Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 9 No. 4 (2025) 706 - 713 e-ISSN: 2580-0760

Optimizing Sentiment Analysis for Lombok Tourism Using SMOTE and Chi-Square with Machine Learning

Hairani Hairani[1]*, Anthony Anggrawan[2], Muhammad Ridho Akbar[3], Khasnur Hidjah[4], Muhammad Innuddin[5]

[1 3 4 5]Department of Computer Science, Faculty of Engineering, Universitas Bumigora, Mataram, Indonesia

[2]Department of Information Technology Education, Faculty of Education, Universitas Bumigora, Mataram, Indonesia

[1]hairani@universitasbumigora.ac.id , [2]anthony.anggrawan@universitasbumigora.ac.id , [3]2001010152@universitasbumigora.ac.id, [4]khasnur72.h@universitasbumigora.ac.id, [5]inn@universitasbumigora.ac.id

Abstract

Tourism is a vital economic sector for Lombok Island, which is renowned for its natural beauty and cultural richness as a top destination. The rapid growth of tourism in Lombok requires a deep understanding of tourists' perceptions and sentiments to ensure an optimal service quality. The sentiment analysis of online reviews is valuable for identifying service strengths and weaknesses and addressing tourists' needs more effectively. This not only enhances tourist satisfaction, but also aids in the design of more effective marketing strategies. However, text data analysis from online reviews presents unique challenges such as noise, class imbalance, and numerous features that may affect classification results. Therefore, this study aims to classify tourist sentiment toward Lombok tourism using machine learning methods combined with feature selection and oversampling techniques. This study focuses on optimizing sentiment analysis of tourism-related tweets using a combination of SMOTE oversampling and Chi-Square feature selection on improving classification performance without hyperparameter tuning. The study applies machine learning methods, such as SVM and Naïve Bayes, with feature selection and oversampling using Chi-Square and SMOTE. The dataset used was sentiment data regarding Lombok tourism obtained from Twitter in 2023, consisting of 940 instances divided into three classes: Negative, Neutral, and Positive. The research findings show that the use of SMOTE and Chi-Square can improve the accuracy of the SVM and Naive Bayes methods. Without optimization, the SVM method achieved an accuracy of 73.93% and a Naive Bayes of 67.02%. After optimization with SMOTE and Chi-Square, the accuracy increased for SVM by 90% and Naive Bayes by 84% to classify tourist sentiment towards Lombok tourism. The implications indicate that combining data balancing using SMOTE with feature selection via Chi-Square effectively improves the performance of sentiment classification models for tourist opinions on Lombok's tourism.

Keywords: chi-square feature selection; optimization of classification methods; SMOTE oversampling; tourism sentiment

How to Cite: Hairani, Anggrawan, A., Muhammad Ridho Akbar, Khasnur Hidjah, & Muhammad Innuddin. (2025). Optimizing Sentiment Analysis for Lombok Tourism Using SMOTE and Chi-Square with Machine Learning. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *9*(4), 706 - 713. https://doi.org/10.29207/resti.v9i4.6623

*Permalink/*DOI: https://doi.org/10.29207/resti.v9i4.6623

Received: May 4, 2025

Accepted: June 23, 2025

Available Online: July 13, 2025

1. Introduction

Tourism is a crucial economic sector for many countries, including Indonesia [1]. Lombok, an island in Indonesia known for its stunning natural beauty and unique cultural heritage, has become a leading tourist destination, attracting thousands of visitors annually [2]. This rapid growth in tourism requires a deep understanding of tourist perceptions and sentiments to ensure optimal service quality and a satisfying travel experience [3]. Sentiment analysis is an essential tool in this context as it allows destination managers to gain valuable insights into tourists' opinions and feelings about various aspects of tourism in Lombok [4].

In today's digital era, tourists increasingly share their experiences and reviews through social media and online review platforms [5]. This data serves as a valuable resource for analyzing tourist sentiments. Sentiment analysis enables tourism managers to identify the strengths and weaknesses of their services and respond more effectively to tourists' needs and expectations [6]. Thus, sentiment analysis helps enhance tourist satisfaction and supports the development of more effective and sustainable marketing strategies [6]. Tourism stakeholders can utilize the results of such analyses to design more targeted service improvement strategies. However, analyzing textual data from online reviews is challenging [7]. Review data often contains noise and class imbalances that can affect the accuracy of the analysis [8]. Several previous studies have employed various approaches to classify tourist sentiments regarding tourism.

Study [9] examined Saudi tourism sentiments using Naïve Bayes and Support Vector Machine (SVM), with SVM demonstrating superior accuracy. In contrast, study [10] analyzed Thailand tourism during the COVID-19 pandemic using SVM, Decision Tree, and Random Forest, where Random Forest yielded the best performance. Similarly, study [11] focused on Saudi tourism sentiment analysis, utilizing Recurrent Neural Networks (RNN), SVM, and Long Short-Term Memory (LSTM) models. Furthermore, research [12] investigated sentiment analysis for Indonesian tourism, utilizing classification methods such as Decision Tree , SVM Naïve Bayes, and K-Nearest Neighbors (KNN). The results indicated that SVM algorithm outperformed the other three classification methods.

Study [13] analyzed customer satisfaction for three major online travel agencies—Traveloka, Tiket, and Agoda using SVM, KNN, and Naïve Bayes for sentiment evaluation. The results showed that KNN had the highest F1-score compared to SVM and Naïve Baye. Study [14] employed the LSTM method for sentiment analysis of reviews about West Sumatra tourist destinations in Indonesia. Study [15] examined tourist sentiments regarding Nusantara Temples on TripAdvisor using Logistic Regression (LR), SVM, and Stochastic Gradient Descent (SGD) classification algorithms. The results indicated that LR outperformed the other methods in sentiment classification for Nusantara Temples. Study [16] applied various classification methods, including KNN, Naïve Bayes, SVM, and Decision Tree, with SMOTE for sentiment classification in cultural and heritage tourism. The findings revealed that SVM with SMOTE performed better than the other three methods. Research [17] conducted sentiment analysis of Bali tourism using the RNN method with 80% accuracy. Research [18] conducting sentiment analysis on Yogyakarta tourism using the Naïve Bayes method.

Despite various approaches being applied, previous studies still have several gaps. One major issue is the lack of attention to class imbalance in tourist review datasets, which can lead to machine learning models being biased and less accurate in predicting minority sentiments [17]. Additionally, suboptimal feature selection affects the model's performance in sentiment analysis. This study addresses these shortcomings by combining several techniques to improve sentiment analysis accuracy for Lombok tourism which is different from previous research. First, the Chi-Square feature selection method is used to identify the most significant features, reducing data dimensions and enhancing model performance. Second, the SMOTE is implemented to handle the problem of imbalanced class distribution by generating synthetic minority data to balance class distribution. Finally, machine learning methods, such as SVM and Naïve Bayes, are applied and compared to identify the best model for classifying tourist sentiments. This study aims to enhance machine learning accuracy in analyzing tourist sentiments on Lombok tourism through feature selection and SMOTE oversampling. This research contributes a novel combination of SMOTE oversampling and Chi-Square feature selection to improve sentiment classification accuracy on tourism-related tweets. The proposed method addresses the common issue of class imbalance and irrelevant features in textual tourism data, resulting in a more reliable sentiment interpretation that can support tourism service providers in tailoring their offerings.

2. Methods

Figure 1 shows that the research process begins with collecting tourist review data related to Lombok tourism. The data is then processed through a text preprocessing stage. Next, feature weighting is carried out using the TF-IDF method to capture the significance of each word in the reviews. To address class imbalance, the SMOTE technique is employed to generate synthetic samples for the minority class. Subsequently, feature selection is performed using the Chi-Square method to identify the most relevant features. In the next stage, various classification methods, such as Naive Bayes and SVM, are applied to classify the sentiment of the reviews. Finally, the evaluation of model performance is conducted through accuracy and recall metrics to identify the best model.

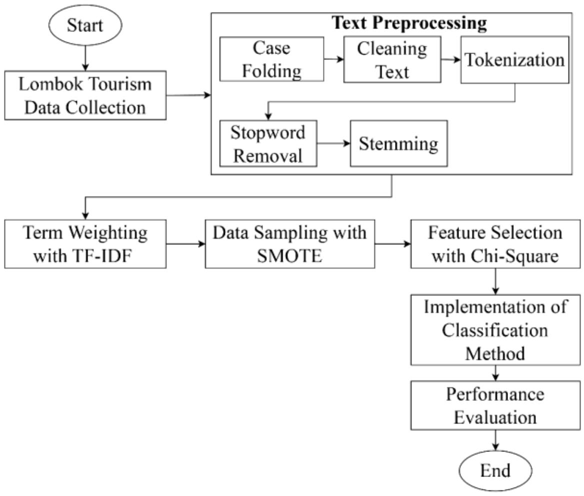

Figure 1. Research Flow

2.1 Text Preprocessing

Text preprocessing is essential for cleaning and preparing textual data for analysis. In this study, techniques such as case folding, text cleaning, tokenization, stopword removal, and stemming are applied to Lombok tourism data, with case folding involving the conversion of all text to lowercase. Text cleaning involves removing unnecessary or irrelevant characters or elements. Common elements removed in this process include: (1) Punctuation marks (such as periods, commas, exclamation points, etc.); (2) Numbers; (3) Special characters (such as #, @, %, etc.); (4) URLs, HTML tags, or other markup elements. Tokenization breaks the text into the smallest units called tokens, such as words or sentences. Stopword removal eliminates common words that frequently appear but typically do not provide significant information for text analysis. Words like "and," "which," and "for" are often considered stopwords. We can focus on more meaningful words in the text by removing stopwords. Stemming is reducing words to their root form by removing suffixes. This study uses the Sastrawi library by applying the Nazief-Adriani method for the stemming process. Meanwhile, the source of the stopword list used comes from Tala's Indonesian stopword list.

2.2 Term Weighting with TF-IDF Method

Term Frequency–Inverse Document Frequency (TF-IDF) is a common technique in information retrieval and text mining used to measure the importance of a term within a document relative to a corpus. It combines TF, which measures term frequency within a document, and IDF, which reflects the term's rarity across the corpus. TF is computed using Equation 1, where $f_{t,d}$ denotes the frequency of term *t* in document *d*, and N_d is the total number of terms in *d*.

$$TF(t,d) = \frac{f_{t,d}}{N_d} \tag{1}$$

IDF aims to assign a higher weight to terms that appear less frequently across the entire corpus. IDF is calculated using Equation 2. After computing TF and IDF, the TF-IDF for a term t in document d is calculated using Equation 3, where $|D|$ is the total number of documents in the corpus, and |d: t ∈ d| represents the number of documents where term t appears at least once.

$$IDF(t,D) = log\left(\frac{|D|}{1+|d:t\in d|}\right) \tag{2}$$

$$TF - IDF(t,d,D) = TF(t,d)xIDF(t,D) \tag{3}$$

2.3. Balancing Data with SMOTE

SMOTE is a method used to address class imbalance in data, particularly in classification problems. In an imbalanced dataset, the minority class has far fewer samples than the majority class. SMOTE generates new synthetic examples for the minority class rather than simply duplicating existing samples, making the model more balanced and better able to recognize patterns from the minority class.

2.4 Feature Selection with Chi-Square

Chi-Square denoted as X^2, is a feature selection method that falls under the filter approach. This method uses statistical analysis to identify unique terms in a document and eliminate less relevant ones, thus reducing computational time. Chi-Square is used for feature selection in documents based on the relationship between two variables: the term t and the class c. Determining which terms contribute to a specific class can be performed using a contingency table, as shown in Table 1. The contingency table helps to determine whether a term in a specific class is independent or dependent. Equation 4 can be derived using the contingency table to calculate the Chi-Square value.

Table 1. Term and Class Contingency Table

	T	Not t
Class	A	C
Not Class	B	D

$$X(t,c) = \frac{N((A \, x \, D) - (B \, x \, C))^2}{(A+B) \, X \, (C+D) \, x \, (A+C) \, x \, (B+D)} \tag{4}$$

The variable *t* denotes the term being evaluated with respect to class *c*. *N* represents the total number of training documents. *A* denotes the number of documents within class *c* that contain the term *t*, while *B* represents the number of documents outside class *c* that also contain *t*. *C* indicates the number of documents within class *c* that do not contain *t*, and *D* refers to the number of documents outside class *c* that likewise do not contain *t*.

2.5 Classification Method

Before the data is classified using the SVM and Naive Bayes methods, it is first divided into training and testing data using a data split of 80% for training and 20% for testing. SVM is a powerful and versatile classification algorithm in machine learning. The main SVM formula aims to maximize the margin $\frac{2}{||w||}$, where w is the weight vector perpendicular to the hyperplane and ||w|| is the length of the vector. The SVM formula is expressed in Equation 5.

$$f(x) = sign(w.x+b) = \begin{cases} +1, & w.x+b \geq 0 \\ -1, & w.x+b < 0 \end{cases} \tag{5}$$

When data is not linearly separable, SVM applies the kernel trick to project the data into a higher-dimensional space, enabling separation by a hyperplane [18]. The kernel function $K(x_i, x_j)$ measures similarity between data points in a higher-dimensional space without explicitly computing their coordinates [19], [20]. Commonly used kernels include linear, radial basis function (RBF), and polynomial kernels [20]. By leveraging these kernels, SVM can address non-linear data and identify the optimal decision boundary within the elevated feature space.

Multinomial Naive Bayes assumes that each feature is independent (the "naive" or "simple" assumption) and that each feature contributes equally to determining the probability of a class. The model calculates the probability of a document $x = (x_1, x_2, \ldots, X_n)$ containing certain features for each class C, then selects the class with the highest probability. This method falls under supervised learning, meaning each data point must be labeled before training. The probability that a document d belongs to class c can be calculated using Equation 6.

$$P(c|d) \propto P(c) \prod_{k=1}^{n} P(t_k|c) \quad (6)$$

Where $P(c|d)$ is the probability of the document d belonging to class c. $P(c)$ represents the prior probability of a document being in class c, The tokens $\{t1, t1, t1, \ldots, tn\}$ are the terms in document d, which form part of the vocabulary consisting of n terms. $P(tk|c)$ is the conditional probability of a term t appearing in a document within class c. $P(c)$ and $P(t_k|c)$ are calculated using maximum likelihood estimation, which is based on the relative frequency of the parameters.

2.6 Performance Evaluation

A confusion matrix is a tool used to evaluate the effectiveness of a classification model. The evaluation of model performance is conducted through accuracy (Equation 7) and recall (Equation 8) metrics to identify the best model.

Table 2. Confusion Matrix

	Negative Prediction	Positive Prediction
Actual Negative	TN	FP
Actual Positive	FN	TP

$$Accuracy = \frac{\mathrm{TN+TP}}{\mathrm{TN+FP+TP+FN}} \quad (7)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{FN+TP}} \quad (8)$$

3. Results and Discussions

This section presents sentiment analysis results on a dataset of tourist reviews about Lombok's tourism, covering preprocessing steps, feature weighting, balancing with SMOTE, and feature selection using the Chi-Square method. The dataset used is sentiment data regarding Lombok tourism obtained from Twitter in 2023, consisting of 940 instances divided into three classes, namely Negative, Neutral, and Positive. Tweets were obtained using several keywords, namely Sembalun, Gili Air, Gili Meno, Gili Trawangan, Sirkuit Mandalika, and Gunung Rinjani. The selection of these keywords aims to capture user opinions and sentiments related to tourist destinations in the Lombok region. The labeling of the dataset was carried out by one Indonesian language expert. The sample dataset on Lombok tourism sentiment is shown in Table 3. Each review contains text expressing opinions about tourism objects in Lombok. Text preprocessing, including case folding, cleaning, tokenization, stopword removal, and stemming, is conducted to prepare the review data for analysis, with the results presented in Table 4.

Table 3. Lombok Tourism Sentiment Dataset

No	Text	Sentiment
1	cuman jalan rinjani itu yangg jalan belak beloknya kalau turunannya itu ada namanya sendiri	Positive
2	Pengen ke Rinjani tapi tidak ada uang https://t.co/JOLLjl7qxL	Neutral
3	Rinjani	Neutral
…	……..	…..
938	@NinaCarla17 Proyek beda kelas Sirkuit Mandalika dengan sirkuit Formula E Yang teriak golongan kadrun doang	Neutral
939	RT @tempodotco: WSBK dan MotoGP di Mandalika Merugi, Erick Thohir: Event yang Memberatkan, Negoisasi Ulang https://t.co/ItaYShT0Y3	Negative
940	bukit merese indah banget	Positive

Table 4. Text Preprocessing Result

No	Technique	Text
1	Data Original	cuman jalan rinjani itu yangg jalan belak beloknya kalau turunannya itu ada namanya sendiri
2	Case Folding	cuman jalan rinjani itu yangg jalan belak beloknya kalau turunannya itu ada namanya sendiri
3	Clean Text	cuman jalan rinjani yangg jalan belak beloknya kalau turunannya namanya sendiri
4	Tokenization	['cuman', 'jalan', 'rinjani', 'yangg', 'jalan', 'belak', 'beloknya', 'kalau', 'turunannya', 'namanya', 'sendiri']
5	Stopword	['cuman', 'jalan', 'rinjani', 'jalan', 'belak', 'beloknya', 'turunannya', 'namanya']
6	Stemming	['cuman', 'jalan', 'rinjani', 'jalan', 'belak', 'belok', 'turun', 'nama']

The preprocessed text is then assigned term weights using TF-IDF. The TF-IDF calculations highlight the importance of certain words in the reviews. For instance, the word "*jalan*" has the highest TF-IDF value, indicating its strong relevance in expressing positive sentiments about Lombok tourism. TF-IDF values provide insights into the influence of reviews in determining sentiment. The results of term weighting using TF-IDF are shown in Table 5.

Table 5. Result of Term Weighting with TF-IDF

Doc ID	Term	TF	DF	IDF	TF-IDF
1	jalan	2	27	1.54	3.08
1	belok	1	1	2.97	2.97
1	turun	1	4	2.37	2.37
1	nama	1	5	2.27	2.27
1	rinjani	1	462	0.31	0.31
1	cuman	1	4	2.37	2.37
1	belak	1	1	2.97	2.37
..	…	…	…	…	…
940	banget	1	48	1.29	1.29
940	merese	1	2	2.67	2.67
940	bukit	1	13	1.86	1.86

Next, data balancing is performed using SMOTE. The SMOTE method is necessary to balance the previously imbalanced Lombok tourism review data, which consists of 91 instances for the Negative class, 331 instances for the Neutral class, and 518 instances for the Positive class. After balancing with SMOTE, each class contains 518 instances. The results before and after applying SMOTE are illustrated in Figure 2.

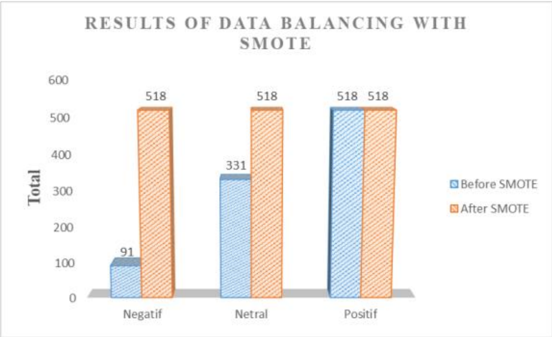

Figure 2. Lombok Tourism Review Data Before and After SMOTE

Next, feature selection uses the Chi-Square method to identify the most significant features in differentiating review sentiments. Table 6 shows the results of the Chi-Square feature selection, where words with the highest Chi-Square values, such as "*indah*" (beautiful), "*dingin*" (cold), and "*buruk*" (bad), were chosen for their significant contribution to sentiment classification. Using these more relevant features helps reduce data dimensions and improve model performance. Words such as "*indah*," "*dingin*," "*buruk*," and "*rinjani*" (mountain) have high Chi-Square values, indicating their strong relevance in sentiment analysis. By reducing the number of irrelevant features, this selection process enhances the model's accuracy in predicting sentiment.

Table 6. Selected Term Result of Chi Square Feature Selection

No	Term	Chi Square Values
1	indah	44.28
2	bagus	43.58
3	Banget	39.28
4	Sukses	20.40
5	Nenek	17.92
..	….	…..
746	Provinsi	0.57
747	Luas	0.57
748	Lombok	0.56
749	Subang	0.56
750	Dusun	0.56

Figure 3 presents the confusion matrix from the results of applying the Naïve Bayes method to the original data without applying optimization techniques. These results indicate that the model has very limited performance in classifying minority classes, especially the Negative class, where none of the 23 instances were predicted correctly (true negative = 0), reflecting a recall value of 0% for that class. In contrast, the Neutral class showed high classification performance, with 95 out of 99 instances being correctly classified, indicating a model bias towards the majority class. For the Positive class, out of a total of 66 instances, only 31 instances were correctly classified, while the remaining 35 were incorrectly predicted as the Neutral class.

Figure 4 presents the confusion matrix resulting from applying the SVM method to the original data without applying optimization techniques. In general, the model shows improved performance compared to the Naïve Bayes method, especially in classifying the Positive class, where 47 out of 66 instances were correctly identified, while the remaining 19 instances were incorrectly classified as Neutral. For the Neutral class, the model recorded quite high accuracy with 89 out of 99 instances correctly classified, although there were still 10 instances incorrectly predicted as Positive. However, performance against the Negative class remains low, with only 3 out of 23 instances being correctly recognized, indicating that this class is still not effectively mapped by the model. This imbalance in the prediction distribution confirms that, although SVM has superior classification capabilities compared to Naïve Bayes, the model still requires the support of data balancing and feature selection strategies to improve performance in identifying minority classes.

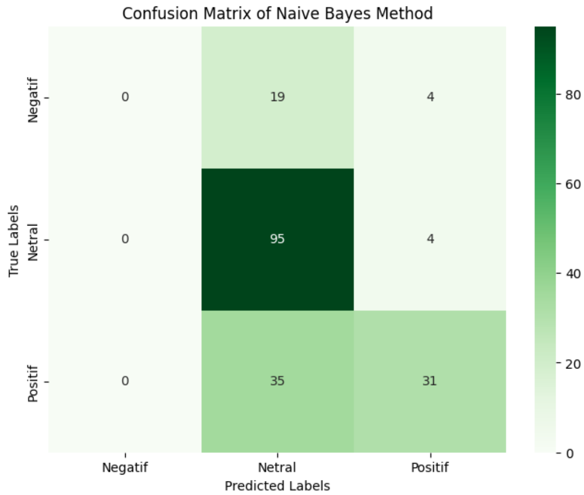

Figure 3. Result Naïve Bayes with Data Original

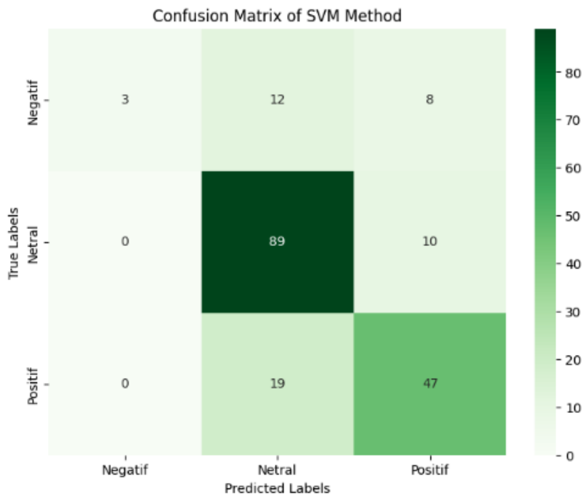

Figure 4. Result SVM with Data Original

Figure 5 shows the overall improvement in classification performance after applying the SMOTE technique to the Naïve Bayes method. With the application of SMOTE, the number of correct predictions for the Negative class increased significantly to 103 out of 105 instances, indicating that the addition of synthetic data helped the model in learning patterns in previously underrepresented classes. In the Positive class, the model also shows significant performance improvement with 79 correct predictions out of 108 instances. Meanwhile, the Neutral class maintains relatively high accuracy with 66 correct predictions out of 98 instances, although it experiences a slight decrease due to the redistribution of the model's attention to other classes.

Figure 6 shows the classification results after applying the SMOTE technique to the SVM algorithm. The model shows excellent performance in classifying the Negative class, with 103 correct predictions out of a total of 105 instances. In the Neutral class, 85 out of 98 instances were successfully classified correctly, although some misclassifications were directed towards the Positive class. As for the Positive class, the model produced 78 correct predictions out of 108 instances, but there were still 30 instances that were misclassified as the Neutral class. Overall, the application of SMOTE has proven effective in improving the model's ability to recognize minority classes and balancing classification performance between classes.

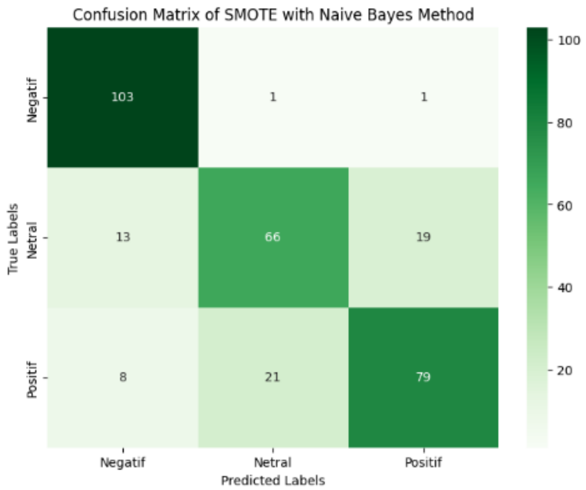

Figure 5. Result Naïve Bayes with SMOTE

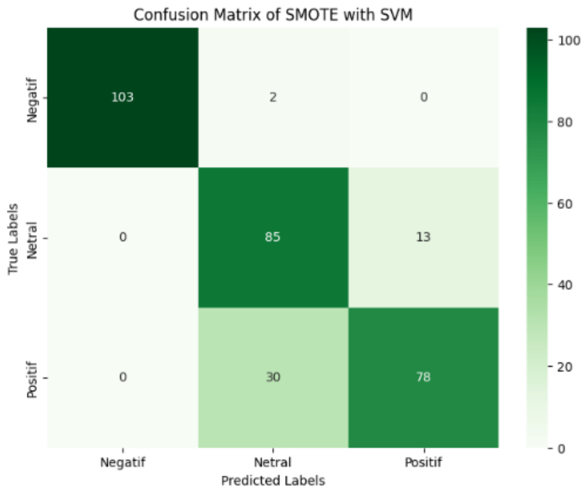

Figure 6. Result SVM with SMOTE

Figure 7 shows the classification results using the Naïve Bayes algorithm after applying the SMOTE technique and feature selection with the Chi-Square method. The model shows excellent performance in the Negative class, with 102 instances successfully classified correctly out of a total of 105 instances. In the Neutral class, 73 out of 98 instances were successfully classified correctly, while 19 instances were classified as Positive and 6 instances as Negative. Meanwhile, in the Positive class, the model was able to classify 85 out of 108 instances accurately, with most of the misclassifications leading to the Neutral class. These findings indicate that the combination of the SMOTE and Chi-Square methods is able to improve the model's capability in recognizing patterns in the minority class and improve the overall distribution of classification results.

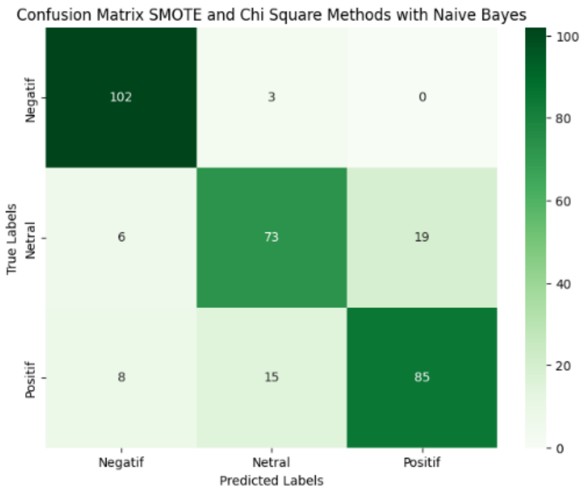

Figure 7. Result Naïve Bayes with SMOTE and Chi Square

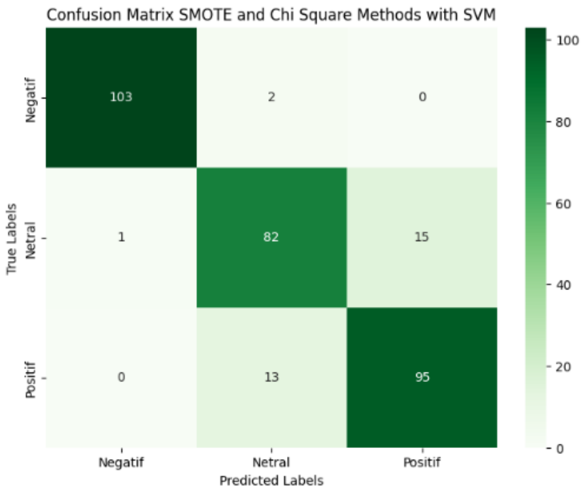

Figure 8. Result SVM with SMOTE and Chi Square

Figure 8 shows the classification results using the SVM algorithm after applying the SMOTE technique and feature selection with the Chi-Square method. The model shows very good performance in the Negative class, with 103 out of 105 instances successfully classified correctly. In the Neutral class, 82 out of 98 instances were successfully classified correctly, while the rest were mostly classified as the Positive class.

Meanwhile, in the Positive class, the model successfully classified 95 out of 108 instances accurately, with 13 instances incorrectly classified as Neutral. Overall, the combination of the SMOTE and Chi-Square methods is able to improve the model's ability to recognize the three classes more balanced, especially in strengthening the classification performance in the minority class.

Figure 9 shows the accuracy comparison between the Naïve Bayes and SVM methods with three types of data: original data, data after applying SMOTE, and data with both SMOTE and Chi-Square. The Naïve Bayes method achieved an accuracy of 67.02% with the original data, increased to 70.74% with SMOTE, and reached 84% with SMOTE and Chi-Square. The SVM method achieved an accuracy of 73.93% with the original data, increased to 85.53% with SMOTE, and reached 90% with SMOTE and Chi-Square.

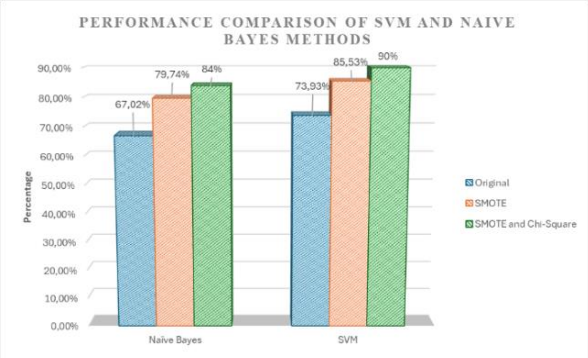

Figure 9. Performance Comparison of SVM and Naïve Bayes Methods

Based on Figure 9, it is evident that the application of SMOTE and Chi-Square feature selection significantly improves accuracy in both the Naïve Bayes and SVM methods. The Naïve Bayes method using SMOTE saw an accuracy increase of 12.72% compared to the original data. Furthermore, applying both SMOTE and Chi-Square improved the accuracy by 16.98% compared to the original data and 4.26% compared to using SMOTE alone. The SVM method using SMOTE experienced an accuracy increase of 11.6% compared to the original data. Additionally, applying SMOTE and Chi-Square boosted accuracy by 16.07% compared to the original data and 4.47% compared to using SMOTE alone.

Using SMOTE as a data balancing technique has proven to improve the accuracy of various classification methods, including Naïve Bayes and SVM, as demonstrated in previous studies [21], [22]. SMOTE helps reduce class imbalance by generating additional samples for the minority class, which ultimately enhances the performance of the classification model. Moreover, combining SMOTE and Chi-Square feature selection further strengthens model accuracy [23]. By eliminating irrelevant features through Chi-Square, the model can focus on the important features that positively impact classification, thus optimizing the model's performance compared to using SMOTE alone [24], [25] Overall, these findings suggest that combining data balancing using SMOTE and Chi-Square feature selection is an effective strategy for improving classification model performance.

4. Conclusions

The application of SMOTE and Chi-Square feature selection effectively improves the accuracy of the Naïve Bayes and SVM models in sentiment classification for Lombok tourism. The combination of SMOTE and Chi-Square results in higher accuracy compared to using SMOTE alone, with an accuracy increase of 16.98% for Naïve Bayes and 16.07% for SVM. SMOTE helps balance the classes, while Chi-Square enables the model to focus on relevant features, thereby optimizing classification performance. This study is limited to Lombok tourism reviews with traditional machine learning methods. Further research is recommended to explore deep learning models such as BERT, given its ability to capture meaning and relationships between words more deeply than traditional approaches. BERT, with its bidirectional mechanism, allows for a more accurate understanding of the context of words in the entire sentence, making it effective in overcoming ambiguity and double meanings in text. In addition, the expansion of the Lombok tourism sentiment dataset is also recommended, by including other data sources such as TripAdvisor and various social media other than Twitter. This aims to improve the generalization capability of the model and strengthen the validity of the sentiment analysis results in various tourist destination contexts in Lombok. Not only that, further research can involve more than one annotator to increase the reliability of labeling on tourism tweets.

Acknowledgements

The author declares no conflicts of interest.

References

[1] M. Azizurrohman, R. B. Hartarto, Y.-M. Lin, and F. H. Nahar, “The Role of Foreign Tourists in Economic Growth: Evidence from Indonesia,” *J. Ekon. Stud. Pembang.*, vol. 22, no. 2, pp. 313–322, 2021, doi: 10.18196/jesp.v22i2.11591.

[2] I. Rahadi, H. Basri, M. Adi Junaidi, and D. Alfatwari, “The most popular tourist destinations in Lombok Timur by univariate and bivariate analysis methods,” *J. Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022, doi: https://doi.org/10.35335/mantik.v6i3.2910.

[3] R. Jangra, S. P. Kaushik, and S. S. Saini, “An analysis of tourist’s perceptions toward tourism development: Study of cold desert destination, India,” *Geogr. Sustain.*, vol. 2, no. 1, pp. 48–58, 2021, doi: 10.1016/j.geosus.2021.02.004.

[4] D. Flores-Ruiz, A. Elizondo-Salto, and M. D. L. O. Barroso-González, “Using social media in tourist sentiment analysis: A case study of andalusia during the Covid-19 pandemic,” *Sustain.*, vol. 13, no. 7, pp. 1–19, 2021, doi: 10.3390/su13073836.

[5] T. Oliveira, B. Araujo, and C. Tam, “Why do people share their travel experiences on social media?,” *Tour. Manag.*, vol. 78, no. November 2019, pp. 1–14, 2020, doi: 10.1016/j.tourman.2019.104041.

[6] O. A. George and C. M. Q. Ramos, “Sentiment analysis applied to tourism: exploring tourist-generated content in

the case of a wellness tourism destination," *Int. J. Spa Wellness*, vol. 7, no. 2, pp. 139–161, 2024, doi: 10.1080/24721735.2024.2352979.

[7] M. Alzate, M. Arce-Urriza, and J. Cebollada, "Mining the text of online consumer reviews to analyze brand image and brand positioning," *J. Retail. Consum. Serv.*, vol. 67, no. November 2021, p. 102989, 2022, doi: 10.1016/j.jretconser.2022.102989.

[8] K. Y. Chan, C. K. Kwong, and H. Jiang, "Analyzing imbalanced online consumer review data in product design using geometric semantic genetic programming," *Eng. Appl. Artif. Intell.*, vol. 105, no. August, pp. 1–15, 2021, doi: 10.1016/j.engappai.2021.104442.

[9] S. M. Alrashidi and A. M. Awadelkarim, "Machine Learning-Based Sentiment Analysis for Tweets Saudi Tourism," *Journla Theor. Appl. Inf. Technol.*, vol. 100, no. 16, pp. 5096–5109, 2022.

[10] N. Leelawat *et al.*, "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," *Heliyon*, vol. 8, no. 10, pp. 1–11, 2022, doi: 10.1016/j.heliyon.2022.e10894.

[11] B. A. Alharbi, M. A. Mezher, and A. M. Barakeh, "Tourist Reviews Sentiment Classification using Deep Learning Techniques: A Case Study in Saudi Arabia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 717–726, 2022, doi: 10.14569/IJACSA.2022.0130685.

[12] C. Steven and W. Wella, "The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter," *IJNMT (International J. New Media Technol.*, vol. 7, no. 2, pp. 102–110, 2020, doi: 10.31937/ijnmt.v7i2.1732.

[13] A. D. Poernomo and S. Suharjito, "Indonesian online travel agent sentiment analysis using machine learning methods," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 1, p. 113, 2019, doi: 10.11591/ijeecs.v14.i1.pp113-117.

[14] A. Muhammad, S. Defit, and G. W. Nurcahyo, "Determining Intent: Sentiment Analysis Based on the Classification of Indonesian Tourist Destination Review Texts," *J. Adv. Inf. Technol.*, vol. 15, no. 10, pp. 1106–1116, 2024, doi: 10.12720/jait.15.10.1106-1116.

[15] Hariyono *et al.*, "Exploring Visitor Sentiments: A Study of Nusantara Temple Reviews on TripAdvisor Using Machine Learning," *J. Appl. Data Sci.*, vol. 5, no. 2, pp. 600–612, 2024, doi: 10.47738/jads.v5i2.208.

[16] Y. A. Singgalen, "Culture and heritage tourism sentiment classification through cross-industry standard process for data mining," *Int. J. Basic Appl. Sci.*, vol. 12, no. 3, pp. 110–120, 2023, doi: https://doi.org/10.35335/ijobas.v12i3.299.

[17] N. K. Shadrina, E. Sutoyo, and V. P. Widartha, "Sentiment Analysis in Reviews About Beaches in Bali on Tripadvisor Using Recurrent Neural Network (RNN)," in *2021 IEEE 7th Information Technology International Seminar (ITIS)*, 2021, pp. 1–6. doi: 10.1109/ITIS53497.2021.9791501.

[18] D. T. Hermanto, M. Ziaurrahman, M. A. Bianto, and A. Setyanto, "Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier," in *Journal of Physics*, 2018, pp. 1–8. doi: 10.1088/1742-6596/1140/1/012037.

[19] H. Hairani, T. Widiyaningtyas, and D. D. Prasetya, "Addressing Class Imbalance of Health Data: A Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 3, pp. 1310–1318, 2024, doi: https://dx.doi.org/10.62527/joiv.8.3.2283.

[20] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol.*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.

[21] X. Gao, N. Jamil, M. I. Ramli, and S. M. Z. S. Z. Ariffin, "A Comparative Analysis of Combination of CNN-Based Models with Ensemble Learning on Imbalanced Data," *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 456–464, 2024, doi: 10.62527/joiv.8.1.2194.

[22] H. Hairani and D. Priyanto, "A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 585–590, 2023, doi: https://dx.doi.org/10.14569/IJACSA.2023.0140864.

[23] M. W. Huang, C. H. Chiu, C. F. Tsai, and W. C. Lin, "On combining feature selection and over-sampling techniques for breast cancer prediction," *Appl. Sci.*, vol. 11, no. 14, pp. 1–9, 2021, doi: 10.3390/app11146574.

[24] L. G. R. Putra, K. Marzuki, and H. Hairani, "Correlation-based feature selection and Smote-Tomek Link to improve the performance of machine learning methods on cancer disease prediction," *Eng. Appl. Sci. Res.*, vol. 50, no. 6, pp. 577–583, 2023, doi: 10.14456/easr.2023.59.

[25] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," *Int. J. Inf. Educ. Technol.*, vol. 13, no. 2, pp. 289–295, 2023, doi: 10.18178/ijiet.2023.13.2.1806.