.76%

Source (Author, 2020)

As shown in Table 4, the VADER lexicon rule-based sentiment analysis method yielded promising results with an accuracy of 72.13%, precision of 97.43%, recall of 98.1%, and an F1 score of 97.76%. The evaluation outcomes demonstrate the excellent performance of the chosen method. The achieved accuracy indicates the model's ability to accurately predict sentiment labels for the sampled data entries. The high precision and recall values further affirm the model's effectiveness in correctly identifying positive sentiments. The F1 score, which combines precision and recall, reflects a strong overall performance of the sentiment analysis method.

The findings of this study align with the research conducted by (Gilbert & E., 2014), which reported an F1 score of 96%. Notably, this study achieves an even higher F1 score despite utilizing Instagram, whereas (Gilbert & E., 2014) focused on Twitter. It is noteworthy that the evaluation was conducted on a test sample comprising less than 50% of the total population. Thus, expanding the test data set would likely yield even more robust results. This indicates the potential for further research and development in this area.

The total count of hashtag keywords used by users in their posts about Labuan Bajo tourism destinations amounts to 7,508 distinct keyword instances derived from 3,351 hashtag combinations, aligning with the number of managed text captions. To identify potential clusters related to specific places, clustering analysis was performed on the hashtag keywords. The selection of keyword hashtags was based on the highest frequency counts, capturing the identification of popular tourist sites among the 250 most frequently occurring keyword hashtags. Commonly encountered keyword hashtags include "padarisland," "komodoisland," and "rincaisland," which refer to scenic locations featuring Komodo dragons. Other notable hashtags include "pinkbeach," "kelorisland," "kanawaisland," and "waerebo." Further exploration of similar hashtags was carried out by applying filters to identify keywords that share identical terms. For example, the keyword "pinkbeach" can be filtered to include hashtags containing the term "pink," while "padarisland" can be filtered for the term "padar" alongside other related terms such as "komodo," "kelor," "kanawa," "rinca," and "waerebo."

To streamline the clustering process, a restriction was imposed to focus on the seven most frequently mentioned tourist destinations derived from the 250 most common keyword hashtags, sorted in descending order of frequency. This approach helps prioritize the analysis of the most frequently mentioned locations. The overall outcome of the clustering process revealed seven highly visited tourist destinations in Labuan Bajo. These prominent locations and hashtags include Komodo National Park (2492), Padar Island (1059), Pink Beach (475), Kelor Island (230), Rinca Island (251), Kanawa Island (143) and Waerebo (123). These findings offer valuable insights into the most popular tourist spots in Labuan Bajo, highlighting the significance of attractions such as Komodo National Park and Padar Island among visitors.

The popular tourist destinations in Labuan Bajo, visited by a significant number of people, are visually presented using a treemap chart. The purpose of this visualization is to provide a comprehensive visual representation of the text data, enabling easier comprehension based on the frequency of keywords. The treemap chart visualization encompasses a collection of all the hashtag keywords. Upon observation, it is evident that the primary hashtag that appears prominently is "Labuan Bajo," which aligns with the data retrieval process that employed this specific keyword, as shown in Figure 3.

After eliminating irrelevant hashtag keywords unrelated to Labuan Bajo's tourist destinations, the treemap chart reveals the popular places. Notably, the destination that stands out is Komodo Island, represented by hashtags such as "komodoisland." Additionally, Padar Island is highlighted by the hashtags "pulaupadar" and "padarisland," while Pink Beach is associated with the hashtag "pinkbeach," as depicted in Figure 4.

Figure 3. Collection of Hashtags Words Before Stophashtags Process

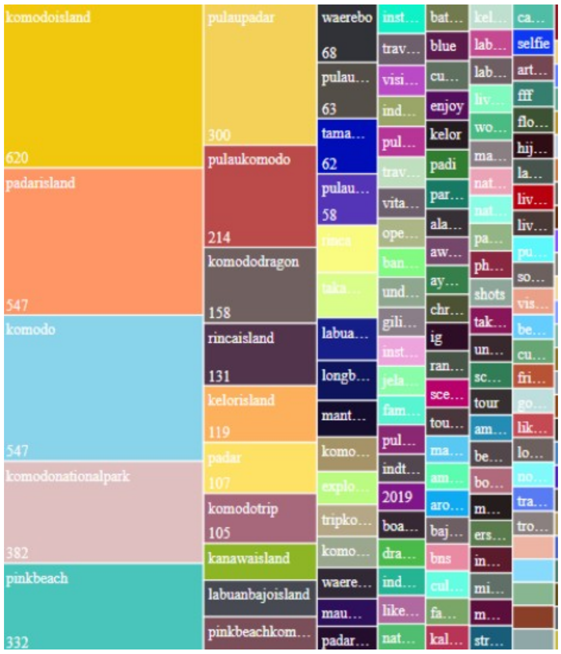

Figure 4. Product Promotion SPAM Data

CONCLUSION AND RECOMMENDATION

Conclusion

Research on sentiment analysis of Labuan Bajo tourism destinations using Instagram data revealed a significant amount of spam or irrelevant data, approximately 70%, which couldn't be effectively managed. Persisting with the inclusion of such data would adversely impact the quality of information. The manual removal of spam data proved challenging due to the diverse patterns in text captions, indicating a potential area for further research in data preparation.

The results obtained through the VADER rule-based sentiment analysis method, which relies on a Lexicon-based approach, demonstrated favorable outcomes with an accuracy exceeding 70%. The test data utilized accounted for less than 50% of the population, suggesting the potential for increased accuracy with larger test datasets. Moreover, future research could explore sentiment analysis utilizing the same method but with different domains, such as Twitter or travel review websites.

The responses from tourists regarding Labuan Bajo's tourist destinations were predominantly positive, as evidenced by 58.55% of the 3,351 text caption data being labeled as positive sentiment. Additionally, the study identified seven popular locations in Labuan Bajo: Komodo Island, Padar Island, Pink Beach, Kelor Island, Rinca Island, Kanawa Island, and the village of Waerebo.

Recommendation

Limitations of the Study:

The study faced challenges in effectively managing a significant amount of spam or irrelevant data, which constituted approximately 70% of the data collected from Instagram. This issue could potentially impact the quality and reliability of the sentiment analysis results, as the presence of spam data may skew the sentiment distribution. Manually removing spam data proved to be difficult due to the varied patterns in text captions. This limitation suggests that the manual approach may not be scalable or efficient for larger datasets. It emphasizes the need for automated techniques or algorithms to address this challenge and ensure the accuracy of sentiment analysis results.

The study relied on a test dataset that represented less than 50% of the population. While the obtained results showed positive outcomes, a larger test dataset would offer a more robust evaluation of the sentiment analysis method's accuracy. Therefore, caution should be exercised when generalizing the findings to the entire population.

The study specifically focused on sentiment analysis of Labuan Bajo tourism destinations using Instagram data. Consequently, the findings may not fully encompass sentiments expressed on other social media platforms or travel review websites. Generalizing the results to the broader sentiment landscape may require conducting similar studies across multiple platforms to obtain a comprehensive understanding of tourist sentiments towards Labuan Bajo.

Implication

The application of sentiment analysis in the context of Labuan Bajo's tourism industry extends beyond marketing and service enhancements. It also offers an opportunity for proactive reputation management. By closely monitoring sentiments and opinions shared by tourists on social media, tourism organizations can promptly address any negative sentiment or complaints that arise. This proactive approach not only demonstrates a commitment to customer satisfaction but also helps in mitigating the potential impact of negative reviews or viral complaints that can quickly spread through online platforms.

Furthermore, sentiment analysis can provide valuable insights into emerging trends and shifting preferences among tourists. By identifying patterns and sentiments associated with specific attractions, activities, or services, tourism stakeholders can adapt their offerings to align with changing demands. This flexibility allows for the continuous improvement and innovation of tourism products and experiences, ensuring that Labuan Bajo remains a compelling and competitive destination in the ever-evolving tourism market.

Future Research :

Future research should concentrate on developing advanced techniques to effectively manage spam or irrelevant data in sentiment analysis of social media platforms. This could involve exploring machine learning algorithms, natural language processing techniques, and advanced data filtering approaches to automate the removal of spam data. Given the diverse patterns in text captions that posed challenges for manual spam data removal, further research should delve into developing algorithms or methodologies that can accurately identify and classify spam or irrelevant data. This could entail the application of machine learning techniques, such as deep learning or ensemble models, to effectively identify patterns and differentiate between relevant and irrelevant content.

While the current study focused on sentiment analysis of Labuan Bajo tourism destinations using Instagram data, future research could expand the scope by applying the same sentiment analysis method to different domains, such as Twitter or travel review websites. This would provide insights into sentiment trends across multiple platforms and validate the effectiveness of the sentiment analysis method across diverse data sources.

The study achieved an accuracy exceeding 70% using the VADER rule-based sentiment analysis method. However, the test data represented less than 50% of the population. Future research should aim to increase the size of the test datasets to enhance the accuracy and reliability of the sentiment analysis results. This could involve collecting and labeling a larger sample of data to provide a more comprehensive representation of sentiments expressed towards Labuan Bajo's tourist destinations.

ACKNOWLEDGMENT

The successful completion of this research has been made possible thanks to the unwavering support and invaluable assistance provided by Universitas Multimedia Nusantara.

REFERENCES

APJII. (2018). *Penetrasi dan Profil Perilaku Pengguna Internet Indonesia* (p. 51).

Fedushko, S., Syerov, Y., & Kolos, S. (2019). Hashtag as a way of archiving and distributing information on the internet. *CEUR Workshop Proceedings*, *2386*, 274–286.

Fuchs, M., Lexhagen, M., Hopken, W., & Schmunk, S. (2013). Sentiment Analysis : Extracting Decision-Relevant Knowledge from UGC. *Inf. Commun. Technol. Tour*, 253–265. https://link.springer.com/chapter/10.1007/978-3-319-03973-2_19

Giannoulakis, S., & Tsapatsoulis, N. (2016). Evaluating the descriptive power of Instagram hashtags. *Journal of Innovation in Digital Ecosystems*, *3*(2), 114–129. https://doi.org/10.1016/j.jides.2016.10.001

Gilbert, C. J. H., & E., E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proc. 8th Int. Conf. Weblogs Soc. Media, ICWSM 2014,*. https://www.mendeley.com/catalogue/cea03574-e257-3e68-89c9-7156ed7e4732/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7Bf37d2106-be67-4d2b-a9b3-277ecdb35437%7D

H. Himawan, W. Kaswidjanti, and G. D. P. (2018). "Metode Lexicon Based Dan Support Vector Machine Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit. *Semin. Nas.Inform.*, *2018*(November), 235–244.

Phoan, V., & Setiawan, J. (2022). Sentiment Analysis Of Comments On Sexual Harrasment In Colleges On Four Popular Social Media. *Journal of Multidisciplinary Issues*, *2*(2), 1–21. https://doi.org/10.53748/JMIS.V2I2.33

Presiden, S. (n.d.). *Sekretariat Presiden - YouTube." [Online].* YouTube. Retrieved April 17, 2022, from https://www.youtube.com/channel/UC_m_NBgf7ieJBHzb6vvJC5A/search?query=labuan+bajo.

Rosanensi, M., Madani, M., Wanggono, R. T. P., Setyanto, A., Selameto, A. A., & Wahyuni`, S. N. (2018). Analysis sentiment and tourist response to Rinjani mountain tour based on comments from photo upload in Instagram. *Proc. - 2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2018*, 184–188.

Sinaga, A. (2017). Pemeringkatan Popularitas Tujuan Wisata Utilization Of Sentiment Analysis For Tourist. *Jurnal Penelitian Pos Dan Informatika*, *7*(2), 109–120.

Tokopedia. (2019). *21 Objek Wisata Labuan Bajo Terbaik & Populer - Tokopedia Blog*. Tokopedia Blog. https://www.tokopedia.com/blog/travel-objek-wisata-labuan-bajo/

Windasari, I. P., & Eridani, D. (2017). "Sentiment analysis on travel destination in Indonesia,. *Proc. - 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2017*, 276–279.