International Journal of Computer Sciences and Mathematics Engineering Vol. 4 No. 1 (2025)
Published by: CV. HAWARI

HWJ Hawari

International Journal of Computer Sciences and Mathematics Engineering
Journal homepage: www.ijecom.org

Analysis of SVM Algorithm And SMOTE Technique on Public Sentiment With Google Maps Reviews Towards The Quality of Education at LP3I Across Indonesia

Maulian Saputra[1], Muhammad Iqbal[2], Khairul[3]
[1,2,3] Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
Email : maulian.saputra@gmail.com, muhammadiqbal@dosen.pancabudi.ac.id, khairul@dosen.pancabudi.ac.id

Article Info

Article history:

Received Mar 15, 2025
Revised Mar 18, 2025
Accepted Mar 27, 2025

Keywords:

Sentiment Analysis
Support Vector Machine
SMOTE
Google Maps Reviews
LP3I Education

ABSTRACT

This study analyzes public sentiment toward LP3I's education quality in Indonesia using the Support Vector Machine (SVM) algorithm and the Synthetic Minority Over-sampling Technique (SMOTE). Reviews from Google Maps were collected, preprocessed, translated, and labeled using the Natural Language Toolkit (NLTK). The classification model was tested with and without SMOTE to address class imbalance. The results show that applying SMOTE improves the model's ability to detect neutral and negative sentiments, increasing accuracy from 86.93% to 88.44% and significantly enhancing recall for minority classes. However, while SMOTE helps create a more balanced classification, it also introduces a potential risk of overfitting due to synthetic data generation. Overall, the integration of SVM and SMOTE improves sentiment classification performance, providing valuable insights for LP3I to enhance its academic and administrative services based on public perception.

Corresponding Author:

Maulian Saputra,
Magister Teknologi Informasi,
Universitas Pembangunan Panca Budi,
Jend. Gatot Subroto Road, 4.5 KM Seikambing, 20122, Medan City, Sumatera Utara, Indonesia.
Email: maulian.saputra@gmail.com

1. INTRODUCTION

Education plays a crucial role in the development of high-quality and competitive human resources in the era of globalization. Along with the rapid advancement of technology and digitalization, society now has easier access to evaluate and provide feedback on educational services through various digital platforms. One of the most widely used platforms is Google Maps [1]. Users provide various reviews on Google Maps, which can be viewed in the review section [2].

LP3I, as one of the vocational education institutions focused on developing job skills, has gained public attention in various regions of Indonesia. However, public perception of LP3I is not always homogeneous. Reviews provided by the public on Google Maps exhibit a diversity of sentiments, ranging from positive and neutral to negative reviews. Sentiment analysis is needed to

E-ISSN 2962-4274

understand and classify public opinions, emotions, or perspectives on a particular topic [3]. By integrating sentiment analysis into promotional strategies, higher education institutions can better understand public reactions for a more effective approach [4].

This study aims to analyze public sentiment toward the quality of education at LP3I across Indonesia based on reviews on Google Maps. The algorithm used, Support Vector Machine (SVM), is an effective text processing method capable of achieving high accuracy in sentiment analysis [5]. It is also a machine learning method used for classification, aiming to find an optimal hyperplane that can separate two different classes in the input space [6]

The SMOTE technique is used to enhance the performance of classification models on imbalanced data [7]. Class imbalance can lead to poor classification results and may not always produce optimal outcomes [8].

In a previous study by [9], The SVM algorithm, when using the SMOTE technique, can improve accuracy by 4.27%. Another study by [10], Testing with a comparison of four data splits showed that the 75:25 ratio achieved the highest accuracy of 85.20% for classification using SVM without SMOTE. Meanwhile, classification using SMOTE-based SVM at a 70:30 ratio resulted in the highest average accuracy of 91%.

2. RESEARCH METHOD

This study is a quantitative research with a secondary data analysis approach. The data used in this study consists of text reviews from users, collected from the Google Maps platform.

The research flow can be seen in Figure 1 below.

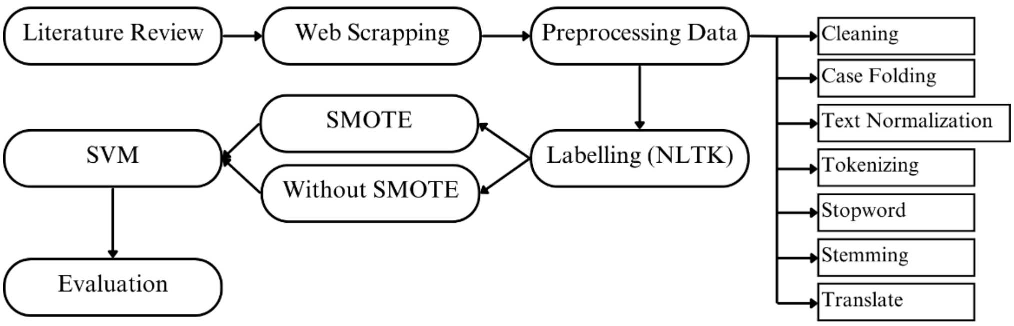

Figure 1. Research flowchart

Based on Figure 1 above, the research procedure can be explained as follows.

2.1. Literature Review

At the literature review stage, this study begins by exploring the fundamental concepts of sentiment analysis [11]. This is the process of identifying and categorizing public opinions based on text into positive, neutral, or negative sentiments.

2.2. Web Scrapping

This study utilizes web scraping techniques to collect public reviews about LP3I from the Google Maps platform. This process is carried out using the Instant Data Scraper extension in Google Chrome, which enables the automated and structured retrieval of review data [12].

2.3. Preprocessing Data

The data preprocessing stage is a crucial step in preparing raw data obtained from web scraping for sentiment analysis [13]. The steps involved are as follows:

2.3.1. Cleaning

It aims to clean the text from elements that are irrelevant to the analysis. In this stage, special characters such as punctuation marks, numbers, symbols, URLs, and emojis are removed from the text.

2.3.2. Case Folding

Converting all letters in the text to lowercase. This is done to avoid differences in meaning due to capitalization, such as "Bagus" and "bagus," which actually have the same meaning.

2.3.3. Text Normalization

It aims to replace non-standard words, slang, or informal terms with standard words according to the Indonesian dictionary. For example, words like "keren banget" can be normalized to "sangat bagus".

2.3.4. Tokenizing

Splitting sentences into individual words. For example, the sentence "LP3I sangat bagus untuk belajar" will be tokenized into ["LP3I", "sangat", "bagus", "untuk", "belajar"].

2.3.5. Stopword

At this stage, common words that frequently appear but do not significantly contribute to sentiment analysis, such as "dan," "yang," "di," or "untuk," are removed from the text [14].

2.3.6. Stemming

The process of converting words into their base form. For example, the word "membelajar" will be changed to "belajar," and "pengajaran" will become "ajar." This process helps reduce word variations that share the same root, making the analysis more efficient and focused.

2.3.7. Translation

At this stage, all text in Indonesian is translated into English. This process is carried out to facilitate data labeling using the NLTK (Natural Language Toolkit) library, which has broader support for text analysis in English.

2.4. Labelling

After the preprocessing stage is completed, the next step is data labeling. In this stage, each review that has been translated into English is assigned a label based on its sentiment: positive, neutral, or negative. This process utilizes the NLTK (Natural Language Toolkit) library, which provides various text analysis features, including labeling functions.

Labeling is performed manually or semi-automatically by leveraging existing algorithms in NLTK, such as sentiment lexicon-based analysis or word pattern analysis. For example, a review like *"LP3I is very good for learning"* will be labeled as positive, while a review like *"LP3I has many issues with the staff"* will be labeled as negative.

2.5. With SMOTE and Without SMOTE

At the SMOTE (Synthetic Minority Over-sampling Technique) stage, this technique is applied to address the issue of class imbalance in the dataset. Class imbalance often occurs when the number of data points in one class (e.g., positive sentiment) is significantly higher than in other classes (e.g., negative or neutral sentiment), which can lead to a machine learning model being biased toward the majority class [15].

Meanwhile, in the non-SMOTE stage, the data used for model training does not undergo the oversampling process to address class imbalance. In this case, the model is trained using a dataset with an imbalanced class distribution, where the majority class has significantly more examples than the minority class [16].

2.6. SVM

Support Vector Machine (SVM) is applied as the primary algorithm for sentiment classification of reviews. SVM is a highly effective machine learning algorithm for handling classification problems, especially in high-dimensional data such as text [17].

SVM works by finding the best hyperplane (decision boundary) that separates data from different classes (positive, neutral, and negative) with the largest margin. In other words, SVM aims to determine the optimal boundary that maximizes the distance between each class.

2.7. Evaluation

This evaluation aims to measure the model's performance in classifying unseen test data. Several evaluation metrics used in this stage include the confusion matrix, accuracy, and classification report. The confusion matrix indicates how well the model classifies data by showing the number of correct and incorrect predictions for each class [18]. Accuracy measures how accurately the model predicts the correct class across the entire test dataset.The classification report provides more detailed information on precision, recall, and F1-score for each class, helping to evaluate the model more comprehensively.

Model evaluation is crucial to determine whether the model is sufficiently effective in classifying sentiment reviews and to decide whether improvements are needed. Further refinements can include hyperparameter tuning or applying additional techniques to handle class imbalance.

3. RESULTS AND DISCUSSIONS

This study aims to analyze public sentiment toward the quality of LP3I education across Indonesia based on Google Maps reviews. The data used in this research consists of public reviews about LP3I across Indonesia, collected from the Google Maps platform in December 2024.

3.1. Web Scraping Results

The data collection process was carried out by accessing individual Google Maps pages of various LP3I branches across Indonesia. The branches scraped include two types: College and Polytechnic. Instant Data Scraper was used to extract available reviews from each LP3I branch, including review text and related information.

The following is an example of the scraped data using Instant Data Scraper from the Google Maps pages of LP3I branches.

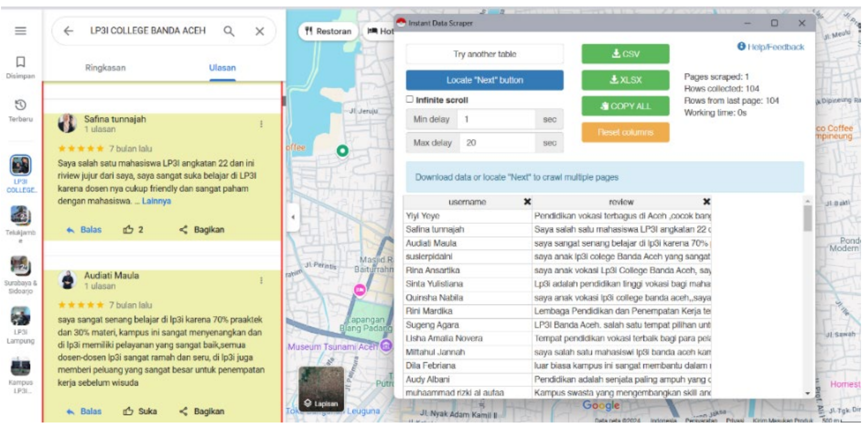

Figure 2. Web Scraping Using Instant Web Scraper

E-ISSN 2962-4274

Figure 2 above illustrates the process of extracting review data from Google Maps, where the scraped data includes only the username and review text. From the data collection process, a total of 2,261 reviews were successfully gathered from LP3I branches across Indonesia.

After collecting all the data from each branch, the next step is merging the data into a single main dataset. At this stage, all data from LP3I College and Polytechnic branches across Indonesia are consolidated into a single CSV file to facilitate further analysis.

Next, data completeness is checked, duplicates are detected, and irrelevant or spam reviews are removed. All processes are carried out using the Python programming language and executed in Google Colab [19], [20].

```python
import pandas as pd

df = pd.read_csv('dataset-LP3I.csv')
df_clean = df.dropna()

df_clean.to_csv('hasil-dataset.csv', index=False)
print("Data telah disimpan ke file hasil-dataset.csv")
```

Figure 3. The process of removing NaN (Not a Number)

Figure 3 above illustrates the results of the dataset cleaning process, which is then saved in CSV format for use in the preprocessing stage.

3.2. Preprocessing Data Results

The preprocessing process is carried out to improve the quality of the data before sentiment analysis. The steps taken aim to ensure that the data is clean, structured, and ready to be analyzed with machine learning algorithms. The preprocessing stages include:

3.2.1. Cleaning

At this stage, special characters, numbers, punctuation marks, and unnecessary symbols are removed from the review text. This process helps eliminate noise in the data that could interfere with the analysis. An example of the cleaning process results can be seen in Table 1 below.

Table 1. Cleaning Results

Reviews	Cleaning
LP3I kekinian banget yaaa, mancap!!! Bangga ja...	LP3I kekinian banget yaaa mancap Bangga jadi l...
Semoga LP3I Balikpapan sukses dan maju ya ges 😊 ...	Semoga LP3I Balikpapan sukses dan maju ya ges
Awesome Place's For Finding A Good Skill and E...	Awesome Places For Finding A Good Skill and Ex..

3.2.2. Case Folding

At this stage, the text is converted to lowercase to ensure consistency in the analysis, preventing differences in word interpretation due to uppercase or lowercase usage. An example of the case folding process results can be seen in Table 2 below.

Table 2. Case Folding Results

Cleaning	Case Folding
LP3I kekinian banget yaaa mancap Bangga jadi l...	lp3i kekinian banget yaaa mancap bangga jadi l...
Semoga LP3I Balikpapan sukses dan maju ya ges	semoga lp3i balikpapan sukses dan maju ya ges
Awesome Places For Finding A Good Skill and Ex...	awesome places for finding a good skill and ex...

3.2.3. Text Normalization

The normalization process is carried out to replace non-standard words or abbreviations with standard words. An example of the word normalization results can be seen in Table 3 below.

E-ISSN 2962-4274

Table 3. Text Normalization Results

Case Folding	Text Normalization
tempat belajar yg...	tempat belajar yang..
tempat yg nyaman untuk kuliah..	tempat yang nyaman untuk kuliah..
politeknik terbaik dan biaya terjangkau boss...	politeknik terbaik dan biaya terjangkau bos...

3.2.4. Tokenizing

At this stage, each review is split into individual words to facilitate word-based analysis. An example of the tokenizing process results can be seen in Table 4 below.

Table 4. Tokenizing Results

Text Normalization	Tokenizing
saya bangga kuliah di lp3i	[saya, bangga, kuliah, di, lp3i]
kampus penempatan kerja	[kampus, penempatan, kerja]
very helpful staff	[very, helpful, staff]

3.2.5. Stopword

At this stage, common words that do not have significant meaning in the analysis, such as "dan" (and), "yang" (which), "di" (in), and other frequent words, are removed to allow the model to focus on more relevant words. An example of the stopword removal process results can be seen in Table 5 below.

Table 5. Stopword Results

Tokenizing	Stopword
[pendidikan, vokasi, terbagus, di, aceh, cocok...	[pendidikan, vokasi, terbagus, aceh, cocok...
[saya, sangat, senang, belajar, di, lp3i]	[senang, belajar, lp3i]
[saya, anak, lp3i, colege, banda, aceh, yang, ..	[anak, lp3i, colege, banda, aceh, terbantu, ka...

3.2.6. Stemming

At this stage, the process transforms words into their root form or base word. An example of the stemming process results can be seen in Table 6 below.

Table 6. Stemming Results

Stopword	Stemming
[pendidikan, vokasi, terbagus, aceh, cocok...	didik vokasi bagus aceh cocok...
[senang, belajar, lp3i]	senang ajar lp3i
[anak, lp3i, college, banda, aceh, terbantu, ka...	anak lp3i college banda aceh bantu

3.2.7. Translation

After the stemming process is completed, the review data is exported into a CSV file. The translation process is carried out using Python with the Translator library in Google Colab, as illustrated in Figure 4 below.

```python
file_path = 'Hasil_Preprocessing_Data_Stemming.csv'
df = pd.read_csv(file_path)
translator = Translator()

def translate_text(text):
    try:
        # Penerjemahan hanya jika teks tidak kosong
        if pd.notnull(text):
            translated = translator.translate(text, src='id', dest='en')
            return translated.text
        return text
    except Exception as e:
        print(f"Error translating: {text}. Error: {e}")
        return text

text_column = 'steming_data'
df['translated_text'] = df[text_column].apply(translate_text)

output_file = 'Hasil_Preprocessing_Data_English.csv'
df.to_csv(output_file, index=False)
```

Figure 4. Python code for translating Indonesian to English

After the preprocessing data is completed, data visualization is performed to understand the patterns of the most frequently occurring words in the reviews. This visualization includes a word cloud and the frequency of the most dominant words in the dataset, as shown in Figure 5 and Figure 6 below.

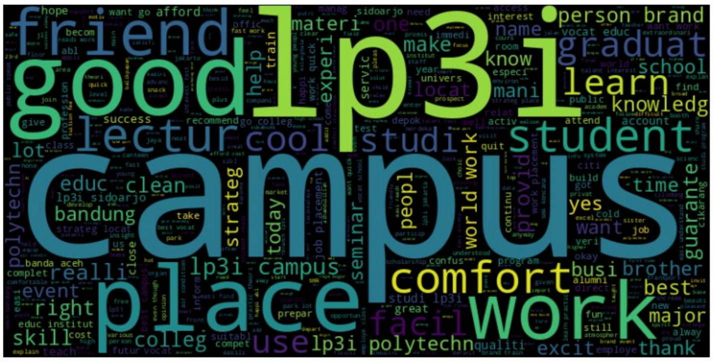

Figure 5. Word Cloud

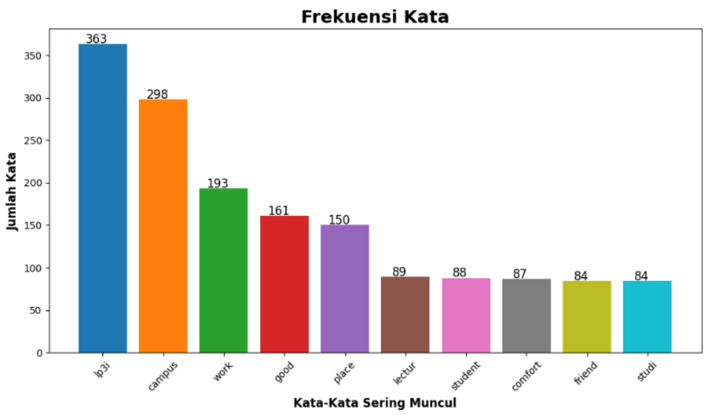

Figure 6. Word Frequency

3.3. Labeling Results

NLTK (Natural Language Toolkit) library is used to label reviews based on the analysis of words contained in the text. This process follows a lexicon-based approach, where each word in the text is assigned a sentiment score—positive, negative, or neutral—based on NLTK's built-in sentiment lexicon. Table 7 presents an example of sentiment labeling.

Table 7. Sentiment Labeling Results Using NLTK

Reviews	Sentiment Score	Sentiment
LP3I is now very proud to pass LP3I skill, rea...	0.8977	Positive
LP3I Balikpapan Location of the Ampal River Di..	0.0000	Neutral
LP3I Didik Student Vocational Fast Work Wrong ...	-4.767	Negative
student vocational good teaching for fast qual...	0.4404	Positive
My Campus	0.0000	Neutral

From Table 7 above, it can be seen that each review is assigned a sentiment score based on the analysis of the words contained within it. Reviews with a positive score (> 0) are categorized as positive sentiment, a neutral score (= 0) as neutral sentiment, and a negative score (< 0) as negative sentiment.

The labeling results are visualized in the form of a bar chart to provide a clearer overview of the sentiment distribution in the reviews, as shown in Figure 7 below.

E-ISSN 2962-4274

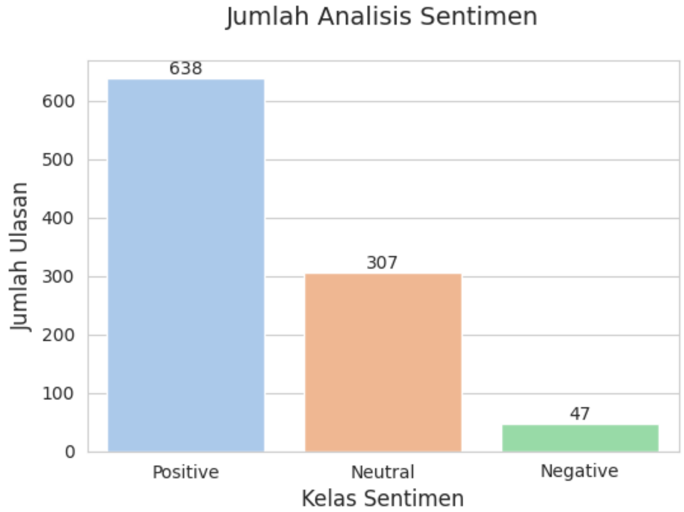

Figure 7. Sentiment Distribution Graph

3.4. Classification Results of SVM and SMOTE

After preprocessing, the review data is converted into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The dataset is then split into 80% training data and 20% testing data to ensure that the model has sufficient data for learning while maintaining a separate dataset for evaluation.

Before applying SMOTE, the data exhibits class imbalance, with a dominance of positive sentiment. After applying SMOTE, the number of data points in the negative and neutral sentiment categories increases to balance with the positive category. Figure 8 below illustrates the sentiment distribution before and after applying SMOTE.

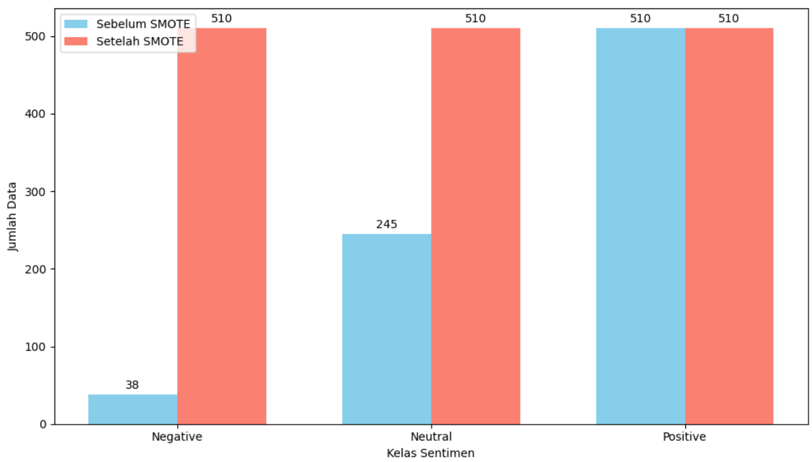

Figure 8. Sentiment Class Distribution Before and After SMOTE

To evaluate the model's performance, a confusion matrix is used to show the number of correct and incorrect predictions for each sentiment class. The model is tested in two scenarios: without SMOTE and with SMOTE. The confusion matrix for the model without SMOTE can be seen in Figure 9 below.

E-ISSN 2962-4274

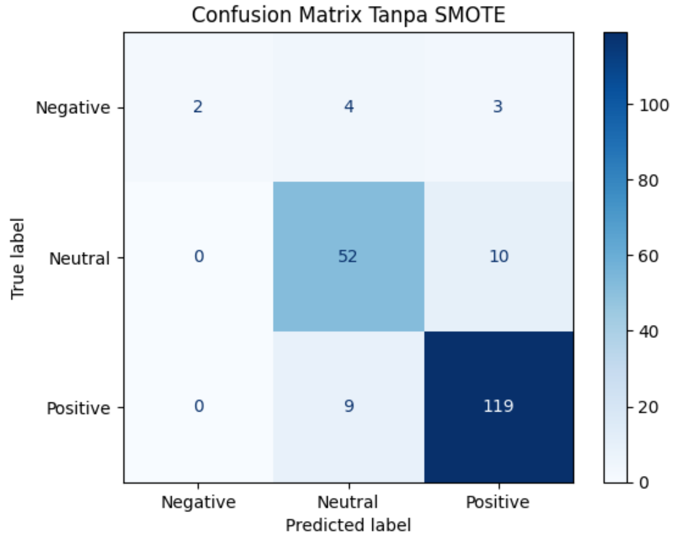

Figure 9. Confusion Matrix Without SMOTE

And the confusion matrix after applying SMOTE can be seen in Figure 10 below.

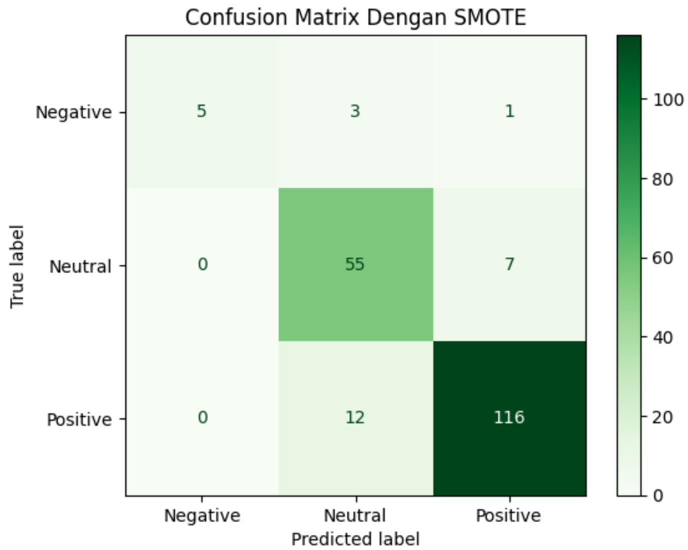

Figure 10. Confusion Matrix With SMOTE

From the results above, the SVM algorithm with SMOTE demonstrates an improvement in detecting neutral and negative classes, which previously suffered from data imbalance.

3.5. Classification Report

The classification report provides a more detailed overview of the model's performance in classifying each class. Tables 8 and 9 below present the classification report results for each model.

Table 8. Clasification Report SVM Without SMOTE

Accuracy of SVM Without Smote : 86,93%				
	Precision	Recall	F1-Score	Support
Negative	1.00	0.22	0.36	9
Neutral	0.80	0.84	0.82	62

Positive	0.90	0.93	0.92	128
Accuracy			0.87	199
Macro Avg	0.90	0.66	0.70	199
Weighted Avg	0.87	0.87	0.86	199

Table 9. Clasification Report SVM With SMOTE

Accuracy of SVM Without Smote : 88,44%				
	Precision	Recall	F1-Score	Support
Negative	1.00	0.56	0.71	9
Neutral	0.79	0.89	0.83	62
Positive	0.94	0.91	0.92	128
Accuracy			0.88	199
Macro Avg	0.91	0.78	0.82	199
Weighted Avg	0.89	0.88	0.88	199

Based on Tables 8 and 9 above, the SVM model with SMOTE demonstrates an improvement in recall and F1-score for neutral and negative classes, indicating better recognition of minority classes.

The increase in precision in the model with SMOTE also suggests that the proportion of correct predictions to total predictions has improved, meaning the model is capable of providing more accurate and relevant results.

The analysis results indicate that applying SMOTE to the SVM model improves data balance and significantly enhances the model's performance in classifying negative and neutral sentiments, which were previously underrepresented in the dataset. Based on the evaluation results, the model without SMOTE achieved an accuracy of 86.93%, while the model with SMOTE showed an improvement, reaching 88.44%. This increase demonstrates that the SMOTE technique helps the model recognize more diverse sentiment patterns instead of focusing solely on the majority class.

Additionally, the recall score for the negative class improved from 0.22 to 0.56, indicating that the model with SMOTE is more capable of identifying negative reviews than the model without SMOTE. However, despite the improvements in accuracy and recall, the application of SMOTE may also increase the risk of overfitting due to excessive synthetic data. Therefore, maintaining a balance in using this technique is essential to ensure that the model retains good generalization for new data.

Overall, the combination of SVM and SMOTE produces better results in analyzing sentiment in LP3I reviews compared to the model without SMOTE. However, for future implementation, additional testing with other data balancing techniques and further parameter tuning is necessary to achieve more accurate and reliable results.

4. CONCLUSION

Based on the sentiment analysis of public perception regarding LP3I's education quality in Indonesia using the SVM algorithm and SMOTE technique, several conclusions can be drawn. Most Google Maps reviews are positive, emphasizing competent teaching staff, an industry-based curriculum, and job opportunities, while negative reviews highlight administrative services, campus facilities, and the academic system. The SVM algorithm demonstrates high accuracy in sentiment classification, with SMOTE improving model performance by balancing sentiment distribution, especially for negative sentiments. The SVM model with SMOTE outperforms the one without it, although synthetic data may introduce a risk of overfitting. Model evaluation using accuracy, precision, recall, and F1-score confirms that SMOTE enhances sentiment classification balance without significantly compromising accuracy. These findings can help LP3I improve academic and administrative services while addressing key public concerns.

REFERENCES

[1] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, Jan. 2022, doi: 10.35969/interkom.v16i4.192.

[2] D. S. Utami, A. Erfina, and M. A. Id, “Analisis Sentimen Ulasan Terkait UNESCO Global Geopark Di Google Maps dengan Algoritma Naive Bayes,” 2022.

[3] Z. Sitorus, M. Saputra, S. N. Sofyan, and Susilawati, “SENTIMENT ANALYSIS OF INDONESIAN COMMUNITY TOWARDS ELECTRIC MOTORCYCLES ON TWITTER USING ORANGE DATA MINING,” *INFOTECH journal*, vol. 10, no. 1, May 2024, doi: 10.31949/infotech.v10i1.9374.

[4] M. Awaludin and A. G. Gani, “PEMANFAATAN KECERDASAN BUATAN PADA ALGORITMA K-MEANS KLASTERING DAN SENTIMENT ANALYSIS TERHADAP STRATEGI PROMOSI YANG SUKSES UNTUK PENERIMAAN MAHASISWA BARU,” *Jurnal Sistem Informasi*, vol. 11, pp. 1–6, 2024, doi: https://doi.org/10.35968/jsi.v11i1.1120.

[5] J. Anggraini and D. Alita, “Implementasi Metode SVM Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 2, pp. 102–111, Aug. 2024, doi: 10.30591/jpit.v9i2.6560.

[6] G. Aliman *et al.*, “Sentiment Analysis using Logistic Regression Sentiment Analysis using Logistic Regression,” *Journal of Computational Innovations and Engineering Applications*, vol. 7, no. 1, pp. 35–40, Jul. 2022.

[7] R. Turaina, R. Saputra, S. Informasi, U. Metamedia, and J. Khatib Sulaiman Dalam, “Optimalisasi Klasifikasi Umpan Balik Mahasiswa Terhadap Layanan Kampus dengan Sinergi Random Forest dan Smote,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 6, no. 6, 2023.

[8] F. Novitasari, E. Haerani, A. Nazir, J. Jasril, and F. Insani, “Sistem Klasifikasi Penyakit Jantung Menggunakan Teknik Pendekatan SMOTE Pada Algoritma Modified K-Nearest Neighbor,” *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3610.

[9] T. A. Dewi and E. Mailoa, “PERBANDINGAN IMPLEMENTASI METODE SMOTE PADA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM ANALISIS SENTIMEN OPINI MASYARAKAT TENTANG MIXUE,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 849–855, Sep. 2023, doi: 10.35870/jimik.v4i3.289.

[10] F. Destiyanti, A. Id Hadiana, F. Rakhmat Umbara, and U. Jenderal Achmad Yani Jl Terusan Jenderal Jenderal Sudirman, “Penerapan Metode Support Vector Machine dan SMOTE untuk Klasifikasi Sentimen Publik Terhadap Polisi Republik Indonesia,” *Jurnal Masyarakat Informatika Unjani*, vol. 8, no. 1, pp. 1–15, 2024, doi: https://doi.org/10.26874/jumanji.v8i1.336.

[11] N. Hafidz and D. Yanti Liliana, “Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 213–219, Apr. 2021, doi: 10.29207/resti.v5i2.2960.

[12] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” 2021.

[13] P. W. Rahayu *et al.*, *BUKU AJAR DATA MINING*, 1st ed., vol. 1. Jambi: PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://www.researchgate.net/publication/377415198

[14] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.

[15] A. Fatkhudin, F. Adi Artanto, N. Abiyu Safli, and D. Wibowo, “Remik: Riset dan E-Jurnal Manajemen Informatika Komputer Decision Tree Berbasis SMOTE dalam Analisis Sentimen Penggunaan Artificial Intelligence untuk Skripsi,” *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 8, no. 2, pp. 494–505, 2024, doi: 10.33395/remik.v8i2.13531.

[16] S. Sofyan and A. Prasetyo, “Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Tingkat Pendapatan Pekerja Informal Di Provinsi D.I. Yogyakarta Tahun 2019,” 2021.

[17] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*, 1st ed. Depok: Gunadarma Penerbit, 2023.

[18] A. Aditya *et al.*, *MACHINE LEARNING*, vol. 1. GLOBAL EKSEKUTIF TEKNOLOGI, 2023. [Online]. Available: www.globaleksekutifteknologi.co.id

[19] A. Mulyoto, *NAÏVE BAYES PADA GOOGLE COLABS*. EUREKA MEDIA AKSARA, 2024.

[20] R. G. Guntara, “Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7,” *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 1, pp. 55–60, Feb. 2023, doi: 10.47233/jteksis.v5i1.750.