Edu Komputika 11 (1) (2024)

Edu Komputika Journal

https://journal.unnes.ac.id/journals/edukom/

Sentiment Analysis of Visitor Reviews on Baturaden Tourist Attraction Using Machine Learning Methods

Mahazam Afrad[1)]✉, Dany Candra Febrianto[2)], Sena Wijayanto[1)], M Yoka Fathoni [1)]

[1]Department of Information System, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia

[2]Department of Informatics, Fakultas of Informatics, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia

Article Info

Article History:
Received: 6 August 2024
Revised: 25 August 2024
Accepted: 29 August 2024

Keywords:
K-Nearest Neighbors, Naive Bayes, Random Forest, Sentiment Analysis, SVM, Tourist Attraction, Visitor Reviews

Abstract

This study evaluates the performance of four machine learning models: Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Naive Bayes in analyzing visitor reviews of the Lokawisata Baturaden tourist attraction. Using 5-fold cross-validation, the study aims to determine which machine learning model best suits sentiment analysis on the Baturaden review data. This study was conducted through several stages, including data preprocessing, feature extraction, and the data training process. Case folding, text cleaning, tokenization, stopword removal, and stemming were performed during the data preprocessing stage. The feature extraction method used was TF-IDF. SMOTE was applied to increase data variation and address the data imbalance in the dataset. The results show that SVM provides the best performance with an accuracy of 0.937, an F1-score of 0.937, a precision of 0.943, and a recall of 0.937. Random Forest also performs well with an accuracy of 0.918 and an F1-score of 0.918, though slightly below SVM. KNN shows the lowest performance with an accuracy of 0.651 and an F1-score of 0.544, while Naive Bayes performs adequately with an accuracy of 0.845 and an F1-score of 0.841. Based on this evaluation, SVM is recommended as the best model for sentiment analysis of reviews, followed by Random Forest as a good alternative. The KNN model is not recommended due to its lower performance, while Naive Bayes can be considered for its speed and simplicity, although its results are not as good as SVM and Random Forest. These conclusions guide the selection of the optimal model to enhance understanding and visitor experience at the Baturaden tourist attraction.

✉ Correspondence address:
Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan No 128, Karangreja, Purwokerto, 53147 Indonesia
E-mail: mahazamafrad@ittelkom-pwt.ac.id

ISSN 2252-6811
E-ISSN 2599-297X

INTRODUCTION

Tourism is an economic sector essential in increasing regional income and the local community's well-being. One of the well-known tourist destinations in Indonesia is Baturaden Tourist Attraction, located at the foot of Mount Slamet, Central Java. To improve the quality of services and the visitor experience, tourism managers must understand the perceptions and sentiments of visitors towards the destination.

In the digital era, online reviews have become an essential source of information for tourists and destination managers. Google Maps is one of the most well-known worldwide services (Mathayomchan & Sripanidkulchai, 2019). Reviews left by visitors on platforms like Google Maps can provide valuable insights into their experiences. Therefore, sentiment analysis of these reviews can help managers identify the positive and negative aspects of the services offered and design more targeted improvement strategies.

Sentiment analysis is a Natural Language Processing (NLP) technique used to identify and classify opinions or emotions in text (Das et al., 2021). By applying sentiment analysis to Google Maps reviews, the managers of Baturaden Tourist Attraction can gain an overall picture of the feelings and opinions of visitors regarding various aspects of the destination, such as facilities, cleanliness, prices, and services. Additionally, the intensity of tourist interactions through various digital communication media creates a new habit in the virtual space, namely the thematic dissemination of information based on popularity. This indicates that positive sentiment from social media users, in the form of published information, can influence their motivation and decision to visit certain tourist attractions and vice versa (Singgalen, 2021).

Research on sentiment analysis of a tourist attraction has been conducted previously. The study by (Merawati et al., 2021)identified public sentiment on social media towards Lombok Island as one of the favorite tourist destinations in Indonesia. This research aimed to classify tourist opinions into positive and negative categories and perform topic modeling on each category. The investigative stages included information collection, cleaning, and change, the Naive Bayes method, and topic modeling with (LDA). The performance results of the Naive Bayes model showed an accuracy of 92%, precision of 100%, recall of 83.84%, and specificity of 100%. Topic modeling with LDA showed the highest coherence score of 0.613 in the positive category and 0.528 in the harmful category.

Another study (Era et al., 2023) analyzed Indonesian open estimation towards government approaches on reviving traveler goals amid the Covid-19 widespread utilizing the Naïve Bayes and K-Nearest Neighbor (KNN) calculations. Information was collected from different social media stages, such as Twitter, Instagram, and Facebook, employing an information slithering preparation. The dataset used consisted of 570 tweets with 12 hashtags related to tourism in Indonesia. The results showed that the Naïve Bayes algorithm performed better than the K-Nearest Neighbor in this sentiment analysis. Naïve Bayes achieved the highest accuracy of 75.53%, positive precision of 71%, negative precision of 25%, positive recall of 99%, and negative recall of 14%.

Meanwhile, KNN showed the highest accuracy of 48.66%, positive precision of 69%, negative precision of 14%, positive recall of 69%, and negative recall of 28%. This study demonstrates that Naïve Bayes is more effective in classifying public opinion about the reopening of tourist destinations during the pandemic, providing more accurate insights into social media sentiment analysis. This study aims to determine which machine learning model is best suited for sentiment analysis on the Baturaden review data. This analysis is expected to provide helpful information for managers to improve the quality of the tourist destination and visitor experience.

RESEARCH METHODS

The research method includes several stages, namely data collection of reviews through web scraping techniques on Google Maps reviews, which are then manually labeled with positive, negative, and neutral labels. Positive is the text that expresses positive feelings or opinions. Negative is the text that conveys negative emotions or criticism. Neutral is text that is neutral or doesn't express strong emotions. The next step is data preprocessing to clean and prepare the review texts and applying sentiment analysis algorithms to classify the sentiments of these reviews. The stages undertaken can be seen in Figure 1, which illustrates the research methodology.

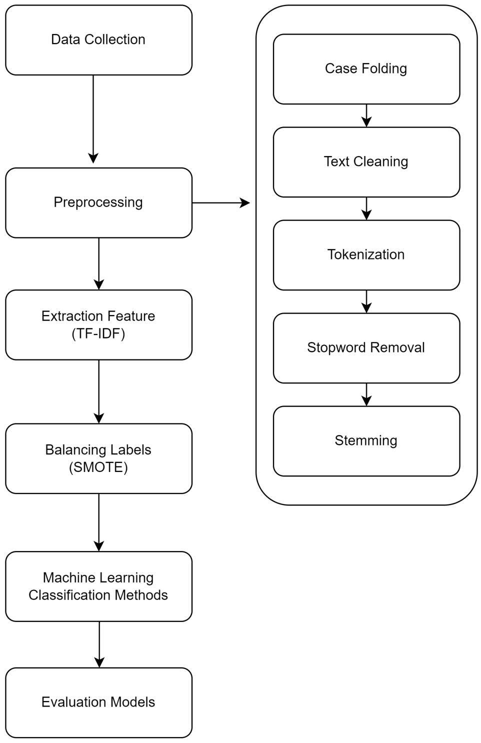

Figure 1. Research methodology diagram

A. Data Collection

Data collection is the first step carried out in this research. Data was obtained from Google Maps reviews of Baturaden Tourist Attraction. Web scraping techniques were used to collect these reviews. A total of 902 reviews were successfully collected. The data in this study were collected in July 2024, from 2020 to July 2024. The collected data consisted of text data in Indonesian. Once the data was collected, each review was manually labeled based on its sentiment. The manual labeling process was conducted to ensure accuracy in sentiment classification. These reviews were grouped into three sentiment categories: positive, neutral, and negative.

Labeling resulted in 623 reviews being labeled as positive, indicating that most visitors had a satisfying experience at Baturaden Tourist Attraction. A total of 228 reviews were labeled as neutral, reflecting visitors did not express strong feelings either positively or negatively. Meanwhile, 36 reviews were labeled as negative, indicating some dissatisfaction or issues experienced by visitors. The distribution of labels can be seen in Figure 2.

Table 1. Sample Data

No	Label	Review
1	Positive	Tempat wisata yang sejuk dan menyenangkan, kalau kesini sih cocok bareng keluarga sih. Tempatnya luas cocok buat piknik, dengan HTM 25 RB /orang. Bisa menikmati indahnya pemandangan dan fasilitas yang tersedia juga sudah memadai, banyak jajanan dan berbagai macam aksesoris yang dijual .. Cocok untuk dijadikan oleh-oleh
2	Positive	tempat yang wajib dikunjungi jika main ke purwokerto. dengan htm 25k sudah sangat worth it karena lokasi nya yg luas. didalam kita bisa melihat taman, air terjun, dan sungai. ad beberapa wahana yg bisa dimainkan namun berbayar seperti perahu angsa, pesawat dan film 3d. ad waterboom juga yg cocok buat anak2 main.
3	Positive	Tempatnya sangat luas dengan indahnya air terjun hawanya juga sejuk, banyak pedagang juga dan fasilitas lainnya apalagi ada 2 kolam berenang ini cocok buat tempat liburan keluarga😉👍🏻🍃
4	Neutral	Untuk tempat luas, untuk tempat bermain atau spot2 fotonya mungkin bisa lebih di upgrade jadi lebih modern hehe, karna suasana old-nya berasa bgt. Cobain juga getuk goreng khas baturaden kalo kesini yaa Htm: 25k/org
5	Neutral	Batu Raden salah satu wisata menarik di Banyumas, tempatnya adem, rapih, banyak pedagang UMKM tapi tertata rapi, masuknya membayar 25 rb, tapi fasilitasnya lumayan komplit petugasnya ramah Disekelilingnya banyak tempat menginap untuk kita singgah, tapi aliran sungainya sudah tidak bisa dimasuki karena pernah kejadian banjir bandang.
…	…	…
902	Positive	Kalau ke purwokerto pasti tidak lupa berkunjung ke baturaden. Entah sudah beberapa kali datang kesini, tapi tetap tidak bosan karena keindahan alamnya yang mempesona.

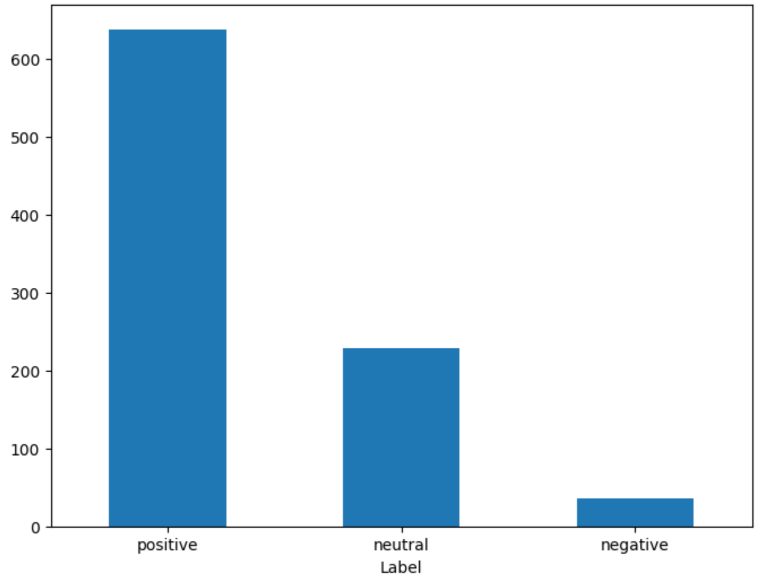

Figure 2. Label distribution

The process of data collection and labeling is crucial to ensure the quality of the data to be used in sentiment analysis. With accurately labeled data, this research can proceed to the next stage, namely data preprocessing and the application of sentiment analysis algorithms. A sample of the dataset can be seen in Table 1.

B. Preprocessing

The preprocessing stage is carried out to process the collected data so that it can be used and has value in classification. Generally, text data has unstructured and multidimensional patterns, requiring a data cleaning process (Fahmi et al., 2020). The preprocessing stages used in this research are case folding, text cleaning, tokenization, stopword removal, and stemming. The case folding, text cleaning, tokenization, and stopword removal stages use the help of the NLTK library with the Indonesian language corpus, and the stemming process uses the help of the Sastrawi library.

1. Case Folding

Case Folding is a process where text consisting of various uppercase and lowercase letters is standardized. An example of a word before and after the case folding stage:

Tempat -> tempat
Tahun -> tahun

2. Text Cleaning

At this stage, text data from the dataset is cleaned by removing symbols, characters, URLs, hashtags, and punctuation marks.

3. Tokenization

Tokenization is a process where sentences are broken down into words or tokens. This step is crucial for text analysis as it converts the unstructured text into a structured format that machine learning algorithms can more easily process. An example of the tokenization process: ‘tempat wisata yang sejuk dan menyenangkan’, becomes ‘tempat’, ‘wisata’, ‘yang’, ‘sejuk’, ‘dan’, ‘menyenangkan’.

4. Stopword Removal

At the stopword removal stage, common words that are considered to have no meaning and, more often than not, appear in huge amounts in a content report will be evacuated (Marutho et al. 2018). Examples of words that are removed include ‘dan,’ ‘atau,’ ‘yang,’ and others.

5. Stemming

Stemming is used to convert words to their root form. In this research, the Sastrawi library, which has a dictionary of Indonesian root words, is used. An example of the stemming process is the word ‘menyenangkan’ becoming the root word ‘senang’.

C. Feature Extraction

The TF-IDF presented in Equation 1 is used to assign weights to words in content documents. This procedure is a compelling text-based highlight extraction methodology that gives precisely what comes around. This technique works by combining two weighting concepts: the repeat of the word occasion in a report and the banter of the repeat of that word's occasion (Gifari et al., 2022).

$$\text{TF-IDF(t)} = \text{TF} \times \log\left(\frac{\text{N}}{\text{df}}\right) \qquad (1)$$

Within the TF-IDF equation over, t speaks to the term, TF is the overall number of times term t shows up in a record, N is the number of archives, and df is the number of archives containing term t (Afrad, 2024).

D. SMOTE

Imbalanced data is when information has an uneven proportion of one course to another, resulting in a more significant part of a lesson and a minority course. Forming forecasts on imbalanced datasets is troublesome since classifiers tend to detect the more significant part of the course instead of the minority course. Resampling strategies are one of the foremost viable ways of understanding this imbalanced information issue. One category of resampling procedures is oversampling. The oversampling method is SMOTE (Haryawan & Ardhana, 2023).

SMOTE is used to address imbalances in class or label distribution within data. The SMOTE algorithm applies an oversampling technique to rebalance the original training set. Rather than merely duplicating examples from the minority class, SMOTE's primary approach is to generate synthetic examples. The use of SMOTE in this research is due to the label imbalance observed, which requires a method to achieve balance among the labels (Magnolia et al., 2023).

E. Machine Learning Classification Methods

This research compares the accuracy results of several machine learning classification models, including Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). The results of this study will determine which classification algorithm method has better accuracy. By evaluating and comparing these models, we aim to identify the most effective algorithm for sentiment analysis of visitor reviews. This comparison is significant for selecting the finest approach to precisely classify estimations and give important experiences for improving the quality of administrations at Baturaden Traveler Fascination.

F. Evaluation

In this consideration, we evaluate the execution of classification models through cross-validation and confusion matrix techniques. These strategies empower us to pick up a more comprehensive understanding of the model's adequacy, guaranteeing more solid results and making strides in generalization.

Cross-validation assesses model performance by partitioning the data into multiple subsets (Rintyarna et al., 2019). In this consideration, we utilize 5-fold cross-validation, which suggests the information is isolated into five similarly measured subsets. The steps are as follows:

1. Data Splitting: The data is divided into five equal-sized segments.
2. Training and Testing Iteration: The model is trained on four segments and tested on the remaining segment. This procedure is repeated five times, each segment serving as the test data once.
3. Performance Averaging: The results from each iteration are averaged to offer a more stable and reliable estimate of the model's performance.

Once the model has been trained and tested using cross-validation, the prediction results are compared to the actual values using a confusion matrix. The confusion matrix offers a comprehensive view of the number of correct and incorrect predictions, which is then used to compute performance metrics such as accuracy, precision, recall, and F1-score. Each matrix can be seen in Equations 2, 3, 4, and 5.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (4)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \qquad (5)$$

Main Components of Confusion Matrix is described as follows:

1. True Positive (TP): Correct predictions for the positive class.
2. True Negative (TN): Correct predictions for the negative class.
3. False Positive (FP): Incorrect predictions where the model predicts positive when it should be negative (Type I error).
4. False Negative (FN): Incorrect predictions where the model predicts negative when it should be positive (Type II error).

RESULT AND DISCUSSION

This study illustrates the overall methodology used for sentiment analysis of visitor reviews on the Baturaden tourist attraction, starting from data collection and progressing through preprocessing, feature extraction, and label balancing. The preprocessing phase includes essential tasks such as case folding, text cleaning, tokenization, stopword removal, and stemming, which prepare the text data for further analysis. Following preprocessing, features are extracted using TF-IDF, and SMOTE is applied to balance the labels.

After SMOTE was applied, the label distribution became balanced, with each class (neutral, positive, and negative) having approximately the same number of samples, around 500. The use of SMOTE in this study aimed to address the issue of data imbalance, which is common in sentiment analysis, where one class may have significantly more data than the others. Such an imbalance can cause the model to overlook the minority class, leading to reduced overall classification accuracy. By using SMOTE, the number of samples in the minority class is synthetically increased, allowing the model to be trained more equitably and resulting in improved performance in recognizing all classes. The distribution of labels after using the smote technique can be seen in Figure 3.

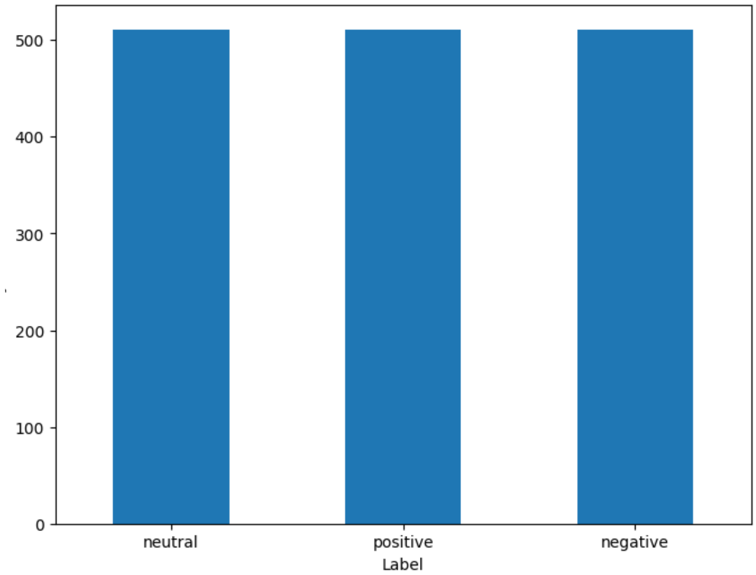

Figure 3. SMOTE label distribution

The processed data is then fed into various machine learning classification methods, and the final models are evaluated to assess their performance. The models used include Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Naive Bayes. Evaluation was performed using 5-fold cross-validation to ensure the results were more accurate and reliable. The following are the results and discussion for each model:

Table 2. Model Performance Comparison

Model	Accuracy	F1-Score	Precision	Recall
SVM	0.937	0.937	0.943	0.937
Random Forest	0.918	0.918	0.947	0.944
KNN	0.651	0.544	0.947	0.944
Naive Bayes	0.845	0.841	0.850	0.845

Table 2 presents a comparison of model performance across four different machine learning algorithms: SVM, Random Forest, KNN, and Naive Bayes. The performance metrics evaluated include Accuracy, F1-Score, Precision, and Recall.

1. SVM demonstrates the highest performance overall, with an Accuracy and F1-Score of

0.937, a Precision of 0.943, and a Recall of 0.937.
2. Random Forest follows closely with an Accuracy and F1-Score of 0.918, slightly higher Precision at 0.947, and a Recall of 0.944, indicating strong performance.
3. KNN shows the weakest performance among the models, with a low Accuracy of 0.651 and an F1-Score of 0.544, despite having high Precision (0.947) and Recall (0.944). This suggests that while KNN is good at predicting positives, it struggles significantly with overall classification.
4. Naive Bayes performs moderately, with an Accuracy of 0.845, an F1-Score of 0.841, a Precision of 0.850, and a Recall of 0.845, showing balanced but not outstanding results across all metrics.

The comparison table above shows that the SVM model provides the best overall performance, followed by Random Forest. The KNN model shows the lowest performance in terms of accuracy and F1-score, while Naive Bayes delivers reasonably good results but still falls below SVM and Random Forest.

Figure 4 shows that the SVM model achieves the highest accuracy compared to the other models, with Random Forest and Naive Bayes following in performance. The KNN model shows the lowest accuracy, indicating that this model may be less suitable for the dataset used in this study. Therefore, the SVM model is the most recommended for sentiment analysis of visitor reviews of the Baturaden tourist attraction.

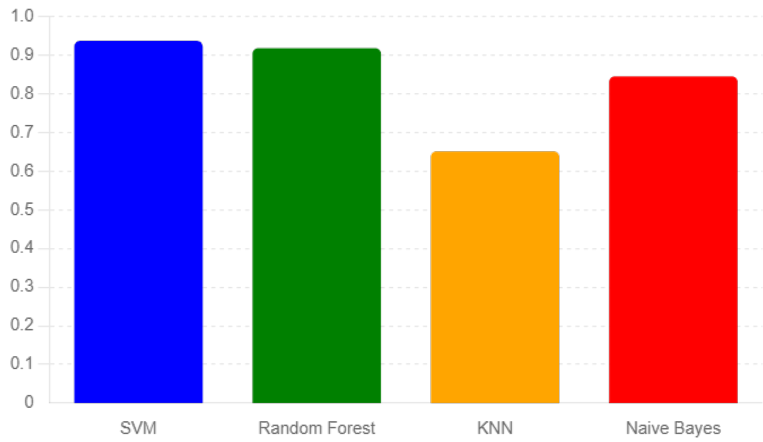

Figure 4. Model accuracy comparison

CONCLUSION

This study has evaluated the performance of various machine learning models for sentiment analysis of visitor reviews for the Baturaden tourist attraction. An evaluation was conducted using cross-validation with a 5-fold split and a confusion matrix analysis. The results indicate that the SVM model delivers the best overall performance, achieving an accuracy score of 0.937, an F1 score of 0.937, a precision of 0.943, and a recall of 0.937. This model is highly effective in classifying the sentiment of visitor reviews. Random Forest also shows excellent performance with high precision and recall, although slightly below SVM in accuracy and F1-score. In contrast, KNN shows less satisfactory performance with lower accuracy and F1-score, while Naive Bayes provides fairly good results with an accuracy score of 0.845 and F1-score of 0.841.

Based on the evaluation results, the Support Vector Machine (SVM) model is the most recommended for sentiment analysis of visitor reviews for the Baturaden tourist attraction, given its superior performance across all key evaluation metrics. Random Forest is also a good alternative, especially due to its high precision and recall. However, K-Nearest Neighbors (KNN) is not recommended due to its lower performance than the other models. Naive Bayes can be considered if speed and simplicity are priorities, although its results are not as good as SVM and Random Forest. Therefore, using the SVM model will provide the most optimal results in sentiment analysis of visitor reviews, helping to understand and improve visitor experiences at the Baturaden tourist attraction.

The findings from this sentiment analysis can be applied to enhance marketing strategies and improve visitor experiences at the Baturaden tourist attraction. Marketers can identify key strengths and areas for improvement by analyzing visitor feedback. Positive aspects can be highlighted in promotional efforts, while negative feedback can guide targeted improvements. This allows for a more tailored marketing approach and better alignment with visitor preferences.

Additionally, the insights can inform product and service development, helping to introduce new features that align with visitor desires. The analysis also aids in crisis management by quickly identifying and addressing issues, thereby protecting the attraction's reputation. Overall, sentiment analysis provides valuable data for making informed decisions that enhance visitor satisfaction and maintain a competitive edge.

REFERENCES

Afrad, M., Muljono, M., & Pujiono, P. (2024). Utilization Of Principal Component Analysis To Improve Emotion Classification Performance In Text Using Artificial Neural Networks. *Journal of Applied Intelligent System*, *9*(1), 8-18.

Haryawan, C., & Ardhana, Y. M. K. (2023). Analisa Perbandingan Teknik Oversampling Smote Pada Imbalanced Data. *Jurnal Informatika dan Rekayasa Elektronik*, *6*(1), 73-78. doi: 10.36595/jire.v6i1.834.

Das, A., Gunturi, K. S., Chandrasekhar, A., Padhi, A., & Liu, Q. (2021, December). Automated pipeline for sentiment analysis of political tweets. In *2021 international conference on data mining workshops (icdmw)* (pp. 128-135). IEEE.

Era, D., Andryana, S., & Rubhasy, A. (2023). Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, *7*(1), 263-272.

Fahmi, S., Purnamawati, L., Shidik, G. F., Muljono, M., & Fanani, A. Z. (2020, September). Sentiment analysis of student review in learning management system based on sastrawi stemmer and SVM-PSO. In *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 643-648). IEEE. doi: 10.1109/iSemantic50169.2020.9234291.

Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, *2*(1), 36-40.

Magnolia, C., Nurhopipah, A., & Kusuma, B. A. (2022). Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter. *Edu Komputika Journal*, *9*(2), 105-113. doi: 10.15294/edukomputika.v9i2.61854.

Marutho, D., Handaka, S. H., & Wijaya, E. (2018, September). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *2018 international seminar on application for technology of information and communication* (pp. 533-538). IEEE. doi: 10.1109/ISEMANTIC.2018.8549751.

Mathayomchan, B., & Sripanidkulchai, K. (2019, July). Utilizing Google translated Reviews from Google maps in sentiment analysis for Phuket tourist attractions. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 260-265). IEEE. doi: 10.1109/JCSSE.2019.8864150.

Putu, N. L. P. M., & Amrullah, A. Z. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *5*(1), 123-131. doi: 10.29207/resti.v5i1.2587.

Setya Rintyarna, B., Sarno, R., & Fatichah, C. (2019). Semantic features for optimizing supervised approach of sentiment analysis on product reviews. *Computers*, *8*(3), 55. doi: 10.3390/computers8030055.

Singgalen, Y. A. (2021). Analisis sentimen dan pemodelan topik dalam optimalisasi pemasaran destinasi pariwisata prioritas di Indonesia. *Journal of Information Systems and Informatics*, *3*(3), 459-470. doi: 10.51519/journalisi.v3i3.171.