Integra: Journal of Integrated Mathematics and Computer Science
Journal Homepage: https://integrajimcs.com/index.php/integra/index

integra
Journal of Integrated Mathematics and Computer Science

Research Paper

Understanding Consumer Sentiments: A TextBlob-Based Sentiment Analysis Study

Dian Kurniasari[1*], Yazid Zinedine Hdiana[1], Favorisen R. Lumbanraja[2], Warsono[1], Normi Abdul Hadi[3]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Lampung, Lampung, 35145, Indonesia*
[2] *Department of Computer Science, Faculty of Mathematics and Natural Science, Universitas Lampung, Lampung, 35145, Indonesia*
[3] *School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, 40450, Malaysia*
[*] **Corresponding author**: dian.kurniasari@fmipa.unila.ac.id

Keywords
Consumer Sentiments, Drug Reviews, Sentiment Analysis, TextBlob

Abstract
This study employs advanced sentiment analysis techniques to enhance the understanding of drug reviews, with a specific focus on TextBlob-based sentiment classification. As the accessibility of health products through pharmacies and online platforms continues to increase, individuals with limited health literacy are increasingly relying on user-generated feedback to inform their decision-making. By utilizing the TextBlob labelling method, this research categorizes user sentiments into positive, neutral, or negative, addressing the limitations inherent in traditional sentiment analysis approaches. The analysis is supported by an innovative model known as BERT, which effectively captures the emotional expression within textual data. The results indicate that the proposed approach consistently achieves a accuracy of 98% across training, validation, and testing phases, highlighting its strong performance in sentiment classification. This accomplishment highlights TextBlob's ability to consistently and reliably assess user sentiment, thereby enriching the understanding of consumer perspectives in the pharmaceutical industry. The results underscore the importance of effective sentiment analysis methods in healthcare, offering valuable insights for both consumers and stakeholders. Moreover, this study provides a foundation for future investigations focused on improving sentiment analysis methods across varied datasets, which will enhance the precision and applicability of classification results in different scenarios.

Received: 5 September 2025, Accepted: 2 November 2025
https://doi.org/10.26554/integrajimcs.20252340

1. INTRODUCTION

Sentiment analysis, also known as opinion mining, plays a crucial role in Natural Language Processing (NLP), focusing on the systematic detection and categorization of sentiments expressed in text. This process enables the evaluation of emotions and opinions expressed in written or spoken language, particularly in relation to specific products or services [1, 2, 3].

Sentiment analysis has been applied across various domains, including healthcare, entertainment, consumer products, travel, and politics, offering substantial advantages to individuals, governments, and businesses alike. Notably, its application within the business sector has garnered significant attention from researchers [4, 5, 6, 7, 8, 9]. For companies, sentiment analysis offers a powerful tool to gauge customer perceptions of products, services, and corporate reputation by mining data from social media and other relevant platforms. Moreover, online reviews, which often incorporate both ratings and detailed feedback, provide crucial insights into user experiences, helping guide potential customers in their decision-making process [10, 11].

In the business sector, consumer feedback is often represented through ratings and reviews. However, this method is susceptible to bias, which can result in discrepancies between the ratings and the actual sentiment expressed in reviews. For instance, it is not uncommon to find instances where high ratings are paired with negative comments or, conversely, low ratings with positive feedback [12]. Kordzadeh [13] has also underscored the potential for bias in rating-based assessments, which may undermine model accuracy. These findings align with recent research by Braja and Kodra [14], who examined reviews of the PUBG Mobile application on the Google Play Store. Their research revealed a significant disparity in model accuracy. While accuracy was 71% when reviews were analyzed based solely on ratings, it increased to 94% when user sentiments were also considered. These results highlight the limitations of relying solely on ratings and underscore the importance of sentiment analysis methods that accurately reflect user perceptions [15]. Consequently, this study adopts TextBlob as a sentiment labelling approach, converting consumer evaluations of a product into positive, neutral, or negative sentiments.

TextBlob is a pre-trained NLP library in Python commonly used to assess the polarity of reviews following the preprocessing phase. Each review is categorized into one of three polarity classes: positive, negative, or neutral, corresponding to values of 1, -1, and 0, respectively. TextBlob's lexicon includes both positive and negative terms, each assigned a specific score. The overall polarity score is derived from the aggregation of individual word scores, which is then utilized to determine the distribution and proportion of positive, neutral, and negative reviews within the dataset [16, 17].

Sentiment analysis methods are generally grouped into four major categories: lexicon-based approaches, traditional machine learning techniques, deep learning architectures, and hybrid models that integrate multiple strategies. Recently, researchers have embraced novel and increasingly prevalent methodologies, one of which is Bidirectional Encoder Representations from Transformers (BERT) [18]. Introduced by Google Research in 2018, BERT, a transformer-based model [19], integrates sentiment information by embedding contextual cues within its language representations, allowing the model to understand intersentential connections and grasp the broader context of the reviews. By leveraging such bidirectional representations, BERT, enhanced with additional output layers, can generate state-of-the-art models that are applicable across diverse scenarios [20].

This research seeks to evaluate the performance of TextBlob in sentiment classification of consumer reviews. TextBlob is used as a preliminary sentiment labelling tool, assigning sentiments—positive, negative, or neutral—based on the ratings present in the reviews. Moreover, to enhance the precision of the analysis, this research employs a transformer-based model, specifically BERT, for sentiment analysis and classification. BERT is utilized to explore intricate patterns within the labelled data, with the expectation of yielding more accurate results and thoroughly capturing the context of the reviews.

2. METHODS

The sentiment analysis in this study begins with data being loaded into Python, with Kaggle serving as the platform for coding and running the analysis. This initial stage is crucial to ensure both the accessibility and feasibility of the data within the selected programming environment. The dataset utilized in this study consists of textual data obtained from the Kaggle platform, containing a total of 184,622 records. The dataset consists of seven variables: uniqueID (the identifier for each respondent), drugName (the name of the medication utilized), condition (the health condition reported by the respondent), review (the feedback provided by the respondent regarding the drug), rating (the assessment of the drug as per the respondent's review), date (the date on which the review was submitted), and usefulCount (the number of respondents who found the review beneficial). Figure 1 presents a sample of the data to be utilized.

The subsequent phase involves data preprocessing, which is designed to identify and select pertinent variables for sentiment analysis. This phase involves determining the variables to be used as features and the variables to be used as targets. Additionally, this process involves cleansing the data of non-alphabetic characters that could disrupt the analytical process. Furthermore, case folding ensures consistency in textual representation, thereby mitigating any potential impact of variations in uppercase and lowercase letter usage on the analysis outcomes.

As noted by Tan et al. [21], preprocessing serves as a critical foundation for ensuring that the data is free from noise and well-structured. The absence of this step may lead to inaccuracies or biases in the analysis outcomes. Similarly, Kiersztyn [22] underscored the importance of implementing appropriate preprocessing techniques, asserting that inadequate data processing can lead to significant errors in interpreting the results of sentiment analysis.

The target variable is subsequently annotated using the TextBlob library. This annotation process is crucial, as it categorizes data into distinct sentiment classes—positive, negative, or neutral—facilitating a more structured and accessible sentiment analysis. Following the generation of these labels, the next step involves transforming the labels into a numerical format through a method known as label encoding. This approach assigns numerical values of 1, -1, and 0 to each predefined sentiment category, enabling the model to process the data more efficiently.

A significant challenge encountered in this study was the issue of data imbalance, characterized by a substantially lower number of samples in one class compared to the other classes. This disparity can adversely influence the analytical outcomes and hinder the model's ability to learn effectively from the available data. To address this issue, we employ a technique known as Random Oversampling (ROS). ROS is a simple resampling technique designed to balance class distribution within a dataset by artificially increasing the number of instances in the minority class. This process is conducted randomly, thereby enhancing the representation of minority classes in the dataset [23, 24, 25].

The dataset was partitioned, with 90% allocated for training and 10% for testing purposes. A portion of the training data was further separated to form a validation set. This tripartite division of the dataset was intended to mitigate the risks of overfitting and model selection bias, as articulated by Muraina [26]. They

	Unnamed: 0	patient_id	drugName	condition	review	rating	date	usefulCount	review_length
0	0	89879	Cyclosporine	keratoconjunctivitis sicca	"i have used restasis for about a year now and...	2.0	April 20, 2013	69	147
1	1	143975	Etonogestrel	birth control	"my experience has been somewhat mixed. i have...	7.0	August 7, 2016	4	136
2	2	106473	Implanon	birth control	"this is my second implanon would not recommen...	1.0	May 11, 2016	6	140
3	3	184526	Hydroxyzine	anxiety	"i recommend taking as prescribed, and the bot...	10.0	March 19, 2012	124	104
4	4	91587	Dalfampridine	multiple sclerosis	"i have been on ampyra for 5 days and have bee...	9.0	August 1, 2010	101	74
...	...	...	...	...	...	...	...	...	...
46103	6103	123432	Apri	birth control	"i started taking apri about 7 months ago. my ...	9.0	August 25, 2010	18	86
46104	6104	159999	Tamoxifen	breast cancer, prevention	"i have taken tamoxifen for 5 years. side effe...	10.0	September 13, 2014	43	97
46105	6105	140714	Escitalopram	anxiety	"i've been taking lexapro (escitaploprgram) si...	9.0	October 8, 2016	11	130
46106	6106	130945	Levonorgestrel	birth control	"i'm married, 34 years old and i have no kids....	8.0	November 15, 2010	7	149
46107	6107	47656	Tapentadol	pain	"i was prescribed nucynta for severe neck/shou...	1.0	November 28, 2011	20	34

184622 rows × 9 columns

Figure 1. Sample of Research Data

propose that the training set should comprise the most significant portion, while the validation and test sets could maintain a similar proportion. The training set enabled the model to learn underlying data patterns, while the validation set played a crucial role in model selection by assessing performance independently of the test data. Ultimately, the test set was reserved for evaluating the efficiency of the finalized model upon the completion of the training and validation phases.

Furthermore, the preliminary application of BERT involved transforming the review text into a structured set of inputs represented as tokens. In this context, the model has a maximum input length of 230 tokens, meaning that no individual input can surpass this limit. The review data exhibits variability in sentence length; consequently, sentences exceeding 230 tokens are truncated to align with the specified constraints, thereby ensuring accurate model processing. Conversely, for sentences shorter than the maximum token limit, padding is employed, which involves adding supplementary tokens to extend the sentence length to 230 tokens. This approach ensures uniformity in input lengths provided to the BERT model, thereby enhancing the model's training efficiency and accuracy.

A BERT model's architecture generally comprises two sequential phases: pre-training followed by fine-tuning [27]. In the pre-training stage, the model is exposed to a large corpus of unlabeled text data, including English literature and web-based content, enabling it to learn inherent patterns, syntactic structures, and semantic relationships within the text. Following pre-training, the BERT model can be tailored for various applications, such as classification or sentiment analysis, by incorporating an additional output layer, thereby minimizing the need for substantial alterations to the overall architecture. The foundational structure of the BERT model is characterized by transformer layers that exclusively feature the encoder component, as illustrated in the accompanying figure. In practical use, the BERT model is available in two configurations: BERT Base and BERT Large. The distinction between these two lies in the number of encoder layers, multi-head self-attention units, hidden layer sizes, and overall parameter counts. Structurally, BERT Base and BERT Large differ significantly. BERT Base is built with 12 encoder layers, each incorporating 12 multi-head self-attention units, and operates with a hidden size of 768, totalling 110 million parameters. BERT Large, on the other hand, scales up the architecture to 24 encoder layers, 16 attention heads, a hidden size of 1024, and 340 million parameters [19]. Figure 2 illustrates the two BERT architectures.

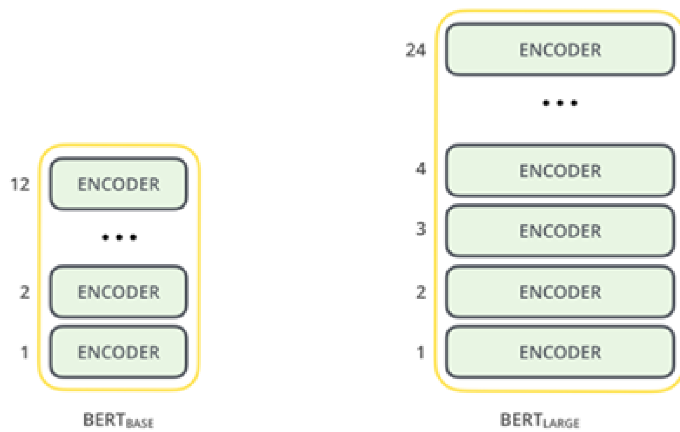

Figure 2. BERT BASE and BERT LARGE Architectures (Source: GitHub)

This study employs the BERT BASE model, selected due to the volume of the research dataset and the appropriateness of its parameter count. During the fine-tuning phase, the BERT Base model is optimized for sentiment classification using BertForSequenceClassification, incorporating several critical parameters, including batch size, dropout rate, learning rate, and adjusted number of epochs. The methodology involves aligning task-specific inputs and outputs with the BERT Base model, followed by comprehensive fine-tuning of all parameters. In the input phase, Sentence A and Sentence B from the pre-training are tailored to suit the objectives of text classification or sentiment analysis. Simultaneously, during the output phase, the [CLS] token representation is utilized as input to the output layer to perform classification tasks, including sentiment analysis. Upon completing the training process, the model's capability to classify novel data is evaluated using test datasets. Figure 3 illustrates the BERT fine-tuning architecture for text classification or sentiment analysis.

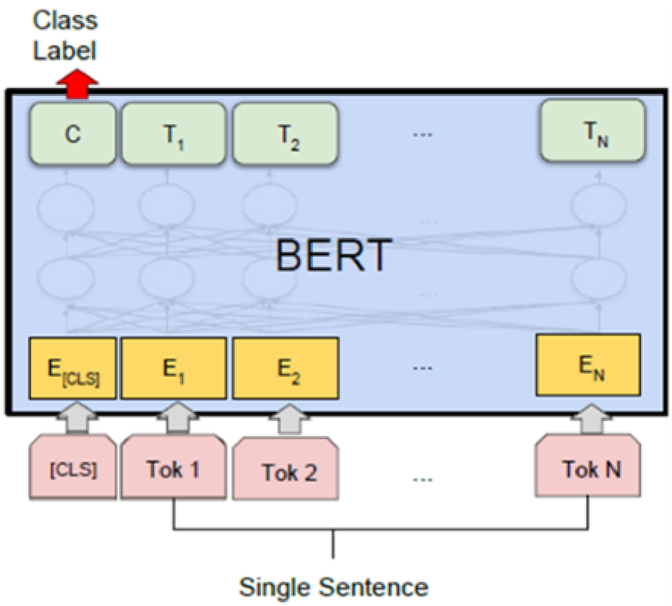

Figure 3. Fine-tuning BERT

The final evaluation phase involves analyzing several key metrics, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates. The term True Positive (TP) denotes instances that truly belong to the positive class and are correctly identified as such by the model. In contrast, TN represents the total of negative instances accurately identified by the model. In contrast, FP denote instances that were incorrectly labelled as positive, while FN correspond to cases that were wrongly classified as negative. These four parameters are integral components of the confusion matrix. The confusion matrix serves as a valuable metric for visualizing the predicted outcomes of a model's classification in relation to the actual classifications. This matrix is typically organized as a 2x2 grid, where the rows represent the predicted classifications and the columns denote the actual classifications. In its simplest form, the matrix accounts for two classes: the positive class and the negative class [28]. From the confusion matrix, key evaluation metrics—including Accuracy, Precision, Recall, and F1-score—can be computed as shown in Equations (1) through (4):

$$\text{Accuracy} = \frac{TP + TN}{TP + FF + TN + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - \text{score} = \frac{2 \times \text{recall} \times \text{precision}}{(\text{recall} + \text{precision})} \tag{4}$$

Accuracy is widely recognized as a fundamental metric for assessing the effectiveness of classification models, which is determined by dividing the number of correct predictions by the total number of instances in the dataset [29]. Precision reflects the proportion of correctly identified optimistic predictions among all predictions labelled as positive by the model. In contrast, recall evaluates the model's ability to capture true positive cases among all actual positives. To balance these two aspects, the F1 score is employed, as it merges precision and recall into a single measure, making it especially valuable when dealing with datasets that have imbalanced class distributions [30].

3. RESULTS AND DISCUSSION

Based on the methodological framework outlined in Section 2, the preliminary phase of this study involves preprocessing. Table 1 illustrates a subset of the dataset that has been processed through the preprocessing pipeline.

Table 1. Data Samples Before and After the Preprocessing

Before Preprocessing	After Preprocessing
"i was prescribed nucynta for severe neck/shoulder pain. after taking only 2, 75mg pills i was rushed to the er with severe breathing problem. i have never had any issues with pain medicines before."	i was prescribed nucynta for severe neckshoulder pain after taking only mg pills i was rushed to the er with severe breathing problems i have never had any issues with pain medicines before

Table 1 illustrates the transformation of the initial sentence, which encompasses critical information regarding the side effects of medications. Prior to preprocessing, the sentence exhibited issues with punctuation and formatting. Following the preprocessing phase, the sentence was refined by removing extraneous punctuation, thereby enhancing its structural clarity for subsequent analysis. The next phase involves applying TextBlob for labelling, where the processed data will be annotated to facilitate further classification and information processing. Figure 4 illustrates the results of the data labelling process, where Class 0 represents negative sentiment, Class 1 represents neutral sentiment, and Class 2 corresponds to positive sentiment.

Nevertheless, subsequent analyses uncovered an imbalance in the distribution of class labels, which may compromise the efficacy of the developed model. Consequently, the researcher implemented balancing techniques to ensure that the model performs optimally across all existing classes. Figure 5 and Table 2 indicate that the data distribution is primarily dominated by neutral sentiment, followed by positive sentiment, and then negative sentiment. This imbalance poses a risk of bias in the analysis if not adequately addressed, as the model may yield less

	review	rating	sentimen_textblob
0	i have used restasis for about a year now and ...	0	1
1	my experience has been somewhat mixed i have b...	3	2
2	this is my second implanon would not recommend...	0	2
3	i recommend taking as prescribed and the bottl...	4	1
4	i have been on ampyra for days and have been s...	4	2
...	...	...	...
46103	i started taking apri about months ago my brea...	4	1
46104	i have taken tamoxifen for years side effects ...	4	1
46105	ive been taking lexapro escitaploprgram since ...	4	1
46106	im married years old and i have no kids taking...	3	1
46107	i was prescribed nucynta for severe neckshould...	0	1

184622 rows × 3 columns

Figure 4. Data After Labelling

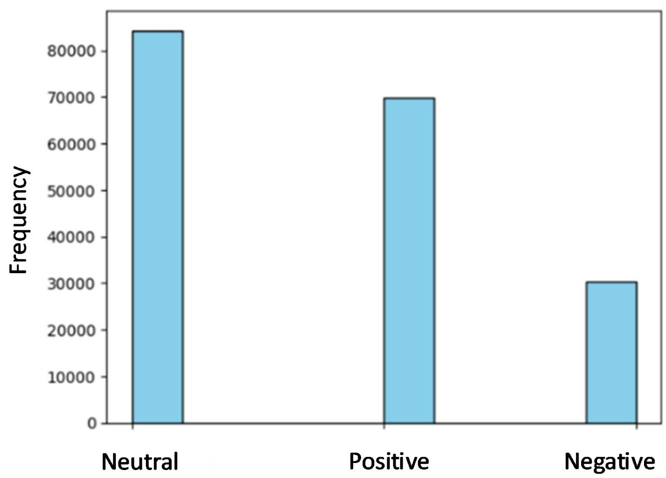

Figure 5. Histogram of Sentiment Class Distribution

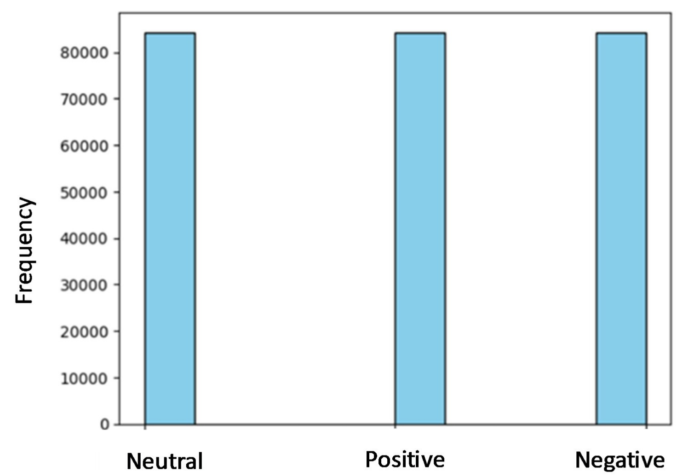

Figure 6. Histogram of Class Distribution After ROS

accurate predictions for minority classes with limited sample sizes.

Table 2. Class Distribution

Sentiment Textblob	Frequency
Negative	30423
Neutral	84258
Positive	69941

The issue of data imbalance, which can significantly impact the performance of classification models, has been effectively addressed through the application of the ROS technique. This method has increased the number of samples in the minority class, resulting in an equal number of samples between the minority and majority classes. Figure 6 and Table 3 present the histogram and class distribution table, respectively, demonstrating that the dataset now contains an equal number of samples across all classes (negative, neutral, and positive). This balanced dataset enables progression to the subsequent phase, specifically the training of the BERT model.

Table 3. Textblob Sentiment Class Distribution After ROS

Sentiment Textblob	Frequency
Negative	84258
Neutral	84258
Positive	84258

Table 4 presents the outcomes of the training conducted with the BERT Base model. These findings highlight the strong performance of the proposed model, which achieves a training accuracy of 99.66%, closely followed by validation and testing accuracies of 98.29% and 98.28%, respectively, indicating excellent generalization capability. These metrics suggest that the model not only excels during the training phase but also effectively generalizes the knowledge it has acquired when encountering new data.

Table 4. BERT Classification Results

Method	Accuracy Training	Accuracy Validation	Accuracy Testing
Textblob	0.9966	0.9829	0.9828

The model's performance in classifying various categories is illustrated more clearly and comprehensively by the confusion matrix. Figure 7 presents performance models.

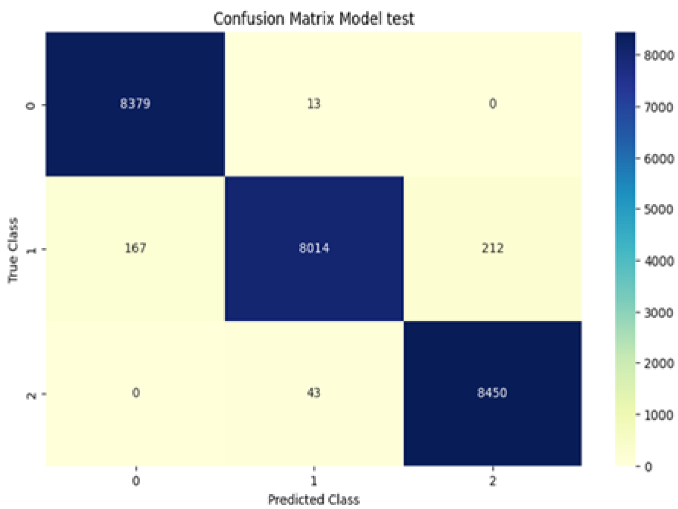

Figure 7. Confusion Matrix

The precise classification occurs along a diagonal line, representing the number of reviews accurately categorized into their respective classes. Conversely, values located outside this diagonal indicate classification errors, where reviews have not been correctly assigned. Consequently, the data illustrated in Figure 6 demonstrates that the BERT model successfully categorizes the majority of consumer reviews, with specific counts of 8,379 reviews classified as class 0 (Negative), 8,014 as class 1 (Neutral), and 8,450 as class 3 (Positive). Furthermore, Table 5 presents a comprehensive evaluation of the model's performance, including precision, recall, and F1-score.

Table 5. Sentiment-based Classification Results on Testing Data

Class	Precision	Recall	F1-score	Support
0	0.98	1.00	0.99	8392
1	0.99	0.95	0.97	8393
2	0.98	0.99	0.99	8493
Accuracy			0.98	25278
Macro Avg	0.98	0.98	0.98	25278
Weighted Avg	0.98	0.98	0.98	25278

The findings presented in Table 5 indicate that the model attains a precision of 98%, a recall of 100%, and an F1-score of 99% for class 0, demonstrating its high accuracy in classifying this category. For class 1, the model achieves an accuracy of 99%, though its recall is slightly lower at 95%, suggesting that there are some misclassifications within this class. Class 2 exhibits performance comparable to that of class 0, with a precision of 98% and a recall of 99%.

Overall, the model achieves an accuracy of 98% on the test dataset, with both the macro mean and weighted average metrics also reflecting a value of 98%. These results indicate that the BERT Base model not only effectively recognizes user sentiment but also exhibits strong generalization capabilities across various data classes.

The results suggest that the developed model is dependable for conducting user sentiment analysis using TextBlob, as evidenced by its low error rate. The model's reliable performance throughout the training, validation, and testing phases demonstrates its relevance in many applications, such as sentiment analysis of customer feedback. The slight variations in precision and recall among classes present opportunities for improvement, either through the use of enhanced datasets or more sophisticated data processing techniques.

4. CONCLUSIONS

The study evaluated the effectiveness of TextBlob in classifying consumer review sentiments, with a focus on detecting positive, negative, and neutral views. The findings indicate that, following data preprocessing and class balancing, the BERT Base model attained an accuracy of 98% in sentiment classification on the test dataset. These findings illustrate the model's ability to generalize knowledge from training to unfamiliar data and comprehensively capture the context of the review. The BERT model consistently achieves high precision, recall, and F1 scores across all sentiment classes, despite minor fluctuations in the individual metrics. This research substantially progresses the development of improved sentiment analysis techniques, which are expected to be widely employed for a deeper understanding of consumer attitudes towards products and services. The findings of this study demonstrate that sentiment analysis, using TextBlob and BERT, can successfully comprehend consumer opinions in various contexts.

5. ACKNOWLEDGEMENT

The author expresses gratitude to the University of Lampung for funding this research through the Basic Research Scheme, Contract Number 606/UN26.21/PN/2024.

REFERENCES

[1] Dhrubajyoti Hazarika, Gitimallika Konwar, Suman Deb, and Dhruba J Bora. Sentiment analysis on twitter by using textblob for natural language processing. In *Proceedings of the International Conference on Research in Management & Technovation 2020*, pages 63–67, 2020.

[2] Purnima Garg, Shweta Tyagi, Anshul Joshi, Ashish Pandey, and Divakar Panwar. Supervised machine learning approaches for customer reviews sentiment analysis. In *Lecture Notes in Networks and Systems*, pages 211–223. Springer Nature Singapore, 2024.

[3] Jannatun Roba Jim, Md Al Amin Talukder, Promit Malakar, Md Monzurul Kabir, Kazi Nur, and Md Faijalah Mridha. Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 6:100059, 2024.

[4] Amara Rashid and Chen Huang. Sentiment analysis on consumer reviews of amazon products. *International Journal of Computer Theory and Engineering*, 13(2):35–41, 2021.

[5] Elham Asani, Hatef Vahdat-Nejad, and Javad Sadri. Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*, 6:100114, 2021.

[6] Weidong Huang, Min Lin, and Yutong Wang. Sentiment analysis of chinese e-commerce product reviews using ernie word embedding and attention mechanism. *Applied Sciences*, 12(14):7182, 2022.

[7] Nausheen Sultana, Pavan Kumar, Subhash Chandra, Mahindra Rajesh Patra, and SK Safdar Alam. Sentiment analysis for product review. *International Research Journal of Modernization in Engineering Technology and Science*, 5(3):2501–2503, 2023.

[8] Varuna Gooljar, Thea Issa, Shri Hardin-Ramanan, and Bilal Abu-Salih. Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review. *Journal of Big Data*, 11(1):107, 2024.

[9] TA Kumar, J Zaafira, P Kanimozhi, R Rajmohan, C Ananth, and SA Ajagbe. Machine learning and sentiment analysis: Analyzing customer feedback. In *AI-Driven Marketing Research and Data Analytics*, pages 245–262. IGI Global, 2024.

[10] Pavan Kumar Jain, Rajasekhar Pamula, and Emre A Yekun. A multi-label ensemble predicting model to service recommendation from social media contents. *The Journal of Supercomputing*, 78(4):5203–5220, 2022.

[11] Qian Amy Xu, Victor Chang, and Chrisina Jayne. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3:100073, 2022.

[12] Zhi-Hong Zhang, Qiong-Biao Ye, Zheng-Yan Zhang, and Yong-Li Li. Sentiment classification of internet restaurant reviews written in cantonese. *Expert systems with applications*, 38(6):7674–7682, 2011.

[13] Nima Kordzadeh. Investigating bias in the online physician reviews published on healthcare organizations' websites. *Decision Support Systems*, 118:70–82, 2019.

[14] Anggita S P Braja and Abdul Kodar. Implementasi finetuning bert untuk analisis sentimen terhadap review aplikasi pubg mobile di google play store. *J I M P-J. Inform. Merdeka Pasuruan*, 7(3):120–128, 2023.

[15] Samer Al-Natour and Ozgur Turetken. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management*, 54:102132, 2020.

[16] Frank Z Xing, Federico Pallucchini, and Erik Cambria. Cognitive-inspired domain adaptation of sentiment lexicons. *Information Processing & Management*, 56(3):554–564, 2019.

[17] Sukruth Mendon, Piyush Dutta, Abhishek Behl, and Stefan Lessmann. A hybrid approach of machine learning and lexicons to sentiment analysis: Enhanced insights from twitter data of natural disasters. *Information Systems Frontiers*, 23(5):1145–1168, 2021.

[18] Yirong Mao, Qian Liu, and Yan Zhang. Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University-Computer and Information Sciences*, 36(4):102048, 2024.

[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[20] An Zhao and Yanyan Yu. Knowledge-enabled bert for aspect-based sentiment analysis. *Knowledge-Based Systems*, 227:107220, 2021.

[21] Ken Li Tan, Chien-Ping Lee, and Kian-Ming Lim. A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7):4550, 2023.

[22] Agnieszka Kiersztyn and Krzysztof Kiersztyn. The impact of data preprocessing on prediction effectiveness. In *Artificial Intelligence and Soft Computing: 21st International Conference, ICAISC 2023, Zakopane, Poland, June 18–22, 2023, Proceedings, Part I*, pages 353–362. Springer, 2023.

[23] Sanaz Fotouhi, Saeideh Asadi, and Mugahed W Kattan. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *Journal of biomedical informatics*, 90:103089, 2019.

[24] Joseph M Johnson and Taghi M Khoshgoftaar. Deep learning and thresholding with class-imbalanced big data. In *2019 18th IEEE international conference on machine learning and applications (ICMLA)*, pages 755–762. IEEE, 2019.

[25] Mardhiah Hayaty, Muthmainah Muthmainah, and Sayed M Ghufran. Random and synthetic over-sampling approach to resolve data imbalance in classification. *Int. J. Artif. Intell. Res*, 4(2):86–92, 2021.

[26] Ibrahim O Muraina. Ideal dataset splitting ratios in machine learning algorithms: General concerns for data scientists and data analysts. In *7th Int. Mardin Artuklu Sci. Res. Conf.*, pages 496–504, 2022.

[27] Subhankar Paul and Srijan Saha. Cyberbert: Bert for cyberbullying identification: Bert for cyberbullying identification. *Multimedia Systems*, 28(6):1897–1904, 2022.

[28] Asti Aprilia Arifiyanti, Ramadhana Mahendra Pradana, and Ivan Fatcha Novian. Klasifikasi produk retur dengan algoritma pohon keputusan c4. 5. *J. IPTEK*, 22(1):79–86, 2018.

[29] Pradeep Kumar Yechuri and Sriram Ramadass. Classification of image and text data using deep learning-based lstm model. *Traitement du Signal*, 38(6):1809–1817, 2021.
[30] Mohammad Hossin and M Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):1–11, 2015.

