Sandy Ananda Dwi[1]
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia, Yogyakarta
email: student.sandy23@mti.utdi.ac.id

Danny Kriestanto
Informatics
Faculty of Information Technology
Universitas Teknologi Digital Indonesia, Yogyakarta
email: danny@utdi.ac.id

Ajie Al Qadri Anwar
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
SMKIT Ibnul Qayyim Makassar
email: students.rajie_anwar24@mti.utdi.ac.id

Syahrur Ro'uf
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
email: students.syahrur_rouf24@mti.utdi.ac.id

Lintang Suci Rochmana
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia
email: students.lintang_sr24@mti.utdi.ac.id

Muhammad Agung Nugroho
Informatics Engineering Study Program
Telkom University, Purwokerto Campus
Central Java,Indonesia
email: magungnugroho@telkomuniversity.ac.id

Performance Analysis of Logistic Regression Algorithm in Opinion Segmentation of INDOSAT Network Service Reviews

In the era of the industrial revolution 4.0, where the use of network services has become a basic need and cannot be separated from daily activities, the massive number of network service users can be proven by the increasing number of people using digital platforms to search for information, express opinions or even just to communicate with each other, currently network services are available in the form of digital platforms that can be used to purchase network data packages or just to monitor the quality of network services, therefore this study aims to analyze user sentiment towards network services that have been launched by the INDOSAT provider based on the results of user reviews sourced from the digital platform using a machine learning approach and a logistic regression algorithm model to determine the segmentation of opinions that are widely expressed on the digital platform. The results of this study indicate that the logistic regression algorithm is able to analyze patterns of consumer characteristics with good accuracy in the algorithm model, and the results of the accuracy of the algorithm model in finding segmentation patterns in sentiment opinions reach an accuracy value of 85%, precision 81%, recall 77% and f1-score 79% to predict an opinion that has negative and positive sentiment during testing, then network speed, connection disruption and network data package prices are one of the factors that can influence an opinion regarding negative and positive sentiment.

***KeyWords**: Network; Sentiment Analysis; Machine Learning; Logisctic Regression*

This Article was:
submited: 20-06-25
accepted: 9-07-24
publish on: 20-07-25

How to Cite:
S. A. Dwi, et al, "Performance Analysis of Logistic Regression Algorithm in Opinion Segmentation of INDOSAT Network Service Reviews", Journal of Intelligent Software Systems, Vol.4, No.1, 2025, pp.16–21, 10.26798/jiss.v4i1.2004

1 Introduction

In the era of the industrial revolution 4.0, the use of network services has become a basic need in daily activities, from the massive number of network service users can be proven by the increasing penetration of people in using digital platforms to search for information, express opinions, communication facilities, or the like. In the midst of the increasing number of network users today, the quality of network access speed, connection disruptions and network data package costs have become serious concerns, based on data quoted from the Speedtest Global Index shows that internet speed per 2024, Indonesia has ranked 117th for broadband networks and 87th for cellular networks [1]. From this ranking, of course, many customers complain that the network they use tends to be slow, with these complaints, many customers express their expressions through the MyIM3 digital platform on the Playstore, so that from these customer expressions, if the data is processed further, it will reflect an opinion and sentiment towards the quality of service provided by the provider [2],[3],[4], therefore it is important to conduct opinion segmentation in order to understand customer perceptions of the service in more depth by using data sources obtained from the reviews that are specifically expressed by INDOSAT network customers. Of course, with this segmentation pattern, it can be used as strategic input for the company to

[1]Corresponding Author.

continue to improve the quality of service for customers. However, to analyze customer data, a dynamic approach is needed by utilizing the artificial intelligence paradigm to analyze opinions on reviews expressed by customers in assessing the quality of network services provided by the provider [5].

In recent years, the scientific field of sentiment analysis has experienced quite significant development and has become one of the methods widely used to process or explore text-based data. This method is an Natural Language Processing(NLP) technique that has the function of assessing the emotional tendencies of an opinion, whether in the form of a positive, negative, or neutral opinion towards an entity or service [6]. As the penetration of networks for the use of digital media continues to grow, the volume of available customer opinions will be very large, so that by considering these aspects, a dynamic and systematic approach is required that is able to classify opinions efficiently and accurately. As a solution to this approach, the approach of the machine learning method in the current era is widely relied on to analyze sentiment data and then classify it in the form of opinion segmentation with the help of an algorithm model to offer better flexibility in recognizing diverse and characteristic natural language patterns [7][8][9]. So in this context, the use of classification algorithm models such as logistic regression can be the right choice because of its simplicity in interpretation and its effectiveness in handling binary text data (positive or negative). Several studies have shown that logistic regression has quite competitive performance compared to other algorithms, especially in short text-based classifications such as customer reviews on digital platforms or others [10][11][12].

As methods in the field of sentiment analysis develop, there are several previous studies that have been conducted to understand public perception of a service or product, one of which was conducted by Muhammad Yusuf Hidayatulloh et al [13] which discusses sentiment analysis towards the BMKG organization using the Naïve Bayes Adaboost and SVM PSO algorithm models to find patterns of informative and uninformative message predictions, the results of this study can be used to predict informative and uninformative information by the BMKG organization, The results of the comparison of the algorithm models used in this study showed that SVM has superior performance than the naïve Bayes algorithm model. Furthermore, research has been conducted by Rian Tineges et al [14] regarding the analysis of customer satisfaction levels towards Indihome network services using the SVM algorithm, the research has proven effective in conducting sentiment analysis on opinions given by customers with an accuracy of 87%, then the research that has been conducted by Vynska Amalia Permadi [15], regarding the classification of sentiment of restaurant visitors in Singapore using the Naïve Bayes algorithm, the results of this study proved effective in classifying sentiment towards visitors to restaurants even though the level of accuracy obtained reached 73.33%, further research conducted by Ibnu Afdal et al [16] regarding the application of the Random Forest algorithm for sentiment analysis on comments on the YouTube digital platform, the results of this study can be used to classify sentiment on YouTube comments about Islamophobia with an accuracy of 79%.

Although sentiment analysis has been widely implemented in various domains, research specifically to evaluate the use of logistic regression algorithm models in segmenting cellular network service review opinions is still relatively limited. Therefore, this study encourages and focuses on analyzing the performance of the logistic regression algorithm model in classifying customer opinions based on reviews obtained from the MyIM3 application digital platform on Playstore. It is hoped that the results of this study can provide new contributions to the development of artificial intelligence-based sentiment analysis, especially in the context of opinion segmentation, and can provide useful insights for network service providers in improving the quality of their services through the utilization of customer feedback so that future services can be provided more optimally and efficiently.

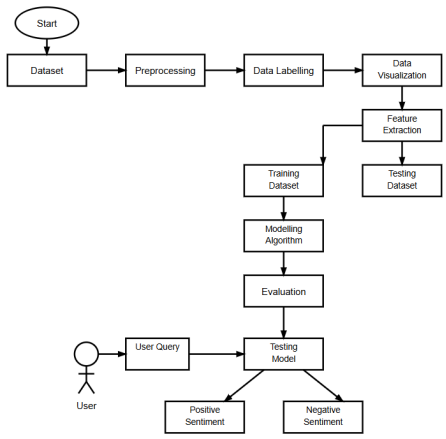

Fig. 1 Research Flow

Fig. 2 Text Preprocessing Pipeline

2 Method

This study was conducted to evaluate the performance of the logistic regression algorithm model in the application of sentiment analysis to user reviews of INDOSAT network services. The research stages include the process of collecting review datasets, preprocessing, labeling, visualization, feature extraction, data training, algorithm modeling, and model performance testing to measure the extent to which this approach is able to group customer opinions effectively. Research Flow can see on Figure 1.

2.1 Dataset. The dataset used in this study was obtained using the web scraping method on the MyIM3 application digital platform on Playstore, the number of datasets successfully collected had 192 thousand reviews expressed by customers with various sentiment labels in the form of free text, so that the results of this data are considered to reflect user opinions on Indosat network services. However, this raw data still contains various data quality problems such as missing values, data duplication, and inaccuracy of sentiment labels. Therefore, before entering the preprocessing stage, data cleaning and preparation are required first so that the dataset can be used for optimal model training.

2.2 Preprocessing. The preprocessing stage is carried out to ensure that the text data obtained in the form of customer reviews can be processed optimally by the machine learning algorithm model. This process aims to clean and prepare the text to be more representative and freer from interference that can affect the accuracy of the model. The preprocessing steps carried out are as follows on Figure 2:

(1) Cleaning Text: Cleaning irrelevant text with several steps, such as removing mentions such as (@username), hashtags, retweets, links, numbers, newline characters, and punctuation. In addition, characters such as newlines will be replaced with spaces, and excess spaces at the beginning and

end of the text will be removed to make the data structure neater.

(2) Case Folding: Converts all letters in a review into lower case to equalize word representation. For example, the words “Internet” and “internet” will be considered the same, thus reducing unnecessary word duplication.

(3) Tokenization: Breaking review sentences into individual words or dividing text data into lists of words grouped into word units (tokens), this is useful for parsing text data into basic text components.

(4) Stopword Removal: Removing affixes that do not have high information value in text data such as "yang", "dan", "di", these affixes will basically be removed because they do not make a significant contribution to the main meaning in the review.

(5) Stopword Removal: Removing affixes that do not have high information value in text data such as "yang", "dan", "di", these affixes will basically be removed because they do not make a significant contribution to the main meaning in the review.Stemming: Returning or reducing words in text data to their basic form, words that have undergone changes in form (for example changes due to affixes) will be returned to their basic form. For example, the words "walking", "running", will be returned to their basic forms such as "jalan", "lari". This aims to group words that have the same basic meaning.

2.3 Data Labeling. In the process of labeling sentiment data in this study, it is carried out automatically using a lexicon-based approach, where each word will be classified into positive and negative sentiment categories based on a sentiment data dictionary that has been standardized with the Indonesian language. This automatic approach is considered efficient in managing large data volumes and accelerating the preprocessing stage for model development so that the labels obtained will be used as targets for the classification of the model training process.

2.4 Data Visualization. Data visualization is done to gain a good understanding of the distribution and pattern of reviews based on sentiment labels. The visualization techniques used in this study include word clouds, histograms, and bar charts. Basically, these diagrams are used to display the frequency of occurrence of the most frequently occurring words in reviews, both positive and negative sentiments. This stage also helps in identifying dominant and relevant words to sentiment labels.

2.5 Data Training and Testing. The dataset used in this study was divided into two parts for the training and testing process with a composition of 80% as training data and 20% as testing data. The purpose of this division is basically to ensure that the machine learning model can be trained using sufficiently large and representative data, then tested on data that has never been used before. This aims to avoid overfitting and ensure that the evaluation of model performance is carried out objectively and realistically.

2.6 Modeling Algorithm. The algorithm model used in this study is logistic regression, which is one of the classification algorithms that is quite widely applied in sentiment analysis. This algorithm was chosen because of its ability to handle high-dimensional text data and produce probabilistic predictions. The model is trained using TF-IDF (Term Frequency-Inverse Document Frequency) based text feature representation that allows the algorithm to recognize the importance of words in the context of the document and the entire corpus [17].

2.6.1 Logisctic Regression. Logistic regression is one of the classification algorithms commonly used in machine learning and data analysis, particularly for binary classification tasks. This algorithm models the relationship between one or more independent variables (features) and a binary dependent variable (target class), whose values are limited to two possibilities, such as positive or negative.

In the context of sentiment analysis, logistic regression is especially suitable due to its ability to handle high-dimensional text data and produce clear probabilistic interpretations of the target class. The model typically employs a text feature representation using TF-IDF (Term Frequency–Inverse Document Frequency), which quantifies the importance of a word in a document relative to a collection of documents.

Logistic regression uses the logistic function (also known as the sigmoid function) to convert the linear output into a probability value between 0 and 1 [18]. The sigmoid function is defined in Equation (1):

$$P = \frac{1}{1 + e^{-z}} \tag{1}$$

Where:

e^{-z}	=	$\beta 0 + \beta 1x1 + \beta 2x2 + \cdots + \beta nxn$
xi		is the 1st feature.
βi		is the model coefficient (weight) learned from the training data

The output value of the sigmoid function, namely $y^{\wedge} = P$ represents the probability that the data belongs to the target class (eg: positive sentiment). Class prediction is done by setting a certain threshold, for example:

$$Class = \begin{bmatrix} 1, & IF\ y^i > 0.5 \\ 0, & IF\ y^i < 0.5 \end{bmatrix} \tag{2}$$

The parameters of the Logistic Regression model are estimated using the Maximum Likelihood Estimation (MLE) approach, which aims to maximize the likelihood function of the training data so that the model can make predictions as accurately as possible.

2.7 Evaluation. Evaluation of the model is carried out using several performance measurement metrics, namely: accuracy, precision, recall, and F1-score. These four metrics are used to measure the extent to which the model is able to classify customer opinions into sentiment categories accurately. This evaluation is very important to assess the effectiveness of the model in the context of accurate and relevant opinion classification for research objectives.

3 Results

Based on the sentiment analysis results graph that has been segmented against the data of INDOSAT network service user reviews, the sentiment categories are divided into two large parts, namely negative sentiment and positive sentiment. The data visualization results are presented in the form of a pie chart to show that the majority of user reviews contain negative sentiment. From the results above, 65.9% of the total analyzed review data were mostly identified as negative sentiment. This certainly shows that most customers convey opinions and narratives about dissatisfaction with the network services provided by Indosat. Meanwhile, the remaining 34.1% were identified as positive, which shows that some customers provide satisfactory opinions and narratives regarding the services provided by Indosat. Based on the findings from the analysis process which includes the data preprocessing stage, sentiment labeling using lexicon, and the application of a logistic regression classification model with TF-IDF-based feature representation. States that the proportion of negative sentiment indicates potential problems in the quality of network services felt by customers. Thus, based on the distribution of sentiment obtained, it can be concluded that the majority of public opinion on INDOSAT network services taken from the MyIM3 digital application platform tends to show negative perceptions. These findings can be used as a basis for evaluation by service providers to identify and

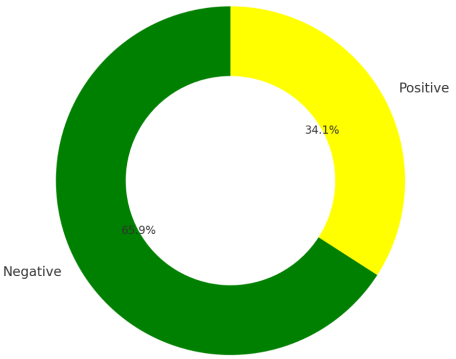

Fig. 3 Sentiment Analysis Result Categories

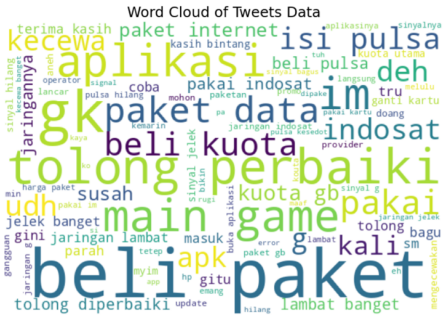

Fig. 4 Word Cloud Analysis Results Against User Reviews

Fig. 6 Word Cloud Analysis Results on Positive Sentiment

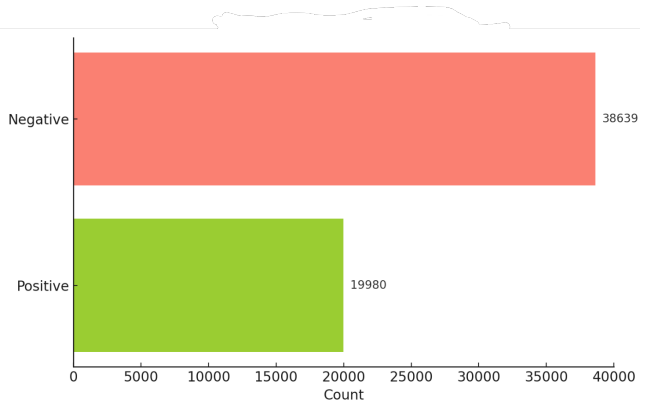

Fig. 7 (a) class distribution

improve aspects of services that are considered less satisfactory by users.

Based on the results of user reviews that have been presented in the form of wordcloud graphics, it has shown the words that most often appear in user reviews of Indosat services. Opinions such as "paket", "beli", "kuota", "data", and "pulsa" dominate and appear the most in the wordcloud, and the results indicate user focus on purchasing and using data services. In addition, words such as "aplikasi", "main", and "game" also appear, which describe user activities that depend on a stable internet connection. Several negative words such as "tolong", "perbaiki", "jelek", "lambat", and "kecewa" indicate complaints about network quality. This visualization supports the results of the previous sentiment analysis which showed that the majority of users gave negative sentiments towards Indosat services, and provides context regarding the main issues complained about.

The results of user reviews that have been presented in the form of negative wordcloud graphs have shown that most complaints from Indosat network service customers are related to network quality, especially in dominant opinions such as "sinyal", "lambat", "jaringan", and "paket". The words "beli", "pakai", and phrases

Fig. 5 Word Cloud Analysis Results on Negative Sentiment

such as "tolong perbaiki" reflect user requests for service improvements. The emergence of the words "gk", "kecewa", and "habis" indicate user dissatisfaction, both in terms of connection, quota, and service price. Words such as "harga paket" and "paket mahal" also highlight complaints about costs that are not commensurate with quality. Therefore, the graph presented in the form of a wordcloud emphasizes that network and price problems are the main focus of negative sentiment. The results of user reviews that have been presented in the form of positive wordcloud graphs show that words such as "indosat", "paket", "sinyal", "jaringan", and "beli" dominate opinions. Although categorized as positive, words such as "lambat", "kecewa", and "tolong perbaiki" still appear, indicating that there is hope for improvement from customers regarding the quality of Indosat's network services. However, the words "terimakasih", "promo", and "bagus" are also present, reflecting a good experience with the service, especially regarding price and promotions. Overall, this positive sentiment shows user recognition of the service, although it is still accompanied by the hope of improving network and signal quality.

The results of the segmentation analysis show three main analysis points presented in graphical form. First, the distribution of sentiment shows that basically negative reviews are expressed more than positive reviews. Second, the length of the majority of review texts is in the short range, which is between 5 and 20 words. And finally, the words that appear most often in the reviews are "jaringan", "sinyal", "kuota", and "paket", of course from the results of the opinion analysis it can reflect that the main concern of users is focused on the quality of network services.

Based on the evaluation results of the performance of the logistic regression algorithm model, the model is able to understand good sentiment analysis patterns in this study, but to assess the performance of the algorithm model used, of course an evaluation must be carried out based on several measurement metrics, including accuracy, precision, recall, and F1-score. Evaluation with these metrics is applied to training data and testing data to ensure stability and the ability to generalize the model in classifying data

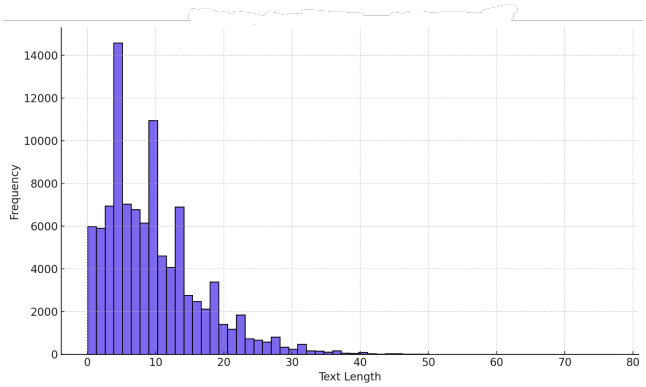

Fig. 8 (b) text length distribution

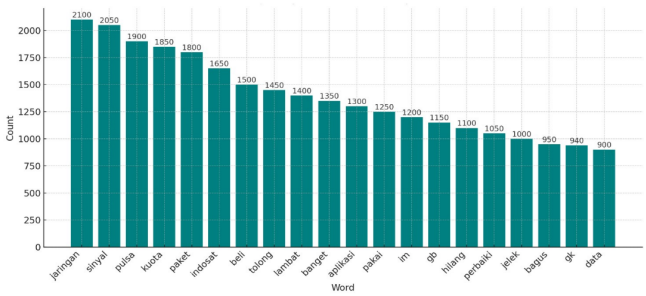

Fig. 9 (c) frequency of dominant words

in a consistent and stable manner, the following are the evaluation results and metrics of the logistic regression algorithm model presented in the Table 1.

The evaluation results of the logistic regression algorithm model showed good performance on the training data (Table 2), with an accuracy of 86.1%. Precision was recorded at 81.3%, which means that the proportion of correct positive predictions is quite high. The recall value of 78.7% indicates that the model is able to capture most of the actual positive data. The F1-Score of 80% reflects a fairly good balance between precision and recall during the training process.

However, during testing, accuracy, precision, recall and f1-score experienced a decrease in percentage value, but the logistic regression algorithm model still showed stable performance. After testing the accuracy data of the algorithm model for accuracy was recorded at 85.6%, with precision 81.8% and recall 77.3%. The resulting F1-Score was 79.5%, with the conditions of table 2 showing that the performance of the logistic regression algorithm model did not decrease drastically when compared to the training data. With this, the model has good generalization capabilities for data that has never been seen before.

The results of the algorithm model evaluation metrics are formatted in graphical form, based on Figure 6 that the performance of the logistic regression algorithm model on training and testing data. The accuracy results on both are almost identical and consistent, namely 0.86, these findings indicate that the model is stable and not overfitting. Precision increases slightly on testing data (0.82) compared to training (0.81), indicating that the prediction accuracy remains consistent. The recall value decreases slightly from 0.79 (train) to 0.77 (test), which is still within reasonable limits. The F1-score remains stable at 0.80, indicating that the balance between precision and recall is maintained in both scenarios.

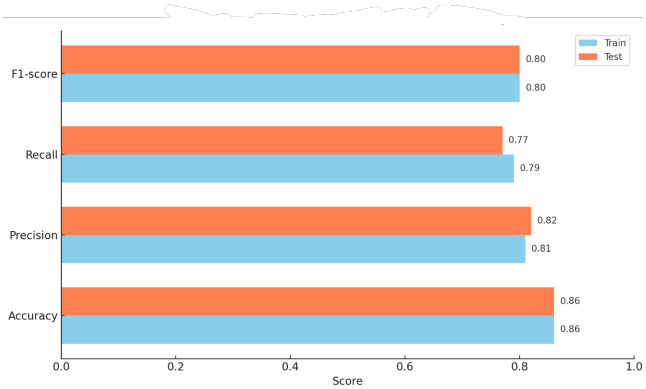

Fig. 10 Performance of Logistic Regression Algorithm Model

4 Discussion

The results of the analysis show that the logistic regression algorithm model is able to classify customer opinion data effectively. This can be seen from the measurement of evaluation metrics during the training stage, with an accuracy value of 0.861, precision 0.813, recall 0.787, and F1-score 0.800. Meanwhile, during the testing stage, the model recorded an accuracy value of 0.856, precision 0.818, recall 0.773, and F1-score 0.795.

The metric values presented in both tables tend to be relatively stable and consistent between training and testing data, indicating that the model used has good generalization capabilities and does not experience significant overfitting.

When compared to previous studies using similar models, these results show competitive levels of accuracy and F1-score. One of the strengths of this study is the consistency of model performance on training and testing data, which strengthens the reliability of the classification. However, there are limitations in the imbalanced distribution of data classes, with a larger number of negative sentiment data (38,639) than positive sentiment (19,980), as well as the dominance of certain words such as "network", "signal", and "quota", which can affect model bias. In addition, recall performance on the testing data decreased slightly, which could indicate that the model is less than optimal in recognizing all minor classes.

Overall, this study successfully proves that the algorithm model of logistic regression can be used to classify and segment customer opinions on network service quality with quite good results. This study contributes to the application of machine learning, especially for short text-based sentiment analysis.

5 Conclusion

This study evaluates the performance of the logistic regression algorithm model in classifying and segmenting user opinions on Indosat network services. The results of this study indicate that the performance of the model is stable and accurate, both on training and testing data, so this model is considered reliable for simple text classification tasks. The findings show that data distribution and feature representation greatly affect the prediction results. Although the model works well, challenges such as class imbalance remain. This study is an important foundation for the development of user opinion classification systems in the future. The Logistic Regression algorithm model has proven to be effective and efficient. It is hoped that further research can be focused on improving data quality and features for more optimal classification results and trying to explore other models such as Random Forest or SVM, applying data balancing techniques, and utilizing word embedding or deep learning models to improve feature representation and classification accuracy.

Table 1 Model Evaluation Results and Training Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	0.861	0.813	0.787	0.800

Table 2 Model Evaluation Results and Training Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	0.856	0.818	0.773	0.795

References

[1] Speedtest, 2025, "Speedtest Global Index Internet Speed Around The World," https://www.speedtest.net/global-index, Accessed: May 15, 2025.

[2] Putri, L. E. and Humaira, M. A., 2024, "Pengaruh Ulasan Produk Terhadap Keputusan Pembelian Konsumen," Karimah Tauhid, **3**(1), pp. 694–702.

[3] Riska, C. F., Hendayana, Y., and Wijayanti, M., 2023, "PENGARUH ULASAN KONSUMEN, KUALITAS PRODUK DAN MARKETING INFLUENCER TERHADAP KEPUTUSAN PEMBELIAN PRODUK ERIGO," Junal Econ., **2**(I), pp. 1–19.

[4] Andriyani, W., Astuti, Y., and Wisesa, B. A., 2024, "Analisis Sentimen pada Ulasan Produk dengan SVM dan Word2Vec," JIKO, (1), pp. 173–185.

[5] Malviya, S., Tiwari, A. K., Srivastava, R., and Tiwari, V. K., 2020, "Machine Learning Techniques for Sentiment Analysis: A Review," SAMRIDDHI A J. Phys. Sci. Eng. Technol., **12**(2), pp. 72–78.

[6] Aryanti, P. A. N. and Mahendra, I. B. M., 2023, "Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine," J. Elektron. Ilmu Komput. Udayana, **12**(1), pp. 2654–5101.

[7] Budianita, E., Cynthia, E. P., Pranata, A., and Abimanyu, D., 2022, "Pendekatan berbasis Machine Learning dan Leksikal Pada Analisis Sentimen," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, pp. 99–104, Online, https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/19137

[8] Kusuma, G. H., Permana, I., Salisah, F. N., Afdal, M., Jazman, M., and Marsal, A., 2023, "Pendekatan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kendaraan Listrik Pada Sosial Media X," JUSIFO (Jurnal Sist. Informasi), **9**(2), pp. 65–76.

[9] Safira, A. and Hasan, F. N., 2023, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," Zo. J. Sist. Inf., **5**(1), pp. 59–70.

[10] Sari, A. K., Irsyad, A., Aini, D. N., Islamiyah, and Ginting, S. E., 2024, "Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif," Adopsi Teknol. dan Sist. Inf., **3**(1), pp. 64–73.

[11] Jahan, I., Islam, M. N., Hasan, M. M., and Siddiky, M. R., 2024, "Comparative analysis of machine learning algorithms for sentiment classification in social media text," Libr. Prog. Int., **44**(3), pp. 8139–8145.

[12] Mohaiminul, M. and Sultana, N., 2018, "Comparative Study on Machine Learning Algorithms for Sentiment Classification," Int. J. Comput. Appl., **182**(21), pp. 1–7.

[13] Hidayatulloh, M. Y., Sunanto, A., Gevin, M. F. A., and Saputra, D. D., 2023, "Optimasi Sentimen Analisis Informatif dan Tidak Informatif dari Tweet di BMKG Menggunakan Algoritma Naive Bayes dan Metode Teknik Pengambilan Sampel Minoritas Sintetis," J. Sains Komput. Inform. (J-SAKTI), **7**(1), pp. 1–12.

[14] Tineges, R., Triayudi, A., and Sholihati, I. D., 2020, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," J. Media Inform. Budidarma, **4**(3), p. 650.

[15] Permadi, V. A., 2020, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," J. Buana Inform., **11**(2), pp. 141–151.

[16] Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., and Syafria, F., 2022, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," J. Nas. Komputasi dan Teknol. Inf., **5**(1), pp. 122–130, Online.

[17] Kholifah, B., Thoib, I., Sururi, N., and Kurnia, N. D., 2024, "Analisis Sentimen Warganet Terhadap Isu Layanan Transportasi Online Berbasis InSet Lexicon Menggunakan Logistic Regression," Kumpul. J. Ilmu Komput., **11**(1), pp. 14–25.

[18] Chen, Y., Liu, P., and Teo, C. P., 2020, "Regularised Text Logistic Regression: Key Word Detection and Sentiment Classification for Online Reviews," http://arxiv.org/abs/2009.04591, Online preprint.