Journal of Engineering Science and Technology
Vol. 21, No. 2 (2026) 649 - 661

SENTIMENT ANALYSIS WITH CLASSIFICATION LEARNING FOR COMMENTS ON THE X (TWITTER) APPLICATION

AGUS NURSIKUWAGUS[1,*], DESHINTA ARROVA DEWI[2],
HILMAN FAUJI ABDILLAH[3], RAYVANKA AZRIFANY NURVIANA NAJMI[3],
RAMADHANA RIZKI ZULKARNAEN[3]

[1]Fakultas Pascasarjana, Universitas Komputer Indonesia, Bandung, Indonesia
[2]Faculty of Data Science and Information Technology, INTI International University, Negeri Sembilan, Malaysia
[3]Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia, Bandung, Indonesia
*Corresponding Author: agusnursikuwagus@email.unikom.ac.id

Abstract

X (formerly Twitter) remains a central yet polarized platform for online interaction, particularly within Indonesia's growing digital landscape. To assess public perception, this study analysed 2,794 cleaned Indonesian reviews from the Google Play Store. Following lexicon-based labelling, data were vectorized using term frequency–inverse document frequency (TF-IDF) and Bag-of-Words. Logistic Regression and Random Forest classifiers were evaluated (80:20 split), utilizing the Synthetic Minority Over-sampling Technique (SMOTE) to mitigate a 24.39% class imbalance. Logistic Regression with TF-IDF and SMOTE achieved optimal performance, yielding 90.2% accuracy and a 92.2% F1-score (90.8% precision, 93.8% recall), significantly outperforming Random Forest (83.0% accuracy). The high recall specifically indicates effectiveness in correctly identifying relevant sentiment. Consequently, this study demonstrates that lightweight linear models combined with over-sampling provide a robust, computationally efficient framework for analysing Indonesian text, enabling developers to derive actionable insights for strategic product improvement and user retention.

Keywords: Classification, Performance, Product innovation, Sentiment analysis, Word cloud.

1. Introduction

With over 600 million users, X (formerly Twitter) is a primary driver of global discourse, yet its influence often triggers polarized public perceptions regarding misinformation and policy changes [1]. Consequently, objectively understanding user sentiment is essential. This study employs classification learning to analyse large-scale user comments, providing a systematic approach to identifying how public perceptions are formed on the platform [2].

Machine Learning compared Logistic Regression and Random Forest for sentiment analysis on Disdukcapil reviews [3, 4], finding that Logistic Regression achieved superior performance with 90% accuracy and an F1-score of 90% on a dataset of 18,810 samples. Addressing feature engineering, the demonstrating that combining Random Forest with TF-IDF outperformed other techniques like BoW and Word2Vector, reaching 99% accuracy on Amazon reviews and 96% on Twitter data [5]. Finally, to resolve performance degradation caused by imbalanced classes, Suandi et al. [6] utilized SMOTE, which significantly improved Random Forest accuracy from 79% to 93%.

This study analyses public perception of X using Google Play Store reviews by comparing Random Forest (ensemble-based) and Logistic Regression (statistical) models. These algorithms were evaluated using TF-IDF and BoW feature extraction across two scenarios: an original dataset and one balanced via SMOTE over-sampling. Key Findings of the research:

(i) Comprehensive text pre-processing significantly enhances machine learning prediction accuracy.

(ii) The combination of Logistic Regression, TF-IDF, and SMOTE yields the highest accuracy for sentiment analysis on this dataset.

The structure of writing in this study consists of Chapter 1, Background on X, literature review, and research objectives. Chapter 2, Data collection, pre-processing, modelling, and evaluation flow. Chapter 3, Outputs from each stage of the research methodology. Chapter 4, In-depth analysis and interpretation of the experimental data. Finally, Chapter 5, Summary of the optimal methods and final research insights.

2. Literature Review

2.1. Data review

Google Play Store reviews offer a real-time, representative look into user perception, though platforms like X (Twitter) introduce complexities such as sarcasm, nonstandard language, and temporal bias [7, 8]. Researchers typically navigate these challenges using lexicon-based, rating-based, or manual annotation labelling approaches, each offering a distinct trade-off between scalability and accuracy [7, 9]. To maintain reliability when using lexicon-based methods, integrating cross-validation through error analysis or manual subsets is essential for validating findings.

2.2. Data preprocessing

Standard Indonesian preprocessing employs a multistage pipeline including case folding, noise cleaning, and slang normalization to transform raw text into a structured format, while researchers find that n-gram and frequency-tuned TF-IDF provides more informative weights than BoW for short, sparse texts [5, 10-15]. Despite the high performance of modern transformers, their heavy computational requirements and sensitivity to domain shift make TF-IDF and BoW the preferred, interpretable baselines for many sentiment studies [16].

2.3. Modelling

In text processing, Logistic Regression and Naive Bayes/SVM serve as robust baselines, whereas Random Forest excels at capturing nonlinear patterns but remains sensitive to noise [17]. To address common class imbalances and ensure rigorous comparison, techniques like SMOTE must be applied strictly to training data to prevent leakage, while performance is validated through stratified k-fold cross-validation and metrics such as Accuracy, F1-score, and McNemar tests [6, 18]. Consequently, this study addresses a gap in the Indonesian-language corpus for Platform X by systematically comparing logistic regression (LR) and random forest (RF) across TF-IDF and BoW features, evaluating the impact of SMOTE resampling on model effectiveness.

3. Method

This study is conducted in a series of structured stages, from initial data collection to final user perception analysis, as shown in Fig. 1, to ensure a consistent and rigorous research flow.

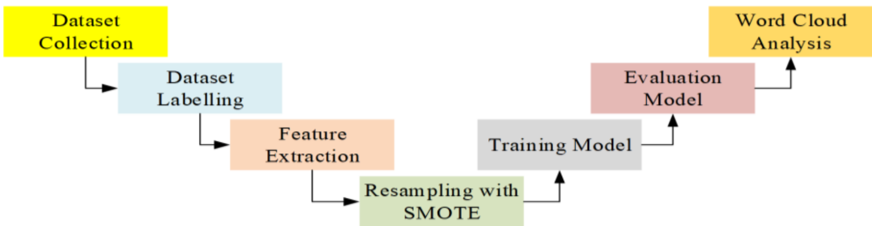

Fig. 1. Method flowchart.

3.1. Data collection

This study uses user comment data from Google Play obtained through data scraping using the Python programming language in the Google Colab environment [19]. The data retrieval process was carried out using the google-play-scraper library [20]. The scraping process was limited to 3,000 comments, which were stored in a pandas DataFrame and saved to a file in Comma Separated Values (CSV) format for further analysis [3].

3.2. Dataset preprocessing

To mitigate noise and structure the raw dataset for machine learning, a multi-stage pre-processing workflow was implemented, finalizing the dataset at 2,794 comments after removing duplicates and missing records [8]. The specific stages included Case Folding, noise cleaning, and word normalization using a specialized GitHub dictionary (ramaprakoso), followed by Tokenization, Stopword Removal, and Indonesian-optimized Stemming via the Sastrawi library to reduce words to their root forms [10-14, 21, 22].

3.3. Data labelling

Pre-processed comments were categorized into positive (1,739) and negative (1,057) sentiments using a lexicon-based approach [7] that weighted tokens based on an Indonesian dictionary of 3,274 positive and 6,609 negative words. Following the methodology of angelmetanosaa, comments were assigned a negative label if their total weighted score was below zero, while those with zero or positive scores were labelled positive [7].

3.4. Feature extraction

The feature extraction stage utilizes TF-IDF and bag-of-words (BoW) to reduce dimensionality by retaining only the most significant data aspects [5, 23]. The TF formula is given in Eq. (1), IDF in Eq. (2), TF-IDF in Eq. (3):

$$tf_{t,d} = \frac{Number\ of\ occurrences\ of\ term\ (t)\ in\ document\ (d)}{Total\ number\ of\ term\ in\ document\ (d)} \tag{1}$$

$$idf_t = log\frac{Total\ number\ of\ documents\ in\ the\ corpus}{Number\ of\ document\ containing\ term\ (t)} \tag{2}$$

$$tfidf_{t,d} = tf_{t,d}\ x\ idf_t \tag{3}$$

$tf_{t,d}$: Term Frequency (frequency weight) value of term/word (t) in document (d). idf_t: the weight value for a term (t) or a specific word. $tfidf_{t,d}$: final weight of term (t) in document (d). Bag-of-Words, this extraction technique is easy to interpret, where words are converted into numerical vectors [24].

3.5. Resampling

To address the imbalance of 1,739 positive and 1,057 negative classes, this study utilizes SMOTE to generate synthetic minority class data and eliminate potential modelling bias [25].

3.6. Modelling

Once the data is pre-processed, it enters the modelling phase where logistic regression (LR) and Random Forest algorithms are applied for classification. Logistic Regression, a discriminative model, predicts class probabilities by calculating weighted feature sums using the mathematical frameworks established in Eq. (4) and Eq. (5) [3, 26].

$$P(Y = 1) = \frac{1}{1+e^{2-(\beta 0+\beta 1x)}} \tag{4}$$

$$P(Y = 0) = 1 - P(Y = 1) \tag{5}$$

In Logistic Regression, the model calculates the probability of a positive class P(Y=1) or negative class P(Y=0) using Euler's number (e), a bias term (β), and coefficients (β) that measure the effect of independent variables. Conversely, **Random Forest** functions as an ensemble learning method that aggregates multiple decision trees and utilizes a majority vote to deliver stable, robust predictions while minimizing overfitting [3, 26].

3.7. Evaluation

Model performance, Table 1, is assessed using an NxN confusion matrix, which provides a detailed breakdown of accurate and misclassified predictions by comparing predicted classes against ground-truth labels [27].

Table 1. Confusion matrix.

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	True Positive (TP)	False Positive (Type I Error)
	Negative	False Negative (Type II Error)	True Negative (TN)

Accuracy is defined as the ratio of correctly predicted instances to the total number of predictions, as calculated in Eq. (6) [28]. Precision, recall, and the F1-score further refine this evaluation by measuring the reliability of positive classifications, the identification of actual positive cases, and the harmonic balance between both metrics through Eqs. (7), (8), and (9).

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{Total\ Prediction} \tag{6}$$

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{7}$$

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \tag{8}$$

$$F1\ Score = 2 * \frac{Prediction * Recall}{Prediction + Recall} \tag{9}$$

3.8. Word cloud analysis

Word cloud visualization highlights the most significant and frequent terms by displaying them in larger sizes to represent specific sentiments [3].

4. Results and Discussion

Our experimental results, presented in alignment with the methodology, confirm that tuning parameters significantly influence prediction accuracy and related performance metrics.

4.1. Data collection

The dataset, visualized in Fig. 2, consists of 3,000 Indonesian-language entries scraped from Google Play Store comments using the google-play-scraper library in Python [19]. These results were exported as a CSV file for structured analysis and subsequently reloaded into a Pandas DataFrame for further processing [3].

4.2. Data Pre-processing

To ensure high-quality input for modelling, raw data from the Google Play Store was refined through a structured preprocessing workflow, resulting in a final dataset of 2,794 comments after removing duplicates and missing values [8]. This sequence included case folding, noise cleansing, slang normalization, tokenization, stop-word removal, and Indonesian-specific stemming to reduce words to their base forms [10-13, 21, 22].

	Review
0	Semakin banyak updatenya semakin buruk pula ca...
1	aplikasi jelek, update malah makin banyak bug ...
2	apk ini semakin lama semakin parah bug nya, sa...
3	udah berkali kali sering lag, aku mau login pu...
4	ini x nya gimana sih abis diupdate malah jadi ...
...	...
2995	Tidak bisa membuat album untuk koleksi foto. D...
2996	gak bisa login blokk
2997	Banyak bug, tidak bisa penelusuran di Twitter ...
2998	kenapa gak bisa buat akun woyyy?!?!?!
2999	Twitter awalnya suka bgt sampe menghabiskan wa...

3000 rows × 1 columns

Fig. 2. The text scrapping by Pandas library in Python.

4.3. Labelling

Each review was assigned a polarity label by summing word scores derived from a public GitHub lexicon, as detailed in Tables 2 and 3 [7]. Following this calculation, reviews were categorized as "positive" for scores greater than or equal to zero and "negative" for scores below zero [7].

Table 2. Text processing result by Scikit-Learn Python.

Review	Case Folding	Cleansing	Word Normalization	Tokenizing	Stopword Removal	Stemming
Semakin banyak updatenya semakin buruk pula ca...	*semakin banyak updatenya semakin buruk pula ca...*	*semakin banyak updatenya semakin buruk pula ca...*	*semakin banyak updatenya semakin buruk pula ca...*	*[semakin, banyak, updatenya, semakin, buruk, p...*	*[updatenya, buruk, berjalannya, aplikasi, asik...*	*updatenya buruk jalan aplikasi asik scroll tib...*
aplikasi jelek, update malah makin banyak bug ...	*aplikasi jelek, update malah makin banyak bug ...*	*aplikasi jelek update malah makin banyak bug n...*	*aplikasi jelek update bahkan makin banyak bug ...*	*[aplikasi, jelek, update, bahkan, makin, banya...*	*[aplikasi, jelek, update, bug, suara, video, d...*	*aplikasi jelek update bug suara video delay da...*
kenapa gak bisa buat akun woyyy?!?!?!	*kenapa gak bisa buat akun woyyy?!?!?!*	*kenapa gak bisa buat akun woyy*	*kenapa tidak bisa buat akun woyy*	*[kenapa, tidak, bisa, buat, akun, woyy]*	*[akun, woyy]*	*akun woyy*
Twitter awalnya suka bgt sampe menghabiskan wa...	*twitter awalnya suka bgt sampe menghabiskan wa...*	*twitter awalnya suka bgt sampe menghabiskan wa...*	*twitter awalnya suka banget sampai menghabiska...*	*[twitter, awalnya, suka, banget, sampai, mengh...*	*[twitter, suka, banget, menghabiskan, dsni, be...*	*twitter suka banget habis dsni hubung bijak ba...*

~~**Table 2 (Continued). Text processing result b**~~

Table 3. Labelled dataset by lexicon-based approach.

Review	Score	Polarity
updatenya buruk jalan aplikasi asik scroll tib...	-8	negative
aplikasi jelek update bug suara video delay da...	-11	negative
akun woyy	0	positive
….	….	….
twitter suka banget habis dsni hubung bijak ba...	13	positive

4.4. Feature extraction

The feature extraction stage converts text into numerical representations using TF-IDF and BoW to facilitate machine learning processing. Using formulas (1), (2), and (3) via the TfidfVectorizer function with specific tuning parameters, the TF-IDF process yielded 746 features by assigning higher weights to words that are uniquely frequent within individual comments compared to the overall dataset [5,24]. Utilizing the CountVectorizer function with parameters for n-grams and frequency thresholds, this process generated a feature space of 1,397 unique columns (see Figs. 3 and 4).

	acak	accord	account	admin	aduh	ajar	aju	akhirakhir	akses	aktif	...	vid	video
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.279768
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2789	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
2790	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
2791	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.175867
2792	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000
2793	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000

2794 rows × 746 columns

Fig. 3. Transformation resulted from word extraction by TF-IDF.

	acak	accord	account	admin	aduh	ajar	aju	aju banding	akhirakhir	akses	...	wawas
0	0	0	0	0	0	0	0	0	0	0	...	0
1	0	0	0	0	0	0	0	0	0	0	...	0
2	0	0	0	0	0	0	0	0	0	0	...	0
3	0	0	0	0	0	0	0	0	0	0	...	0
4	0	0	0	0	0	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
2789	0	0	0	0	0	0	0	0	0	0	...	0
2790	0	0	0	0	0	0	0	0	0	0	...	0
2791	0	0	0	0	0	0	0	0	0	0	...	0
2792	0	0	0	0	0	0	0	0	0	0	...	0
2793	0	0	0	0	0	0	0	0	0	0	...	0

2794 rows × 1397 columns

Fig. 4. Transformation resulted from word extraction by BoW.

4.5. Resampling data by SMOTE

The dataset labelling process revealed a 24.39% imbalance between classes, potentially introducing model bias. Consequently, the data was partitioned into an To address class imbalance, the dataset was resampled using the SMOTE technique to generate 543 synthetic samples for the minority class [25]. By applying this method to both TF-IDF and BoW training sets, the distribution was equalized to 1,389 data points for both positive and negative classes, ensuring a balanced input for the modelling phase.

4.6. Modelling

To evaluate the most effective approach, model training was conducted by comparing datasets with and without SMOTE across both TF-IDF and BoW feature extractions. Logistic Regression was configured with specific regularization parameters (C=4, L1 penalty) to predict class probabilities [3], while the Random Forest ensemble utilized 500 decision trees to achieve robust classification [26]. 80:20 split, yielding 1,389 training and 559 testing data points as illustrated in Figs. 5 and 6.

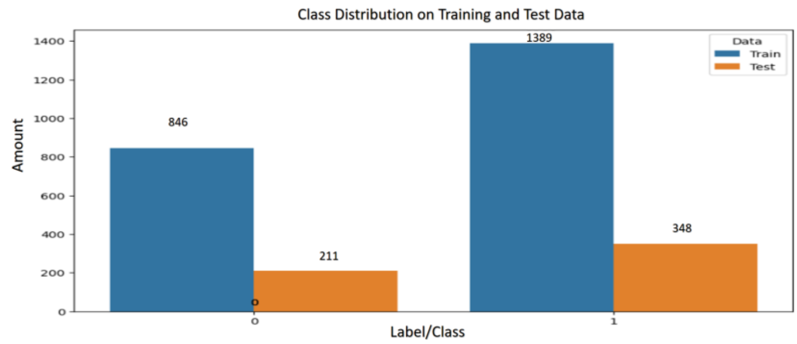

Fig. 5. Distribution dataset by class before SMOTE process.

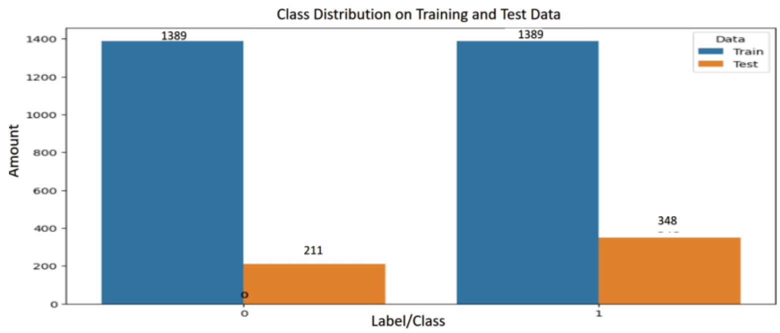

Fig. 6. Distribution dataset by class after SMOTE process.

4.7. Discussion

During the evaluation stage, the performance of various model and feature extraction combinations was measured using a confusion matrix to identify the most accurate configuration [27, 28]. As illustrated in Fig. 7 and Fig. 8, the matrix categorizes outcomes into TP, TN, FP, and FN, which serve as the basis for calculating accuracy, precision, recall, and F1-score via formulas (6) through (9) [29-31].

The Logistic Regression model using TF-IDF and SMOTE achieved the best results with 90.16% accuracy and a 92.22% F1-score, correctly predicting 506 data points. Errors primarily occurred when stop-word removal deleted critical sentiment context (Error Type I) or when labelling inconsistencies led to misclassifications between positive and negative sentiments (Error Type II) (see Figs. 9 and 10). Word cloud at Fig. 11, visualizations were generated for both positive and negative sentiments to identify frequently occurring terms and gain deeper insights into user experiences [3]. As shown in Fig. 12, the negative word cloud is dominated by terms such as "*masuk*," "login," and "*akun*," indicating that authentication and account access issues are the primary drivers of user dissatisfaction [28, 29] (see Table 4).

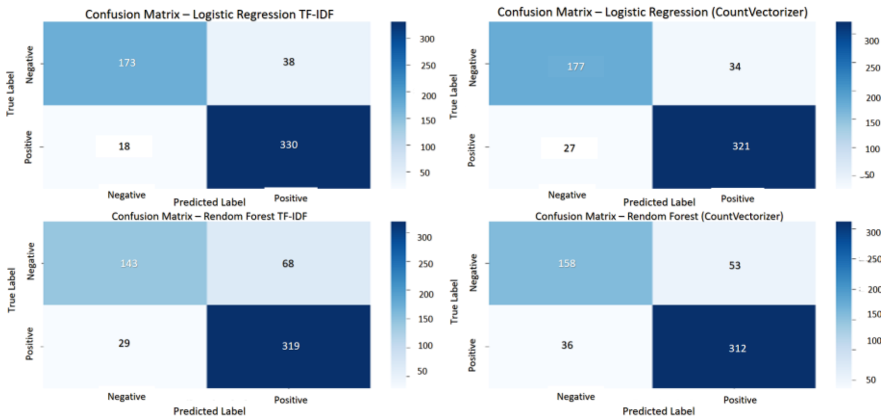

Fig. 7. Distribution dataset without SMOTE by confusion matrix.

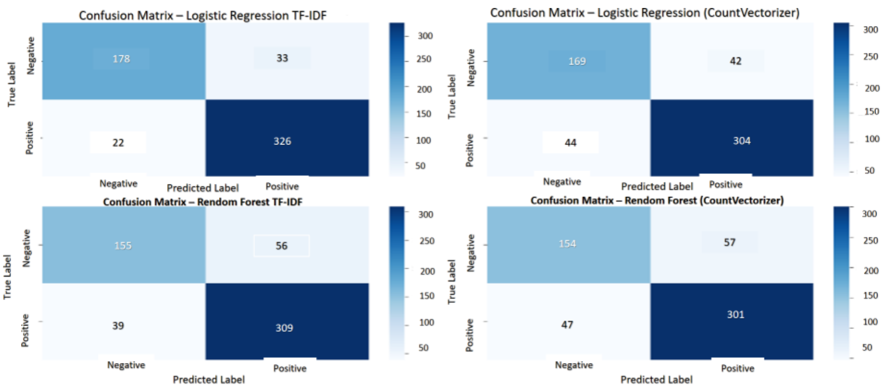

Fig. 8. Distribution dataset with SMOTE by confusion matrix.

Table 4. Comparison performance model.

Model	Acc.	Prec.	Rec.	F1-Score
LR TF-IDF	0.899	0.897	0.948	0.922
LR BoW	0.890	0.904	0.922	0.913

RF TF-IDF	0.826	0.824	0.917	0.868
RF BoW	0.840	0.855	0.897	0.875
LR TF-IDF ST	0.901	0.908	0.937	0.922
LR BoW ST	0.846	0.879	0.874	0.876
RF TF-IDF ST	0.830	0.847	0.888	0.867
RF BoW ST	0.813	0.841	0.865	0.853

	Review	polarity	Predicted_Polarity	Stemmed_Text
2541	Mohon maaf izin ngasih sedikit ulasan .Aku barusan ganti username di twitter web (laptop), terus twitter yg dihp otomatis keluar, nah aku mau masuk lagi kok dibilang ada kesalahan teknis terus? Di...	negative	positive	mohon maaf izin ulas barusan ganti username twitter web laptop twitter dihp otomatis masuk bilang salah teknis suruh nyoba
593	di suruh ngisi verifikasi kode lah gk di kirimin nmrnya emng bisa ngisi udah lama malah gk bisa	negative	positive	suruh isi verifikasi kode kirimin nmrnya isi
2726	senang karna berita yg di sampaikam kurang bohong	negative	positive	senang berita sampaikam bohong
343	jelek, apk nya gabisa login terus sering bug, makin rusak deh	negative	positive	jelek apk login bug rusak deh
262	udah bayar premium, saldo udah kepotong, malah dimintain lagi buat bayar, parah banget bug nya. rugi lah saya	negative	positive	bayar premium saldo potong dimintain bayar parah banget bug rugi
1592	ngak ada beda nya setelah di update 👎👎	negative	positive	ngak beda update
2547	Mau daftar susah sekali, masukin nomor hp tp kode verifnya ga muncul2 , pdhal sinyal aman pulsa aman. Akhirnyaaa, notifikasinya tidak dapat mendaftarkan nomor ini . Trs dicoba menggunakan nomor ba...	negative	positive	daftar susah masuk nomor hp kode verifnya muncul pdhal sinyal aman pulsa aman akhirnyaa notifikasi daftar nomor coba nomor hasil hadeehh capee

Fig. 9. An example instance found for error type I.

	Review	polarity	Predicted_Polarity	Stemmed_Text
2529	Gajelas udah 5 kali di suruh verifikasi akun X itu pun ga jadi jadi ganti akun sama aja ;(	positive	negative	gajelas kali suruh verifikasi akun x ganti akun
2032	susah untuk mendaftar, otp mau lewat email atau no hp tidak kunjung datang, gimana mau masukin otpnya kalau otp nya saja tidak ada terus, dicoba keseringan malah diblok, ga bisa dicoba lagi harus ...	positive	negative	susah daftar otp email nomor hp kunjung masuk otpnya otp coba sering blok coba tunggu apk sombong daftar sulit
238	aplikasi jelek ga usah di download sering error sering muncul terjadi masalah, padahal koneksi wifi lancar, paketan pun juga lancar	positive	negative	aplikasi jelek download error muncul koneksi wifi lancar paket lancar
2252	Saya kan mau login di akun yang lama, terus disuruh masukin nomer telepon, pas udah dimasukkan biasanya langsung muncul SMS kode dari pihak Twitter, tapi kok tadi punyaku gak ada sama sekali notif...	positive	negative	login akun suruh masuk nomer telepon pas masuk langsung muncul sms kode twitter punya notifikasi sms twitter bingung plus emosi coba kali tetep muncul notifikasi smsnya
2134	"Aplikasi X luar biasa! Saya sangat terkesan dengan antarmuka yang intuitif dan fitur-fitur yang canggih. Pengalaman pengguna yang lancar membuat saya semakin terpikat untuk menggunakan aplikasi i...	positive	negative	aplikasi x kes antarmuka intuitif fiturfitur canggih alam guna lancar pikat aplikasi fungsi sedia aplikasi x bantu tingkat produktivitas hidup mudah rekomendasi aplikasi cari solusi andal efisien ...
574	keluar sendiri giliran di login malah gak bisa	positive	negative	gilir login

Fig. 10. An example instance found for error type II.

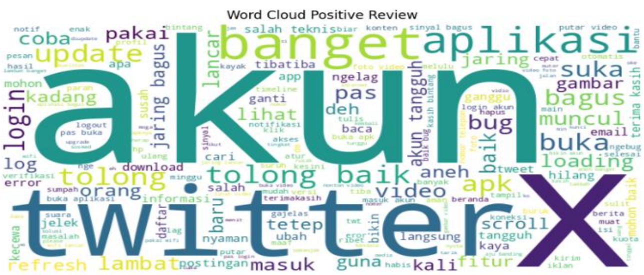

Fig. 11. The big word is predicted as a positive sentiment.

Fig. 12. The big word is predicted as a negative sentiment.

5. Conclusion

Based on a sentiment analysis of 2,794 Google Play Store reviews, the Logistic Regression model combined with TF-IDF and SMOTE emerged as the superior approach, achieving a 90.2% accuracy and 92.2% F1-Score. This configuration consistently outperformed Random Forest, which reached only 83.0% accuracy, proving that Logistic Regression is the most reliable method for this specific dataset. Future research may consider using other preprocessing techniques to improve data quality so that the model can predict classes more accurately.

References

1. Sengul, K. (2025). How are far-right online communities using X/Twitter spaces? Discourse, communication, sharing. *Discourse, Context and Media*, 65, 100884-100898.
2. Tanjung, M.R.; Heryana, N.; and Siska. (2024). Analisis sentimen pengguna platform X terhadap Indonesia Emas 2045 menggunakan algoritma support vector machine. *Jurnal Mahasiswa Teknik Informatika*, 8(4), 7657-7665.
3. Junianto, H.; Saputro, R.E.; Kusuma, B.A.; and Saputra, D.I. (2024). Comparison of logistic regression and random forest in sentiment analysis of Disdukcapil application reviews. *Jurnal Teknik Informatika (JUTIF)*, 5(6), 1539-1547.
4. Ningrum, M.P.; Mutia, R.; Azmi, H.; and Khalifah, H.D. (2025). Sentiment analysis of Twitter reviews on Google Play Store using a combination of convolutional neural network and long short-term memory algorithms. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(2), 107-115.
5. Semary, N.A.; Ahmed, W.; Amin, K.; Pławiak, P.; and Hammad, M. (2024). Enhancing machine learning-based sentiment analysis through feature extraction techniques. *Plos One*, 29(2), 1-19.
6. Chan, A.S.; Husna, M.; and Putra, P.P. (2025). A performance enhancement strategy for sentiment classification models on political social media using hyperparameter tuning and boosting. *International Journal of Advances in Data and Information Systems,* 6(3), 763-772.
7. Firda, H.; Putra, P.; Oktadini, N.R.; Sevtiyuni, P.E.; and Sevtiyuni, P.E. (2025). Perbandingan pelabelan rating-based dan inset lexicon-based dalam

analisis sentimen menggunakan SVM (Studi kasus: Ulasan aplikasi GoBiz di google play store). *Sistemasi: Jurnal Sistem Informasi*, 24(2), 516-529.

8. Palomino, M.A.; and Aider, F. (2022). Evaluating the effectiveness of text pre-processing in sentiment analysis. *Applied Sciences*, 12(17), 8765-8786.
9. Ayuningtyas, P.; Khomsah, S.; and Sudianto, S. (2024). Pelabelan sentimen berbasis semi-supervised learning menggunakan algoritma LSTM dan GRU. *Jurnal Informatika Sunan Kalijaga*, 9(3), 217-229.
10. Manuaba, I.B.; Dantes, G.R.; and Indrawan, G. (2022). Analisis sentimen data provider layanan internet pada Twitter menggunakan support vector machine (SVM) dengan penambahan algoritma Levenshtein distance. *Jurnal Sistem Komputer dan Kecerdasan Buatan*, 5(2), 9-17.
11. Yutika, C.H.; Adiwijaya, A.; and Faraby, S.A. (2021). Analisis sentimen berbasis aspek pada review Female Daily menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 415-421.
12. Işık, M.; and Dağ, H. (2020). The impact of text preprocessing on the prediction of review ratings. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(3), 1405-1421.
13. Siino, M.; Tinnirello, I.; and Cascia, M.L. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, 121, 1-19.
14. Ayub, N.; Talib, M.R.; Hanif, M.K.; and Awais, M. (2023). Aspect extraction approach for sentiment analysis using keywords. *Computer, Materials and Continua*, 74(3), 6799-6892.
15. Setiawan, Y.; Maulidevi, N.U.; and Surendro, K. (2024). The optimization of n-gram feature extraction based on term occurrence for cyberbullying classification. *Data Science Journal*, 23(1), 31-53.
16. Flamholz, Z.N.; Crane-Droesch, A.; Ungar, L.H.; and Weissman, G.E. (2022). Word embeddings trained on published case reports are lightweight, effective for clinical tasks, and free of protected health information. *Journal of Biomedical Informatics*, 125, 103971-103987.
17. Aftab, F.; Bazai, S.U.; Marjan, S.; Baloch, L.; Aslam, S.; Amphawan, A.; and Neo, T.K.; (2023). A comprehensive survey on sentiment analysis techniques. *International Journal of Technology*, 14(6), 1288-1298.
18. Anggraeni, A.; Cholissodin, I.; and Marji. (2021). Sentimen analisis layanan produk Indihome menggunakan information gain dan metode K-nearest neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(8), 3616-3624.
19. Guntara, R.G.; Kashira, F.B.; Amri, T.K.; Restu, L.B.; and Susanto, F.R. (2024). Analisis penjualan handphone di Tokopedia dengan teknik web scraping menggunakan Python pada Google Colab. *Ulil Albab: Jurnal Ilmiah Multidisiplin*, 3(4), 69-75.
20. Onyenwe, I.; Onyedinma, E.; Nwafor, C.; and Agbata, O. (2021). Developing products update-alert system for e-commerce websites users using html data and web scraping technique. *International Journal on Natural Language Computing*, 10(5), 1–7.

21. Putri, N.A.; Srirahayu, A.; and Sudibyo, N.A. (2025). Analisis sentimen terhadap aplikasi KitaLulus menggunakan metode Naive Bayes dari ulasan Google Play Store. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 14(2), 269-279.
22. Naufal, M.F.; and Kusuma, S.F. (2022). Analisis sentimen pada media sosial Twitter terhadap kebijakan pemberlakuan pembatasan kegiatan masyarakat berbasis deep learning. *Jurnal Edukasi dan Penelitian Informatika*, 8(1), 44-49.
23. Putra, K.T.; Hariyadi, M.A.; and Crysdian, C. (2023). Perbandingan feature extraction TF-IDF dan BoW untuk analisis sentimen berbasis SVM. *Jurnal Cahaya Mandalika*, 3(2), 1449-1463.
24. Suryanti, R.; and Prasetyaningrum, P.T. (2025). Perbandingan metode TF-IDF dan bag of words dalam analisis sentimen diet kopi Americano di media sosial Twitter menggunakan Naïve Bayes. *Building of Informatics, Technology and Science (BITS)*, 7(1), 104-115.
25. Budaya, I.G.; and Suniantara, I.K. (2024). Comparison of sentiment analysis algorithms with SMOTE oversampling and TF-IDF implementation on Google Reviews for public health centers. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1077-1086.
26. Özmen, C.G.; and Gündüz, S. (2025). Comparison of machine learning models for sentiment analysis of big Turkish web-based data. *Applied Sciences*, 15(5), 2297-2317.
27. Sathyanarayanan, S.; and Tantri, B.R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, 27(4), 4023-4031.
28. 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. *World Journal of Advanced Research and Reviews*, 23(3), 2842-2852.
29. Momshad, A.M.; Islam, M.H.; Datta, S.D.; Sobuz, M.H.R.; Aayaz, R.; Kabbo, M.K.I.; and Khan, M.M.H. (2025). Assessing the engineering properties and environmental impact with explainable machine learning analysis of sustainable concrete utilizing waste banana leaf ash as a partial cement replacement. *Cleaner Engineering and Technology*, 24(2025), 10088-10101.
30. Hiba, A.; Adarsh, S.; Md Munir, H.K.; and Ali, D.M. (2025). Drought modeling and forecasting using shallow and deep machine learning techniques. *Modeling Earth Systems and Environment*, 11(2025), 41-61.
31. Alazaidah, R.; Al-Batah, M.S.; Samara, G.; Alluwaici, M.; Safaqah, M.A.; Aljaidi, M.; Mohammad, S.I.; Vasudevan, A.; and Zhang, Y. (Tahun). Prediction of landslides using data mining techniques. *Applied Mathematics and Information Sciences*, 19(2), 437-477.