J-KOMA Journal of Computer Science and Applications

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

Open Access

E-ISSN: *2620 - 4872*
Vol.*08*, ***No***.*01*
Doi: *https://doi.org/10.21009/j-koma.v8i1.05*

Recived: *06 April 2025*
Accepted: *15 Mei 2025*
Published: *22 Juni 2025*

Keywords:
Sentiment Analysis;
Naïve Bayes;
SVM;
TF-IDF.

Correspondence Email:
ria.arafiyah@unj.ac.id

Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes

Verra Budhi Lestari[1*], Carli Apriansyah Hutagalung[2]

[1]Universitas Media Nusantara Citra, Jakarta
[2] Computer Science Study Program, Universitas Negeri Jakarta

Abstract
This study evaluates the application of TF-IDF feature extraction in sentiment analysis of Indonesian-language marketplace product reviews using Logistic Regression, Naïve Bayes, and Support Vector Machine (SVM) algorithms. The dataset, sourced from Kaggle, comprises 831 reviews (385 positive, 446 negative), processed through preprocessing steps including text cleaning, tokenization, stopword removal, and stemming. The data was split into 80% training and 20% testing sets. Results show that Logistic Regression with TF-IDF achieved the highest performance, with 90.4% accuracy, 91.8% precision, 90.4% recall, and 90.9% F1-measure, outperforming Naïve Bayes (87.4% accuracy) and SVM (89.8% accuracy). Logistic Regression effectively captures linear relationships in TF-IDF features, while Naïve Bayes struggles with emotional context, and SVM requires complex parameterization. TF-IDF is efficient for explicit reviews but limited in handling complex semantic contexts like sarcasm. This study confirms that Logistic Regression combined with TF-IDF is the most effective approach for sentiment analysis of Indonesian marketplace reviews, with recommendations for future exploration of methods like word embedding.

INTRODUCTION

Based on the rapid development of internet technology, various activities can now be conducted online, including shopping. At present, online shopping has become the primary medium for users to conduct purchasing transactions and consume products [1]. In expressing satisfaction with online shopping activities, users have the right to share their opinions related to these activities, ranging from store services, product quality, product prices, seller attitudes, to the delivery process [2]. Moreover, consumer reviews are highly valuable for businesses to assess seller performance and beneficial for customers as they provide insights into what can be expected from a new product [3].

In addition, marketplace users rely on reviews from other users to evaluate a product's value before making a purchase. These reviews influence consumer sentiments and emotions when grouping product reviews to determine the polarity between positive and negative reviews [2]. Sentiment analysis can analyze user reviews to extract important knowledge that can be utilized by others in decision-making [4], [5]. Thus, user satisfaction can be effectively improved by leveraging sentiment analysis techniques applied to large-scale reviews on marketplace platforms [1].

However, sentiment analysis in the Indonesian language still faces various challenges. One of the main issues is the linguistic characteristics of user reviews, which are often informal, non-standard, or

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

contain slang expressions. Unlike English, which is well supported by NLP (Natural Language Processing) resources and extensive prior research, Indonesian sentiment analysis remains underexplored, particularly in terms of feature representation and algorithms [6].

Previous studies on sentiment analysis have examined product reviews in several languages, including English [7], [8], Bangla [9], Chinese [10], and Indonesian [11]. Several works have employed methods such as SVM, Logistic Regression, and Naïve Bayes [2], [7], [9], [11] to perform sentiment classification on product reviews. Various feature extraction methods have also been applied in these studies.

A study proposed the application of the Support Vector Machine (SVM) algorithm [12] to classify sentiments in e-commerce reviews with high variability. The dataset consisted of customer reviews from platforms such as Daraz, Picabbo, Bikroy.com, and Ajkerdeal, collected via web scraping in Python using the BeautifulSoup library. Preprocessing steps were applied to handle unstructured data, including stopword detection, tokenization, and data normalization. The application of SVM achieved an accuracy of 82.92%.

Another study used a dataset of 2006 positive and 4783 negative reviews related to products such as cameras, laptops, mobile phones, tablets, televisions, and surveillance systems, obtained from Amazon [13]. Preprocessing steps included tokenization, stopword removal, stemming, and punctuation removal. The classification process employed Naïve Bayes and Support Vector Machine (SVM). Among the tested algorithms, Naïve Bayes achieved the best performance with an accuracy of 98.17%, while SVM achieved 93.54%.

Several prior studies demonstrated that TF-IDF is a simple yet effective text representation method for text classification tasks, including sentiment analysis. TF-IDF assigns weights to words based on their importance within a document relative to the entire corpus. When combined with algorithms such as Support Vector Machine (SVM) and Logistic Regression, TF-IDF successfully improved classification accuracy in Indonesian texts [14], [15]. Nevertheless, there is still a lack of systematic studies evaluating the performance of multiple machine learning algorithms (Naïve Bayes, SVM, and Logistic Regression) with TF-IDF features specifically for Indonesian marketplace reviews. Each algorithm has its own strengths and limitations in handling complex and unstructured text data.

This study proposes the application of TF-IDF feature extraction combined with several machine learning algorithms, namely Logistic Regression, Multinomial Naïve Bayes, and Support Vector Machine, to evaluate their performance and effectiveness in sentiment analysis of Indonesian-language marketplace product reviews. The evaluation will be conducted using accuracy, precision, recall, and F1-score as performance metrics..

METHOD

The motivation of this study is to analyze customer reviews on marketplace platforms using supervised learning methods under the machine learning approach for sentiment analysis representation. Several steps are carried out in this research, and to facilitate understanding, these steps are illustrated in the diagram presented in Figure 1.

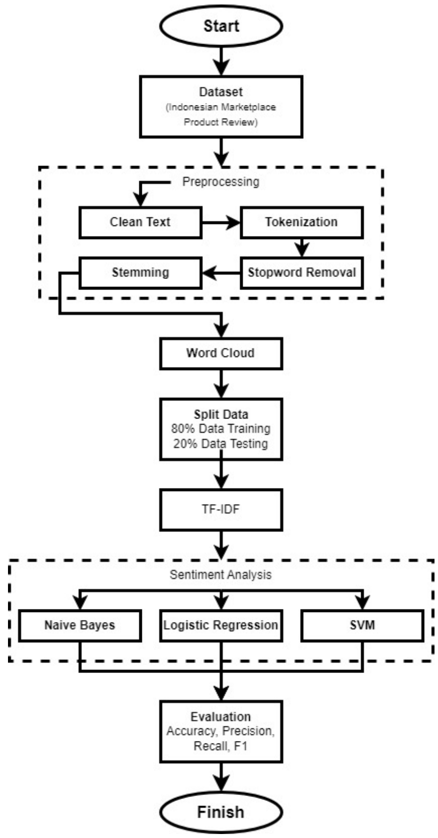

FIGURE 1. Research Flow

The process of sentiment analysis on marketplace user reviews begins with the use of a dataset obtained from Kaggle.com as the primary data source. This dataset already contains labels for each sentiment category, namely “1.0” for positive sentiment reviews and “0.0” for negative sentiment reviews. The researchers added additional labels to further clarify the sentiment of each review. Before applying the dataset to the model, several preprocessing steps were conducted, including text cleaning, tokenization, stopword removal, and stemming, which will be discussed in the Data Preprocessing section. After preprocessing, the dataset was divided into two parts: 80% for training data and 20% for testing data. The data was then trained and tested using three models: Naïve Bayes with TF-IDF, Logistic Regression with TF-IDF, and Support Vector Machine with TF-IDF. The performance of these three methods was evaluated and validated using accuracy, precision, recall, and F1-Measure to determine the best-performing method for sentiment analysis of marketplace user reviews.

Sentiment Analysis

Sentiment analysis is the study of analyzing people’s expressions toward an object expressed in written text; therefore, sentiment analysis can also be defined as opinion mining. The objects of sentiment analysis may include products, services, individuals, events, and many others. Several types of analysis within the concept of sentiment analysis include emotion analysis, influence analysis, subjectivity analysis, and review mining [1].

Dataset and Labeling

The dataset used in this study is titled “Indonesian Marketplace Product Review.” This dataset was developed for sentiment analysis research using machine learning methods. The data consists of product review texts and their corresponding sentiment labels. The dataset was collected from one of the most widely used marketplaces in Indonesia. The data collection method employed was web scraping. The dataset was collected on May 18, 2022, and subsequently labeled manually by annotators.

The dataset comprises 831 product reviews, which are divided into two sentiment categories: 385 positive reviews and 446 negative reviews. This dataset can be accessed via Kaggle.com (API command: kaggle datasets download -d taqiyyaghazi/indonesian-marketplace-product-reviews) or through the URL: *https://www.kaggle.com/datasets/taqiyyaghazi/indonesian-marketplace-product-reviews*. An example of five data entries from the dataset is presented in Table 1.

TABLE 1. Confusion Matrix

No	Ulasan	Rate
1	kemeja nya bagusss bgtttt 😍😍😍 aaaa mauuu nngisssss 😩😩😩 knpa ga dri dlu beli kemeja ditoko ini😜, ini kemejanya asli emg bagus, bahannya jga adem ga gerah,and ga nerawang jga... itu krna camera nya jelek jdi ga trlalu jelas kemejanya.. asliny baguss bgttt ga bhong sumpaha	1.0
2	Jahitannya sih rapi,cuman ada benang yang ikut ke jahit juga jadi agak jelek	0.0
3	Sesuai harga. Agak tipis tapi masih oke kok. Warnanya abu tapi kalo difoto emang kayak biru dikit. Thanks seller	0.0
4	Wah gila sihhh sebagus itu, se worth it, se lembut itu bajunya,,,, kirain bakal terlalu tipis ky kemeja ku yg lama ternyata sedikit ky kaos bajunya baklgus bettt dah dgn harga segitu sih worth it thankyou yh,, next bakal order lgi	1.0
5	Kain nya bagus halus Tapi kok di bukak kotor ya warna putih lagi	0.0

Term Frequency-Inverse Document Frequency (TF-IDF)

The labeled review data underwent several preprocessing stages before being divided into training and testing subsets. First, the clean text step was applied to remove unnecessary elements such as punctuation, numbers, emoticons, and symbols, while also converting all text into lowercase form to maintain consistency. Next, the tokenization process was performed to split the textual data into smaller units, or tokens, based on each word that appears adjacently in the text [19]. Afterward, stopword removal was carried out to eliminate frequently occurring words that do not contribute meaningful information to sentiment analysis, such as Indonesian conjunctions (atau, dan, di, untuk, dari, yang, ke). Finally, the stemming process was conducted to reduce words to their root form by removing affixes, thereby minimizing word variation. In the Indonesian language, this includes affixes such as meng-, me-, ke-, kan-, di-, se-, pe-, peng-, -nya, and -kan.

$$idf(i) = log\left(\frac{D}{D_i}\right) \quad (1)$$

$$tf_i = \left(\frac{n_i}{N}\right) \quad (2)$$

$$TF - IDF = tf_i * idf(i) \quad (3)$$

Logistic Regression

Another supervised machine learning approach commonly used for binary classification problems is Logistic Regression. The best way to conceptualize Logistic Regression is as a linear regression technique applied to classification tasks. Unlike Linear Regression, which produces outputs over a continuous range, Logistic Regression constrains its outputs to values between 0 and 1 through the use of a logistic function. This enables it to model binary outcome variables effectively. Furthermore, Logistic Regression does not require a linear relationship between input variables and the target variable, as it applies a nonlinear transformation to the odds ratio, making it highly suitable for categorical prediction problems.

Naïve Bayes

Naïve Bayes is a probabilistic classification technique that works effectively with high-dimensional feature spaces. Given a dataset represented as an n-dimensional feature vector, the classifier estimates the probability of a data instance belonging to each possible class and assigns it to the class with the highest posterior probability. The model assumes conditional independence between features given the class label, which simplifies computation and allows the classifier to handle large-scale datasets efficiently. Despite its "naïve" assumption, Naïve Bayes has been widely applied in sentiment analysis tasks and has demonstrated competitive performance, particularly with textual data.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a robust classification algorithm that can perform both linear and non-linear classification. The core idea of SVM is to find an optimal hyperplane that maximizes the margin between two classes, ensuring that the closest data points, called support vectors, are as far as possible from the decision boundary. When data is not linearly separable, SVM employs kernel functions to map the data into higher-dimensional spaces, enabling more effective separation of complex patterns. By maximizing the geometric margin between classes, SVM achieves strong generalization capabilities, making it one of the most reliable algorithms for sentiment analysis tasks [19].

RESULTS AND DISCUSSION

Data and Preprocessing

The product review dataset used in this study consists of two categories: reviews with positive sentiment and reviews with negative sentiment. As shown in Figure 2, the dataset comprises 385 positive reviews and 446 negative reviews.

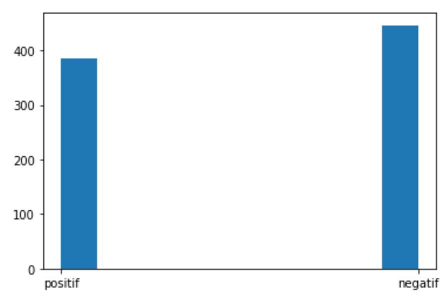

FIGURE 2. Distribution of Review Data

Next, the review data is cleaned from unnecessary elements. The process begins by converting all text into lowercase, followed by the removal of punctuation marks and numbers. The review data is then converted and broken down into components in the form of tokens during the tokenization stage. Subsequently, the stopword removal and stemming stages are applied using the Sastrawi library for the Indonesian language. After completing the preprocessing steps, as shown in Table 3, the review data can be used for the main stage of this study, namely the classification process.

TABLE 3. Stages and Results of Preprocessing

Preprocessing	Result
Before Preprocessing	kemeja nya bagusss bgtttt 😍😍😍 aaaa mauuu nngisssss 😩😩😩 knpa ga dri dlu beli kemeja ditoko ini😜, ini kemejanya asli emg bagus, bahannya jga adem ga gerah,and ga nerawang jga... itu krna camera nya jelek jdi ga trlalu jelas kemejanya.. asliny baguss bgttt ga bhong sumpaha
Clean Text	kemeja nya bagusss bgtttt 😍😍😍 aaaa mauuu nngisssss 😩😩😩 knpa ga dri dlu beli kemeja ditoko ini😜, ini kemejanya asli emg bagus, bahannya jga adem ga

	gerah,and ga nerawang jga... itu krna camera nya jelek jdi ga trlalu jelas kemejanya.. asliny baguss bgttt ga bhong sumpaha
Tokenization	[['kemeja', 'nya', 'bagusss', 'bgtttt😍😍😍aaaa', 'mauuu', 'nngisssss😩😩😩knpa', 'ga', 'dri', 'dlu', 'beli', 'kemeja', 'ditoko', 'ini😜', 'ini', 'kemejanya', 'asli', 'emg', 'bagus', 'bahannya', 'jga', 'adem', 'ga', 'gerahand', 'ga', 'nerawang', 'jga', 'itu', 'krna', 'camera', 'nya', 'jelek', 'jdi', 'ga', 'trlalu', 'jelas', 'kemejanya', 'asliny', 'baguss', 'bgttt', 'ga', 'bhong', 'sumpah']
Stopword Removal	['kemeja', 'nya', 'bagusss', 'bgtttt😍😍😍aaaa', 'mauuu', 'nngisssss😩😩😩knpa', 'ga', 'dri', 'dlu', 'beli', 'kemeja', 'ditoko', 'ini😜', 'kemejanya', 'asli', 'emg', 'bagus', 'bahannya', 'jga', 'adem', 'ga', 'gerahand', 'ga', 'nerawang', 'jga', 'krna', 'camera', 'nya', 'jelek', 'jdi', 'ga', 'trlalu', 'jelas', 'kemejanya', 'asliny', 'baguss', 'bgttt', 'ga', 'bhong', 'sumpah']
Stemming	['kemeja', 'nya', 'bagusss', 'bgtttt aaaa', 'mauuu', 'nngisssss knpa', 'ga', 'dri', 'dlu', 'beli', 'kemeja', 'toko', 'ini', 'kemeja', 'asli', 'emg', 'bagus', 'bahan', 'jga', 'adem', 'ga', 'gerahand', 'ga', 'nerawang', 'jga', 'krna', 'camera', 'nya', 'jelek', 'jdi', 'ga', 'trlalu', 'jelas', 'kemeja', 'asliny', 'baguss', 'bgttt', 'ga', 'bhong', 'sumpah']
Next Preprocessing	'kemeja nya bagusss bgtttt aaaa mauuu nngisssss knpa ga dri dlu beli kemeja toko ini kemeja asli emg bagus bahan jga adem ga gerahand ga nerawang jga krna camera nya jelek jdi ga trlalu jelas kemeja asliny baguss bgttt ga bhong sumpah'

The visualization of words containing positive and negative sentiments is presented according to their sentiment, as shown in Figure 3 and Figure 4.

Figure 3. Positive Word Visualization

Figure 4. Negative Word Visualization

In this study, the dataset was divided into two parts: 80% (664 data points) for training and 20% (167 data points) for testing, with the data randomized using random state 42. The classification results were then obtained from the implementation of the Naïve Bayes algorithm (Multinomial Naïve Bayes) with TF-IDF feature extraction, as shown in Figure 5, the Logistic Regression algorithm with TF-IDF feature extraction, as shown in Figure 6, and the Support Vector Machine algorithm, as shown in Figure 7.

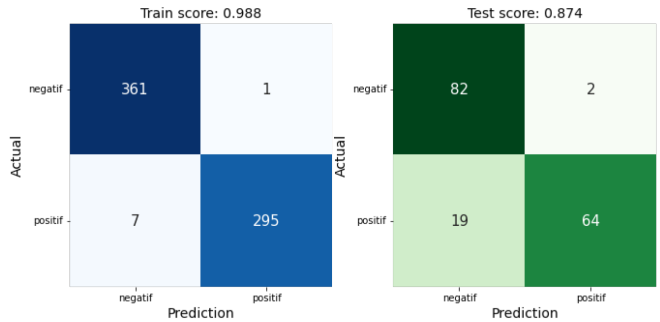

FIGURE 5. Confusion Matrix of Naïve Bayes – TF-IDF

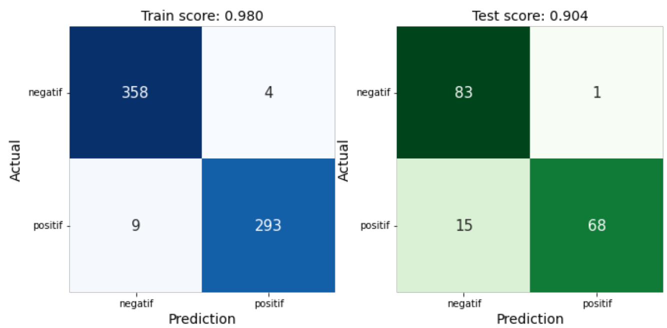

FIGURE 6. Confusion Matrix of Logistic Regression – TF-IDF

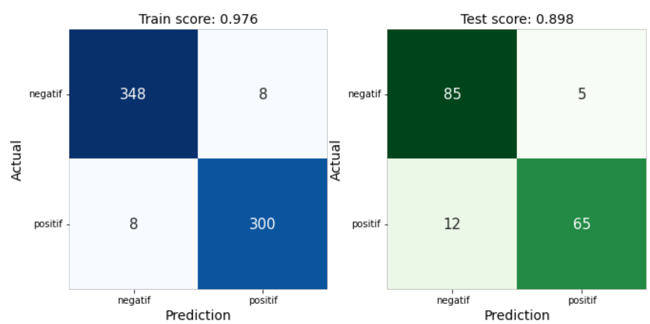

Figure 7. Confusion Matrix of Support Vector Machine – TF-IDF

This study employed evaluation metrics, namely accuracy, precision, and recall, to assess the performance of each algorithm in conducting sentiment analysis. The performance results are presented in Table 4 and visualized in Figure 8. The combination of Naïve Bayes with TF-IDF achieved an accuracy of 87.4%, a precision of 89.5%, a recall of 87.4%, and an F1-score of 88.4%. Logistic Regression with TF-IDF achieved an accuracy of 90.4%, a precision of 91.8%, a recall of 90.4%, and an F1-score of 90.9%. Meanwhile, SVM achieved an accuracy of 89.9%, a precision of 90.2%, a recall of 89.9%, and an F1-score of 89.9%. These results indicate that Logistic Regression with TF-IDF feature extraction is capable of effectively classifying sentiment analysis of product reviews in the Indonesian language. Furthermore, Logistic Regression outperforms the other methods evaluated in this study.

TABLE 4. Model Performance Comparison

Method	Performa			
	Accurasy	Presisi	Recall	F1-Measure
Naïve Bayes – TF-IDF	0,874	0,895	0,874	0,884
Logistic Regression – TF-IDF	**0,904**	**0,918**	**0,904**	**0,909**
Support Vector Machine – TF-DF	0,898	0,902	0,898	0,899

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

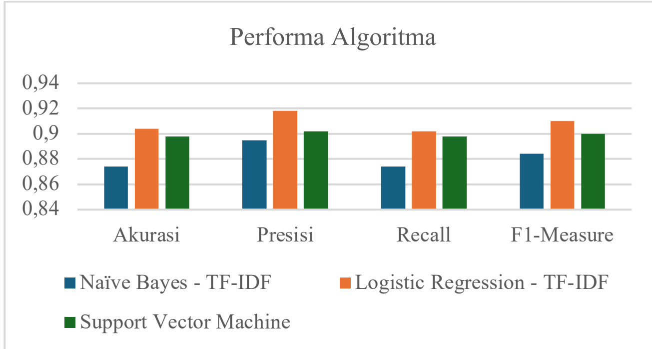

Figure 8. Model Performance

Discussion

In this study, the TF-IDF feature extraction approach was used to represent Indonesian-language marketplace review data into numerical vectors that can be processed by machine learning algorithms. TF-IDF was chosen because of its ability to assign relevant weights to important and frequently occurring words in a document while reducing the influence of common words that appear across the entire corpus.

TF-IDF is a statistical text representation method that works by calculating the relative frequency of a word within a document compared to the entire corpus. In the context of sentiment analysis, TF-IDF is effective in capturing explicit opinion words such as "good," "bad," "cheap," or "disappointing" that have strong associations with sentiment polarity.

However, TF-IDF has limitations in capturing deeper semantic meaning or context, particularly when sentiments are expressed indirectly or through sarcasm, metaphors, or complex sentence structures—common in Indonesian reviews, especially in informal or colloquial writing styles. The richness of the Indonesian language, with its wide vocabulary variations, non-standard words, and flexible sentence structures, poses unique challenges for TF-IDF.

While TF-IDF is fairly accurate in capturing explicit sentiments, it remains less optimal in understanding contextual and nuanced meanings compared to semantic representation approaches such as word embeddings or contextual embeddings (e.g., BERT). Nevertheless, for short reviews with limited vocabulary and direct opinions, which are common on marketplace platforms, TF-IDF remains an efficient and competitive choice, particularly when combined with algorithms that can effectively handle linear relationships among features, such as Logistic Regression.

The results of this study show that Logistic Regression achieved the best performance among the three algorithms. Logistic Regression consistently outperformed others in terms of accuracy and F1-score, especially in handling complex and non-standard sentences frequently found in Indonesian reviews. This model was more stable in capturing linear relationships among TF-IDF features that are relevant to sentiment labels.

On the other hand, the Naïve Bayes algorithm tended to experience accuracy degradation when dealing with words that carry multiple meanings or implicit emotional context. This is due to the independence assumption between features in Naïve Bayes, which does not always align well with informal Indonesian text data. Although SVM demonstrated competitive performance, it required more complex parameter tuning and, in the context of this dataset, performed slightly worse than Logistic Regression, particularly in detecting ambiguous word classes.

Overall, the findings indicate that the combination of TF-IDF and Logistic Regression is the most effective approach for sentiment analysis of Indonesian-language marketplace reviews, at least within the context and characteristics of the dataset used in this study.

CONCLUSION

Based on the analysis and discussion presented, it can be concluded that the Logistic Regression method in this study demonstrated strong performance when implemented for sentiment analysis of marketplace product reviews in the Indonesian language. With an accuracy of 90.4%, Logistic Regression with TF-IDF feature extraction outperformed Naïve Bayes and Support Vector Machine, which achieved accuracies of 87.4% and 89.8%, respectively.

For future research, it is recommended to apply alternative methods or feature extraction techniques to further improve accuracy, such as combining TF-IDF with lexicon-based sentiment scoring. In addition, validation with other datasets is suggested to strengthen the findings. Further studies are also needed to investigate whether the algorithms can overcome the limitations of imbalanced data in embedding calculation and weight assignment to enhance the performance of NLP models.

REFERENCES

1. Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., . . . Ruder, S. (2022). One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. Computation and Language. doi:https://doi.org/10.48550/arXiv.2203.13357
2. Bhatia, S., Sharma, M., & Bhatia, K. K. (2017). Sentiment Analysis and Mining of Opinions. Bhatia, S., Sharma, M., & Bhatia, K. K. (2017). Sentiment AnalyInternet of Things and Big Data Analytics Toward Next-Generation Intelligence, 503-523.
3. Cahyo, D. N., Farasalsabila, F., Lestari, V. B., Hanafi, Lestari, T., & Al Islami, F. R. (2023). Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method. Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, 8(2), 90-95.
4. Fauzi, M. A. (2019). Word2Vec model for sentiment analysis of product reviews in Indonesian language. International Journal of Electrical and Computer Engineering (IJECE), 9 No.1, 525-530.
5. Hariguna, T., Baihaqi, W. M., & Nurwanti, A. (2019). Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes Classifier Algorithm. International Journal of Informatics and Information Systems, 2 No.2, 48-55.
6. Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. CIRP Journal of Manufacturing Science and Technology, 23, 128-144.
7. Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2018). Sentiment Analysis on Product Reviews Using Machine Learning Techniques. Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, 768, 639-647.
8. Kanna, P. R., & Pandiaraja, P. (2019). An Efficient Sentiment Analysis Approach for Product Review using Turney Algorithm. INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019. Karur, India.
9. Kumar, S., Yadava, M., & Roy, P. P. (2019). Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Information Fusion, 52, 41-52.
10. Onan, A. (2020). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurrency Computation Practice and Experience.
11. Prihatini, P. M., Indah, K. A., Sukerti, G. N., Indrayana, I., & Sudiartha, I. (2023). Feature extraction performance on classified methods for text sentiment analysis. n Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science (iCAST-ES 2021), (pp. 1235-1243). doi:10.5220/0010962900003260
12. Sarowar, M. G., Rahman, M., Ali, M. N., & Rakib, O. F. (2019). An Automated Machine Learning Approach for Sentiment Classification of Bengali E-Commerce Sites. 2019 5th International Conference for Convergence in Technology (I2CT) . Pune, India.
13. Shafin, M. A., Hasan, M. M., Alam, M. R., Mithu, M. A., Nur, A. U., & Faruk, M. O. (2020). Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language. International Conference on Computer and Information Technology (ICCIT). Dhaka, Bangladesh.
14. Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021). Sentiments Detection for Amazon Product Review. 2021 International Conference on Computer Communication and Informatics (ICCCI -2021). Coimbatore, India.
15. Yang, L., Li, Y., Wang, J., & Siherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. IEEE Access, 23526.
16. Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Information Processing and Management.