Jurnal Teknik Informatika (JUTIF)
Vol. 6, No. 1, February 2025, pp. 159-172

DOI: https://doi.org/10.52436/1.jutif.2025.6.1.4178
p-ISSN: 2723-3863
e-ISSN: 2723-3871

PERFORMANCE COMPARISON OF NAIVE BAYES AND BIDIRECTIONAL LSTM ALGORITHMS IN BSI MOBILE REVIEW SENTIMENT ANALYSIS

Hannatul Ma'we*[1], Ario Yudo Husodo[2], Budi Irmawati[3]

[1]Master of Information Technology, Faculty of Engineering, University of Mataram, Indonesia
[2,3]University of Mataram, Indonesia
Email: [1] i2s02310003@student.unram.ac.id, [2] ario@unram.ac.id, [3] budi-i@unram.ac.id

(Article received: December 5, 2024; Revision: December 18, 2024; published: February 20, 2025)

Abstract

Currently, almost all banks have used mobile banking in conducting banking transactions, one of which is Bank Syariah Indonesia (BSI). BSI mobile is still classified as a new mobile banking application compared to other mobile banking, this certainly still has a low rating and really needs feedback from users which can be seen through reviews on the Google Play Store application. Input in the form of criticism and suggestions from BSI mobile users can be used by BSI mobile as a suggestion for careful supervision and evaluation material in improving its services. This study aims to find the best algorithm to analyze review sentiment on the Google Play Store for the BSI mobile application and provide an overview of the response of application users to application developers based on the results of review data processing. The data mining methodology used in this study is CRISP-DM, using a dataset collected for 6 years (2018-2023) which is annotated into positive and negative labels manually, then modeled using 2 algorithms, namely Naïve Bayes (NB) and Bidirectional LSTM (BiLSTM). The contribution of this study is to test, evaluate and compare the two algorithms (NB and BiLSTM) using the K-Fold Cross Validation (NB) testing model and over-sampling techniques to the minority class (negative) then provide recommendations for the best algorithm. The conclusion of the study is that the BiLSTM algorithm is superior to NB with an accuracy of 94.90 % while the NB algorithm is 94%. In addition, the over-sampling technique is more optimal in increasing the accuracy of the algorithm's performance compared to without over-sampling.

Keywords : *sentiment analysis, bidirectional LSTM, BSI mobile, deep learning, machine learning, naive bayes*

PERBANDINGAN KINERJA ALGORITMA NAÏVE BAYES DAN BIDIRECTIONAL LSTM DALAM ANALISIS SENTIMEN ULASAN BSI MOBILE

Abstrak

Saat ini hampir semua bank telah menggunakan mobile banking dalam melakukan transaksi perbankan salah satunya yakni Bank Syariah Indonesia (BSI). BSI mobile masih tergolong aplikasi mobile banking baru dibandingkan dnegan mobile banking lainnya, hal ini tentunya masih memiliki rating yang rendah dan sangat membutuhkan umpan balik dari para pengguna yang bisa dilihat melalui ulasan pada aplikasi Google Play Store. Masukan berupa kritik dan saran dari pengguna BSI mobile dapat digunakan pihak BSI mobile sebagai saranan pengawasan cermat dan bahan evaluasi dalam memperbaiki layanannya. Penelitian ini bertujuan untuk menemukan algoritma terbaik untuk menganalisis sentimen ulasan di Google Play Store terhadap aplikasi BSI mobile dan memberikan gambaran respon para pengguna aplikasi kepada pengembang aplikasi berdasarkan hasil pemrosesan data ulasan. Metodologi penambangan data yang digunakan pada penelitian ini yaitu CRISP-DM, menggunakan dataset yang dikumpulkan selama 6 tahun (2018-2023) yang dianotasi ke dalam label positif dan negatif secara manual, kemudian di modelkan menggunakan 2 algoritma yaitu Naïve Bayes (NB) dan Bidirectional LSTM (BiLSTM). Kontribusi penelitian ini yaitu menguji, mengevaluasi dan membandingkan kedua algoritma (NB dan BiLSTM) menggunakan model pengujian K-Fold Cross Validation (NB) dan teknik over sampling kepada kelas minority (negatif) kemudian memberikan rekomendasi algoritma terbaik. Kesimpulan penelitian adalah algoritma BiLSTM lebih unggul dari NB dengan akurasi sebesar 94,90% sedangkan algoritma NB sebesar 94%. Selain itu teknik over sampling lebih optimal dalam meningkatkan akurasi dari performa algoritma dibandingkan tanpa over sampling.

Kata kunci: *analisis sentimen, bidirectional LSTM, BSI mobile, deep learning, machine learning, naïve bayes*

1. INTRODUCTION

Currently, almost all banks use mobile banking for transactions, one of which is Bank Syariah Indonesia (BSI). BSI mobile is still a new mobile banking application compared to other mobile banking applications, of course it has a low rating and of course it also requires a lot of feedback from customers that can be seen, one of which is through Google Play Store reviews and ratings. For companies that offer mobile banking services, it is important to know whether the quality of mobile banking services provides convenience in transactions or not, whether it provides customer satisfaction or not, what needs are desired by users and what needs to be done. Receive and respond to input from customers or BSI Mobile users. Criticism and input from customers or BSI Mobile users can be used as evaluation material by BSI Mobile or BSI Mobile App developers to improve their services. From this assessment, a strategic plan can be made to evaluate, improve and aggregate many services. BSI users or potential users to compete with other mobile banking competitors. One way to reconcile this process is through sentiment analysis.

Sentiment analysis is part of the Natural Language Processing (NLP) process and information retrieval [1]. Sentiment analysis is also a field of data mining [2] that is used to find out a person's feelings or emotions towards something, such as a product or service so that they can determine marketing strategies, business focus, product optimization, whether in the form of text, images, audio or video [3].

The algorithms used in previous research [4] are supervised learning algorithms including Naive Bayes, SVM, KNN, and Decision Tree. The amount of data used is still relatively small, namely 2554 materials, the percentage of positive opinions is 59.2% and the percentage of negative opinions is 40.8%. With this amount of data, the results of the classification model are still considered poor in predicting negative opinions because the proportion of training data in each class is not balanced. Therefore, researchers developed this sentiment analysis by adding quite a lot of data, namely 98,820 data from 2018 to 2023.

This study implements two classification algorithms for sentiment analysis, namely Naïve Bayes (NB) and Bidirectional LSTM (BiLSTM), then compares the performance of the two algorithms to find the best algorithm. The NB algorithm is one of the techniques for evaluating the accuracy of the model built based on the dataset used [5] . While the BiLSTM algorithm relies on the classification labels in the training documents with the test documents [6] and BiLSTM is one of the algorithms widely used in analyzing text sentiment [7] .

There are several previous studies that have applied the NB and BiLSTM algorithms in analyzing sentiment in mobile banking application reviews and other applications and social phenomena, including sentiment analysis to obtain customer satisfaction analysis of Jenius, Jago and Blue banks using the SVM, Linear Regression, Random Forest, NB and LGBM algorithms. Based on the results obtained, the NB algorithm produces an accuracy of 74.67% [8] . Sentiment analysis of the Skype application by detecting similar topics before and after the update using the experimental method can find users' opinions about the advantages and disadvantages of the Skype application [9] . Sentiment analysis of user needs and the quality of bank applications by [3] with the Naïve Bayes and LDA algorithms produces the best algorithm with an accuracy of 93.47% at a value of k = 5. Comparison of the NB and KNN algorithms in analyzing the sentiment of KRL Commuter Line Jabodetabek users on Twitter application data [10] using the NB, KNN and Decision Tree algorithms produces an NB algorithm accuracy of 80%.

Sentiment analysis of Gopay, Ovo and LinkAja user customers was conducted by [11] with the NB and KNN algorithms. The results of the NB algorithm on the application were sequentially 62.53 % , 75.60% and 69.5%. The NB algorithm is suitable for use in research [12] to see public opinion about the Sinovac and Pfizer vaccines in Indonesia with an accuracy of 85% and 69%. To find out the public mood [2] conducted a sentiment analysis of the BIST30 stock market in Istanbul with NB algorithm results of up to 68%. In addition, to find functions and improve business quickly [13] used the Random Fores and Naïve Bayes algorithm approaches.

Sentiment analysis on Bibit and Bareksa application reviews using the NB and KNN algorithms obtained an accuracy of 85.14% and 81.70% [14] . In the sentiment analysis of the Shopee application, the NB algorithm is superior to KNN with an accuracy of 80% [15] and in the E-Government application, an accuracy of 89% was also produced using the same algorithm [16] . Subsequent research on sentiment analysis of the CAPCUT application using the NB algorithm produced the best accuracy of 79.41% [17] and in the Mypertamina application, an accuracy of 70.73% was obtained. [18] . Classification of depressive sentiment on Youtube using the NB algorithm produced an accuracy of 84.11% [19] .

Recently [20] successfully detected user views on the Covid-19 pandemic using the BiLSTM algorithm with an accuracy of 99.33%. In research [21] the BiLSTM algorithm can also classify sentiments related to Covid-19 from social media. In mobile electronic products [22] using the BiLSTM algorithm succeeded in analyzing user sentiment up to 91.40%.

Analysis of the level of satisfaction of Grab Indonesia application violations was carried out [23] with the BiLSTM and LSTM algorithms. The results of this analysis stated that BiLSTM is better and more reliable to use than the LSTM algorithm, namely an accuracy of 91%. The BiLSTM algorithm is also used to detect news of community safety crisis incidents on the internet. The proposed model produces an accuracy of up to 90.93% on text data and 82.10% on voice data [24] . [25] classifying mobile users' application security opinions with the BiLSTM algorithm achieved 90% accuracy.

45 participants were recruited using a mobile banking application by [26] with the aim of authenticating users on the device. The test was successful on raw data from motion sensors and touch screens with up to 99.85 % accuracy.

Based on the background above, this study conducted optimization of sentiment analysis performance from previous research [4] using the Naive Bayes algorithm of the Multinomial type using a larger amount of data, namely 98,820 data and using the Bidirectional Long Short Term Memory (BilLSTM) algorithm with a larger amount of data (98,820). The Naïve Bayes and Bidirectional LSTM methods are often used, more effective and proven to be suitable for use in conducting text sentiment analysis with a fairly large amount of data [3] , [11] , [13] , [8] , [20] , [7] , [21] , [26], [22] , [23] , [25] , [24] , [27] , [28] , [29] , [30] , [31] .

The purpose of this study is to find the best algorithm to analyze the sentiment of BSI mobile user reviews on the Google Play Store and provide recommendations for application developers based on the results of data processing.

2. RESEARCH METHOD

This research was conducted from October 2023 to July 2024. The design of this research was in the form of testing using data obtained from the Google Play Store as a test. This research stage uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The stages of this research include: data scraping, data source analysis, data preprocessing, sentiment analysis using the Naïve Bayes and Bidirectional LSTM algorithms, method evaluation with confusion matrix. The research methodology flow diagram is shown in Figure 1 below:

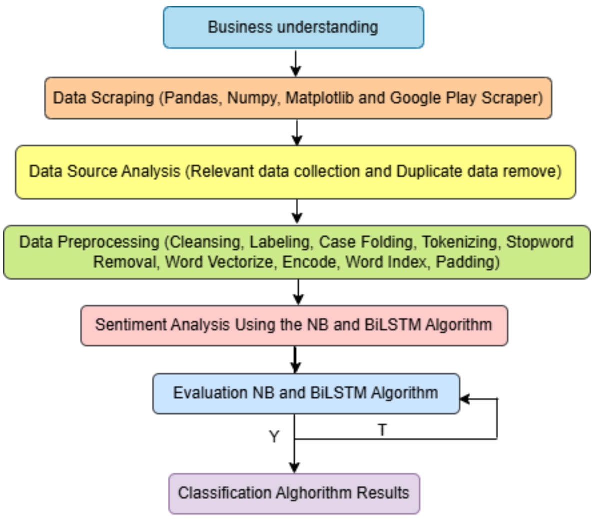


Figure 1. Research Methodology

2.1 Business Understanding

Business understanding contains literature studies to find information about previous research references that have been carried out previously related to sentiment analysis using the Naïve Bayes (NB) and Bidirectional LSTM (BiLSTM) algorithms. The NB and BiLSTM algorithms are algorithms that are suitable for the characteristics of the data used.

2.2 Data collection

The data collection method in this study is Scraping. In this study, the data collection process was carried out using the scraping technique using regular expressions with the Python programming language in Google Colab. The libraries used are Pandas, Numpy, Matplotlib and Google Play Scraper. Data was taken using the keyword of the BSI Mobile website page on Google Play, keyword "com.bsm.activity".

2.3 Data Source Analysis

The next stage is the analysis of data sources by taking relevant data and deleting duplicate data. The relevant data used in this study are only content attributes or BSI Mobile user reviews.

2.4 Data Preprocessing

In the preprocessing stage, data labeling, cleansing, case folding, tokenizing, stopword removal and stemming are carried out. Furthermore, the TF-IDF process is carried out on the Naïve Bayes algorithm while in the Bidirectional LSTM algorithm, the set is divided between training data and test data, data vectorization, word index and padding are carried out.

2.5 Sentiment Analysis

The sentiment analysis process is carried out using the Naïve Bayes and Bidirectional LSTM algorithms. Multinomia Naïve Bayes and Bidirectional LSTM are popular algorithms used to classify sentiment analysis text with a large amount of data.

After the preprocessing process, sentiment analysis is carried out using the BiLSTM algorithm by dividing the data into 2 parts, namely the training and testing sets with a proportion of 80% training data and 20% test data. The architecture of the BiLSTM algorithm in conducting sentiment analysis is shown in Figure 2 below.

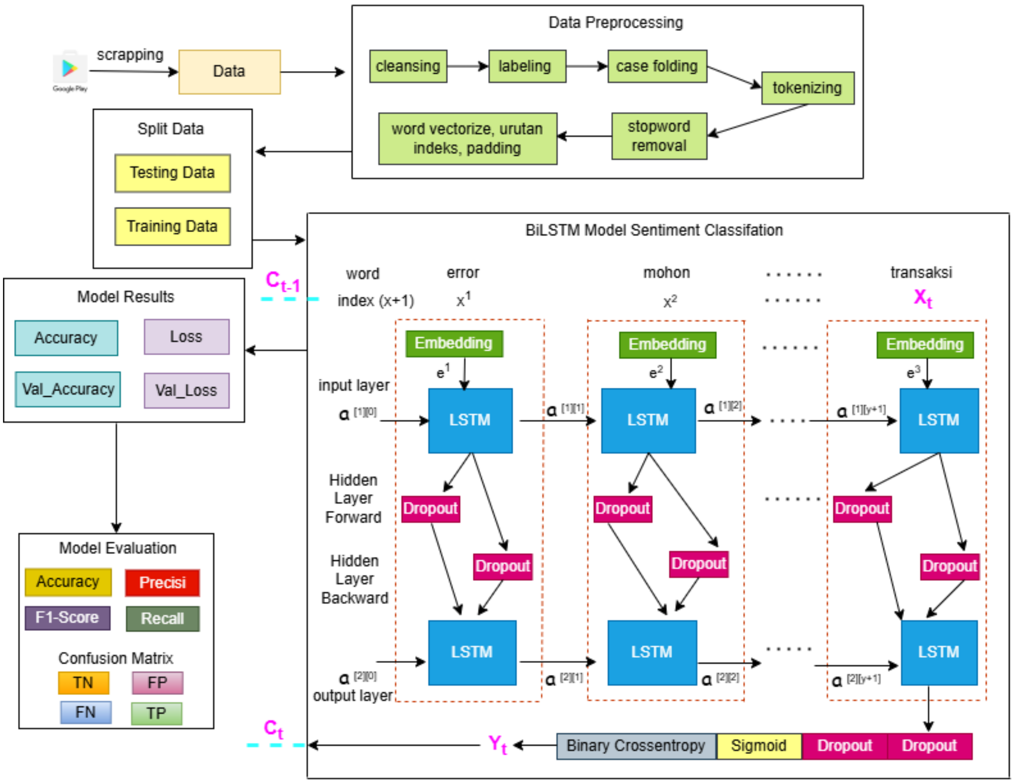


Figure 2. BiLSTM Algorithm Architecture

Notation Description:

X_t	: input to network
Y_t	: output from the network
a_{t-1}	: previously hidden state
C_t	: current state of the cell
C^*_t	: candidate function
LSTM/σ	: sigmoid function (0 or 1/ negative or positive)
W_i, W_f, W_c,	: weight of each input gate, forget gate, candidate function and output gate
	: bias of each input gate, forget gate and output gate
b_i, b_f, b_y	

The BiLSTM algorithm formula is:

$$F_t = \sigma (W_f \,.\, [a_{t-1}, X_t] + b_f) \quad \ldots\ldots\ldots\ldots (1)$$

$$i_t = \sigma (W_i \,.\, [a_{t-1}, X_t] + b_i) \quad \ldots\ldots\ldots\ldots (2)$$

$$C^*_t = \tanh (W_c \,.\, [a_{t-1}, X_t] + b_f) \quad \ldots\ldots\ldots\ldots (3)$$

$$C_t = F_t \,.\, C_{t-1} + i_t \,.\, C^*_t \quad \ldots\ldots\ldots\ldots (4)$$

$$Y_t = \sigma (W_y [a_{t-1}, X_t] + b_y) \quad \ldots\ldots\ldots\ldots (5)$$

$$a_t = Y_t \,.\, \tanh (C_t) \quad \ldots\ldots\ldots\ldots (6)$$

$$H_t BiLSTM = h_t \text{ forward} + h_t \text{ backward} \quad \ldots\ldots\ldots\ldots (7)$$

The formula of the Naïve Bayes algorithm is:

$$P\,H/X = \frac{P(X/H).P(H)}{P(X)} \quad .... (8)$$

Notation Description:

	X	=	Unknown *class* data
	H	=	Data hypothesis of a specific *class*
	P(H|K)	=	Probability of hypothesis H based on condition X / *posteriori probability*
	P(H)	=	Hypothesis probability H/ *prior probability*
	P(X|H)	=	Probability of X based on hypothesis condition H
	P(X)	=	Probability X

2.6 Evaluation Method

Method evaluation is done with confusion matrix to determine the accuracy performance of Naïve Bayes and Bidirectional LSTM. The confusion matrix of this study consists of accuracy, precision, recall and F1-score with their respective formulas, namely:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP\ \_\ FP} \quad (2)$$

$$Recall = \frac{TP}{TP\ \_\ FN} \quad (3)$$

$$F1{=}Score = 2\ x \frac{Recall\ x\ Precision}{Recall+Precision} \quad (4)$$

2.7. Data Visualization

The visualization of the words of this research data is presented using a “cloud” by highlighting the frequency of words in the text. This term is often referred to as a word cloud to visually describe words that have high frequencies in user reviews on the BSI Mobile application.

3. RESULTS

This section shows the results and analysis based on sentiment analysis conducted on BSI Mobile application reviews using Naïve Bayes and Bidirectional LSTM. The main source of data comes from the results of data scraping on the BSI Mobile application on the Google Play Store. After carrying out the data scraping process, 98,820 user review data were obtained. Furthermore, the selection of relevant data attributes was carried out using "content/reviews" and reviews that were indicated as duplicates were removed using the RapidMinner Studio tool. The duplicate removal process produced 74,570 clean data datasets. This dataset was manually labeled by a linguist into 2 sentiment labels, namely positive and negative sentiments. The results of data labeling can be seen in Table 1.

Table 1. Data

Total Sentiment	
Positive	45,806
Negative	28,764

The data that has been manually labeled is then processed further, namely data cleansing, case folding, stopword removal, tokenizing, stemming. The next process in the Naïve Bayes algorithm is the TF-IDF calculation process, while in the Bidirectional LSTM algorithm, the data vectorization process, word indexing and padding are carried out.

The data that has been generated from a series of previous processes is reprocessed by dividing the data into 2 sets of divisions, namely test data and training data and applied to the Bidirectional LSTM algorithm. While the Multinomial Naïve Bayes algorithm in this study uses cross validation techniques in determining data randomly.

The findings show that BSI Mobile application users express their sentiments into sentiment groups, namely sentiment statistical data can provide a picture of reviews in 2 categories and classification accuracy is described using a confusion matrix consisting of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

The analysis results are graphically depicted using word clouds in 2 review categories. This word cloud can provide convenience in finding the identification and characteristics of user reviews. This provides a useful understanding of sentiment in a more nuanced way.

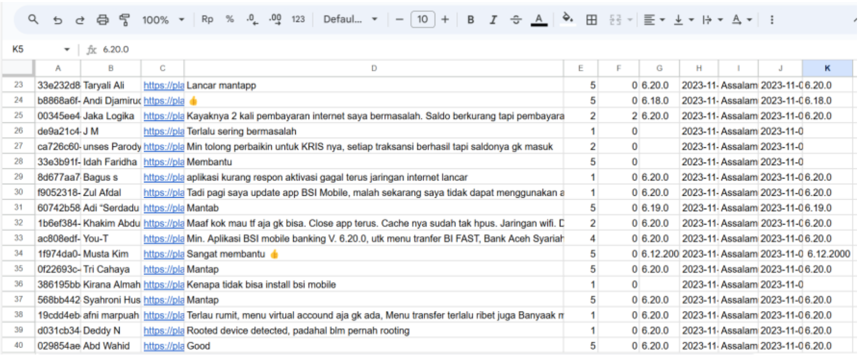
3.1. Data collection

The scraping data from the Google Play Store BSI Mobile in this study obtained 98,820 data in CSV format. Google Colab and the code used to obtain the data can be illustrated in Figure 3.

```
[ ] pip install google-play-scraper

Requirement already satisfied: google-play-scraper in /usr/local/lib/python3.10/dist-packages (1.2.6)

[ ] # Import necessary libraries
    import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    from google_play_scraper import app, Sort, reviews_all

[ ]  #Define and configure Google Play Scraper library
    us_users_reviews = reviews_all(
        #1https://play.google.com/store/apps/details?id=jd.cdyjy.overseas.market.indonesia&showAllReviews=true dari ini diambil jadi ini
        #https://play.google.com/store/apps/details?id=com.bsm.activity2
        #1'id=jd.cdyjy.overseas.market.indonesia',
        'com.bsm.activity2',
        sleep_milliseconds=0,
        lang='id', # Default language is 'en', set language to Chinese.
        country='id', # Default country is 'us', set country to Hong Kong.
        sort=Sort.NEWEST, # Default is Sort.MOST_RELEVANT.
    )
```

Figure 3. Scraping Data with Google Colab

The results of data scraping using Google Colab are shown in Figure 4.



	A	B	C	D	E	F	G	H	I	J	K
23	33e232d8	Taryali Ali	https://pla	Lancar mantapp	5	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
24	b8868a6f-	Andi Djamiruc	https://pla	👍	5	0	6.18.0	2023-11-	Assalam	2023-11-0	6.18.0
25	00345ee4	Jaka Logika	https://pla	Kayaknya 2 kali pembayaran internet saya bermasalah. Saldo berkurang tapi pembayara	2	2	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
26	de9a21c4-	J M	https://pla	Terlalu sering bermasalah	1	0		2023-11-	Assalam	2023-11-0	
27	ca726c60-	unses Parody	https://pla	Min tolong perbaikin untuk KRIS nya, setiap traksansi berhasil tapi saldonya gk masuk	2	0		2023-11-	Assalam	2023-11-0	
28	33e3b91f-	Idah Faridha	https://pla	Membantu	5	0		2023-11-	Assalam	2023-11-0	
29	8d677aa7	Bagus s	https://pla	aplikasi kurang respon aktivasi gagal terus jaringan internet lancar	1	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
30	f9052318-	Zul Afdal	https://pla	Tadi pagi saya update app BSI Mobile, malah sekarang saya tidak dapat menggunakan a	1	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
31	60742b58	Adi "Serdadu	https://pla	Mantab	5	0	6.19.0	2023-11-	Assalam	2023-11-0	6.19.0
32	1b6ef384-	Khakim Abdu	https://pla	Maaf kok mau tf aja gk bisa. Close app terus. Cache nya sudah tak hpus. Jaringan wifi. D	2	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
33	ac808edf-	You-T	https://pla	Min. Aplikasi BSI mobile banking V. 6.20.0, utk menu tranfer BI FAST, Bank Aceh Syariah	4	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
34	1f974da0-	Musta Kim	https://pla	Sangat membantu 👍	5	0	6.12.200	2023-11-	Assalam	2023-11-0	6.12.2000
35	0f22693c-	Tri Cahaya	https://pla	Mantap	5	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
36	386195bb	Kirana Almah	https://pla	Kenapa tidak bisa install bsi mobile	1	0		2023-11-	Assalam	2023-11-0	
37	568bb442	Syahroni Hus	https://pla	Mantap	5	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
38	19cdd4eb	afni marpuah	https://pla	Terlau rumit, menu virtual accound aja gk ada, Menu transfer terlalu ribet juga Banyaak m	1	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
39	d031cb34	Deddy N	https://pla	Rooted device detected, padahal blm pernah rooting	1	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0
40	029854ae	Abd Wahid	https://pla	Good	5	0	6.20.0	2023-11-	Assalam	2023-11-0	6.20.0

Figure 4. Data Scraping Results with Google Colab

3.2. Data Source Analysis

The results of the data analysis obtained the attribute "content" as material in analyzing sentiment and removing duplicates using the Rapid Miner Studio tool produced clean data of 74,570 datasets. The results of the attribute selection and the process of removing duplicates are illustrated in Figure 5.

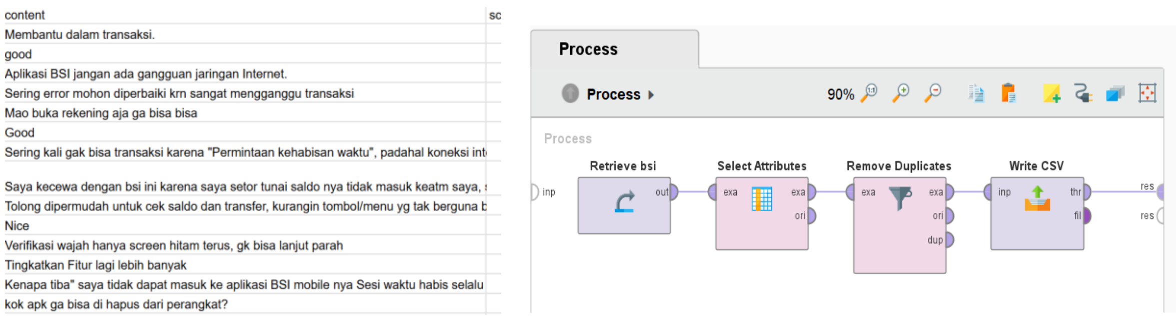


Figure 5. Attribute Selection Results and Remove Duplicate Process

3.3. Data Preprocessing

The next process is to clean the data from all types of punctuation, emoticons and other signs. The results of the cleansing and case folding process are shown in Figure 6.

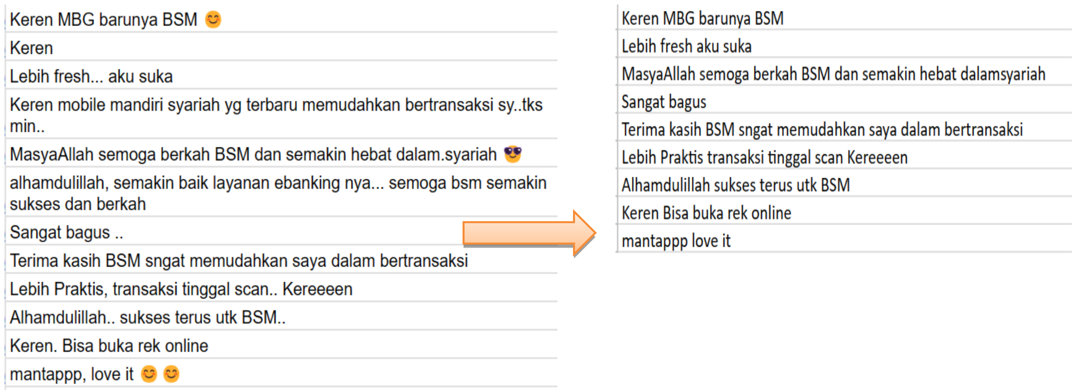


Figure 6. Data Cleansing Results

The deletion of empty and meaningless content data is carried out next. The total data until the duplicate removal process stage is 74,823 data to 74,570 datasets after the deletion of empty data and meaningless content. Figure 7 shows the Meaningless Content Row.

	CONTENT
55126	aGus suparru
57780	ssh
62647	hu ui okk i i
64037	ufh
64129	L O
64180	Yoi mamae
64847	Anteekk Maruuuun
65625	Bla Bla Bla
65837	Widodo Budi darmo
65905	Suntoro A
65949	Gak
66006	Waalaikumsalam
66010	Sugiyanto
67825	Nomrr saty
68229	Budinya ngakak
68420	Kontuik
68573	bodola
69258	Totemo benri desu
69270	Angel
69335	33nm

Figure 7. Data Content Rows That Have No Meaning

74,570 datasets were manually labeled into 2 label categories, namely positive labels and negative labels. Figure 8 shows the Data Labeling Results.

Rooted device detected padahal blm pernah rooting	negatif
Baik	positif
Mantul banget	positif
Baru mau buka Apkikasi belum apa-apa malah keluar sendiri	negatif
Aplikasinya sering eror Tidak bisa digunakan padahal jaringan bagus	negatif
Ada bug tidak bisa transfer Tolong diperbaiki	negatif
Elor teruss	negatif
Mohon di tindak lanjuti lagiaplikasi sering close sendiri berkali2 jadi susah transksi dan demi keamanan mohon segera di betulkan lagi sistemnya Terima kasih	negatif
Biaya administrasi TF ke bank lain buset mahal bgt	negatif
Giliran mau dipake malah eror Jelek kualitas aplikasinya	negatif
Bsi kenapa sering banget gangguannya Mau ngirim gk bs buka aplikasi nya mlm ini	negatif
Apk bisa bug	positif
Mengecewakan mau transfer malah ngebug Giliran pakai kartu di ATM bilang gagal transaksi Dahlah	negatif
Transfer tidak bisa selalu keluar sendiri	negatif
Kenapa selalu logot sebelum trx selesai Menghambat dikala semua serba cepat hmmm	negatif
Kecewa banget saya bukan pengguna bank ini saya pengguna bank sebelah kemarin tanggal 31-10-23 saya TF dari bank ke bank bsi punya sa	negatif

Figure 8. Data Labeling Results

Case folding is a process of changing all uppercase characters in the data into lowercase. This process is shown in Table 2.

Table 2. Case Folding Process Results

Content	Case Folding Results
Membantu dalam transaksi	membantu dalam transaksi
good	good
Aplikasi BSI jangan ada gangguan jaringan Internet	aplikasi bsi jangan ada gangguan jaringan internet
Sering error mohon diperbaiki krn sangat mengganggu transaksi	sering error mohon diperbaiki krn sangat mengganggu transaksi
Mao buka rekening aja ga bisa bisa	mao buka rekening aja ga bisa bisa

Tokenizing is the process of separating text into features. Tokenizing results in Table 3.

Table 3. Tokenizing Process Results

Content	Tokenizing Results
Membantu dalam transaksi	membantu,dalam,transaksi
good	good
Aplikasi BSI jangan ada gangguan jaringan Internet	aplikasi,bsi,jangan,ada,gangguan,jaringan,internet
Sering error mohon diperbaiki krn sangat mengganggu transaksi	sering,error,mohon,diperbaiki,krn,sangat,mengganggu,transaksi
Mao buka rekening aja ga bisa bisa	mao,buka,rekening,aja,ga,bisa,bisa

The Stopword Removal stage is done by removing meaningless words that often appear in the text. The results of this process are shown in Table 4.

Table 4. Stopword Removal Process Results

Content	Stopword Removal Results
Membantu dalam transaksi	membantu transaksi
good	good
Aplikasi BSI jangan ada gangguan jaringan Internet	aplikasi bsi gangguan jaringan internet
Sering error mohon diperbaiki krn sangat mengganggu transaksi	error mohon diperbaiki krn mengganggu transaksi
Mao buka rekening aja ga bisa bisa	mao buka rekening aja ga bisa bisa

The last stage of data preprocessing is Stemming by changing words into basic word forms using a lemmatizer. The results of the stemming stage are shown in Table 5.

Table 5. Stemming Process Results

Content	Stemming Results
Membantu dalam transaksi	bantu dalam transaksi
good	good
Aplikasi BSI jangan ada gangguan jaringan Internet	aplikasi bsi jangan ada ganggu jaring internet
Sering error mohon diperbaiki krn sangat mengganggu transaksi	sering error mohon baik krn sangat ganggu transaksi
Mao buka rekening aja ga bisa bisa	mao buka rekening aja ga bisa bisa

As an additional stage of the process using the Bidirectional LSTM algorithm, namely word vectorize by changing the form of the sentiment label category into a numeric form, word index to provide numbering for each word and padding to change the word numbering into a vector form. The form of word vectorize, word index and padding is shown in Figure 9.

```
{'bsi': 1, 'dan': 2, 'aplikasi': 3, 'di': 4, 'bisa': 5, 'sangat': 6, 'ini': 7, 'saya': 8, 'mudah': 9, 'nya': 10,

Buka  folder  informasi  rek  yg  muncul  cm  daftar  mutasi  Mana  cek  saldonya  Baik  pergi  ke  ATM  klo
[30, 9055, 301, 390, 36, 211, 1200, 215, 331, 266, 147, 719, 68, 758, 26, 119, 290, 161, 23, 3, 36, 935, 70,

[    0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0     0     0     0     0
     0     0     0     0     0     0     0     0    30  9055   301   390
    36   211  1200   215   331   266   147   719    68   758    26   119
   290   161    23     3    36   935    70   361    35 35727   253    26
    36  1792     7   135   927    16   279     3   273    60   279   136
 35728   319   655   334    35  3776    86    18    34   241   604    70
     4   215   331    17  2329   315    51   552    43   615   290    74
   547    70     4     3  1205    34    77  1196  2429  1314    36  2241
  4707   224  2808  3394  1196   331    94   241  6064   172   668   772
    78   279     3    34   499     4 35729   240]
```

Figure 9. Word Vectorize, Index/Text-To-Sequence and Padding Results

3.4. Analysis and Evaluation

Before being analyzed using the Bidirectional LSTM algorithm, the data is first divided into 2 parts, namely test data and training data into 80% test data and 20% training data. While in the Naïve Bayes algorithm, data division is carried out using the cross validation technique. The data is then analyzed using the Naïve Bayes and Bidirectional LSTM algorithms. Because the number of positive labeled data is greater than negative, the SMOTE over sampling method is used to balance the data class to the minority class (negative) which is combined with different hyperparameter configurations, Epoch 3, namely using Dropout 0.2 and 0.5, both using

SpatialDropout1D 0.25, 100 Hidden layers and 0.5 on LSTM convergence and Reccurent-Dropout LSTM on both types of Dropout (0.5 and 0.2).

Table 6. is the result of sentiment analysis using the Bidirectional LSTM algorithm.

Table 6. Results of Sentiment Analysis Using the Bidirectional LSTM Algorithm

Results Evaluation Model (%)	Dropout =0.2		Dropout = 0.5	
	Batch Size		Batch Size	
	32	64	32	64
Accuracy	94,77	94,74	94,86	**94,90**
Precision	94,02	94,40	94,33	94,12
Recall	92,39	92,92	92,35	92,62
F1-Score	93,20	93,20	93,33	93,33

The results of sentiment analysis using the Naïve Bayes algorithm are shown in Table 7.

Table 7. Results of Sentiment Analysis Using the Naïve Bayes Algorithm

Test Results	10 Cross Validation
Train Set	67.113
Test Set	7.457
Mean Score	0.9403
True Negative	27,457
False Positive	3.199
False Negative	1,307
True Positive	42,607
Negative Precision	90%
Positive Precision	97%
Negative Recall	95%
Positive Recall	93%
F1-Score Negative	92%
F1-Score Positive	95%
Accuracy	94%

Based on Table 6. the best test results using the Bidirectional algorithm were produced using Dropout 0.5 rather than using 0.2, which is 94.90%. With the cross validation technique, the Naïve Bayes algorithm produces the highest accuracy of 94%.

The final stage for testing the Bidirectional LSTM model is to conduct analysis testing using new data input on the model. The results of model testing with new data input on the Bidirectional LSTM algorithm model are depicted in Figure10.

Gambar 10. Hasil pengujian model dengan penginputan data baru pada model algoritma Bidirectional LSTM

3.5. Data Visualization

From the preprocessing of the data obtained, the results are in the form of words that are often expressed by users of the BSI Mobile application. The difference in the word cloud results produced by the Naïve Bayes and Bidirectional LSTM algorithms is shown in Figures 11 and 12.

Figure 11. Word Cloud with Naïve Bayes Algorithm

Figure 12. Word Cloud with Bidirectional LSTM Algorithm

At first glance, the visualization results of word cloud data generated by the Naïve Bayes and Bidirectional LSTM algorithms look the same , but there are interesting analysis results. Found as an additional contribution in this study, it produces analysis findings in the form of the performance of each algorithm used. In the Naïve Bayes algorithm, the word “Transaction” falls into 2 sentiment categories. The word "Transaction" is indicated as an ambiguous word because the word "Transaction" appears in both positive and negative classes. This happens because the Naïve Bayes algorithm modeling at the data preprocessing stage uses TF-IDF in the form of n-grams and word similarity with the cosine vector technique so that it cannot capture semantic words well and without considering the context of the sentence. While the word "Transaction" in the Bidirectional LSTM algorithm is included in the negative category only or does not appear in the positive class because the way the Bidirectional LSTM algorithm works categorizes word classes based on the context of the sentence or the entire sentence using semantic vectors in the form of word embedding and word2vec which function to determine the length of the word dimension in a sentence even without calculating TF-IDF.

Therefore, in terms of the context of the sentence, the word "Transaction" is mostly found in the context of sentences that are negative or negative class and also in the training process of the word model "Transaction" is more categorized as a negative class. This is a valuable finding that in terms of text sentiment analysis, the Bidirectional LSTM algorithm is better than the Naïve Bayes algorithm. This finding is still rare and no one has even discussed this in sentiment research.

4. DISCUSSIONS

The testing of this study is in line with previous research [6]-[26] which consistently shows that the NB and BiLSTM algorithms produce high accuracy in conducting sentiment analysis.

Based on the testing of this study, the BiLSTM algorithm is superior to NB in analyzing BSI mobile review sentiment with an accuracy rate of 94.90% when using a dropout value of 0.5, and a batch size of 64 at epoch 3 and using the SMOTE sentiment class balance technique over sampling to the minority class. Meanwhile, the NB algorithm produces the best accuracy in the 10th cross validation with an accuracy of 94%.

The performance of the model in the form of a word cloud from the percentage of word occurrences can also work very well. The Naïve Bayes and Bidirectional LSTM methods can provide an understanding of which method is better at displaying the sentiment analysis results of BSI Mobile users.

Comparison of the results of sentiment analysis performance optimization with big data using the Naïve Bayes and Bidirectional LSTM methods is shown in Figure 13.

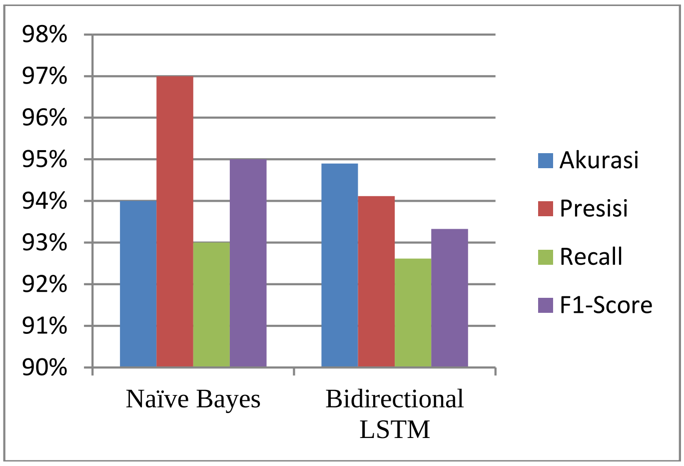


Figure 13. Comparison of Performance Results of Methods

5. CONCLUSION

This study conducted a sentiment analysis of BSI mobile user reviews on the Google Play Store by comparing two algorithms, namely NB and BiLSTM. Based on the results obtained with the sentiment class balance technique, namely SMOTE to the minority class (negative), a dropout value of 0.5 and a batch size of 64 produced superior accuracy of 94.90% compared to NB Cross Validation which is only 94%.

The results of the study show that the performance of the BiLSTM algorithm is better in producing sentiment analysis visualizations compared to the NB algorithm.

This study using Naïve Bayes and Bidirectional algorithms is effective in analyzing the sentiment of BSI Mobile user reviews. The findings of the analysis show the efficacy of the methodology in generating, classifying and differentiating sentiment analysis of reviews that can provide valuable insights into user satisfaction and can present future prospects for improving public services on mobile banking applications from a deeper understanding of user complaints. These findings have significant implications for policy makers in formulating policies to improve services to the BSI Mobile application.

Future work can be recommended to be refined by using more data not only from Google Play Store but also from other media sources. Testing can use different hyperparameter tuning configurations and use different algorithms such as transformer-based models to improve sentiment analysis capabilities.

REFERENCES

[1] R. Adawiyah and J. Nugraha, "Sentiment Analysis on Mobile Banking Application Using Naive Bayes Classifer and Association Methods," *Int. J.Eng. Technol.* , vol. 7, no. 4.15, p. 244, 2018, doi: 10.14419/ijet.v7i4.15.22998.

[2] H. Cam, AV Cam, U. Demirel, and S. Ahmed, "Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers," *Heliyon* , vol. 10, no. 1, p. e23784, 2024, doi: 10.1016/j.heliyon.2023.e23784.

[3] M. Eksa Permana, H. Ramadhan, I. Budi, A. Budi Santoso, and P. Kresna Putra, "Sentiment analysis and topic detection of mobile banking application review," *2020 5th Int. Conf. Informatics Comput. ICIC 2020* , 2020, doi: 10.1109/ICIC50835.2020.9288616.

[4] AA Arifiyanti, NR Shantika, and AO Syafira, "Sentiment Analysis of Bsi Mobile User Reviews on Google Play Using a Supervised Learning Approach," *J. Inform. Polinema* , vol. 9, no. 3, pp. 283–288, 2023 , doi: 10.33795/jip.v9i3.1003.

[5] M. V. Kusuma, S. Juanita, and U. B. Luhur, "COMPARISON OF K-NEAREST NEIGHBORS AND NAÏVE BAYES CLASSIFIER ALGORITHMS IN SENTIMENT ANALYSIS OF USER REVIEWS FOR PERBANDINGAN ALGORITMA K-NEAREST NEIGHBHORS DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI," vol. 5, no. 5, pp. 1433–1441, 2024.

[6] H. Akbar, D. Aryani, M. K. M. Al-shammari, and M. B. Ulum, "SENTIMENT ANALYSIS FOR E-COMMERCE PRODUCT REVIEWS BASED ON FEATURE FUSION AND BIDIRECTIONAL LONG SHORT-TERM MEMORY," vol. 5, no. 4, pp. 1385–1391, 2024.

[7] W. L. Roldan-Baluis, N. A. Zapata, and M. S. M. Vásquez, "The Effect of Natural Language Processing on the Analysis of Unstructured Text: A Systematic Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 5, pp. 43–51, 2022, doi: 10.14569/IJACSA.2022.0130507.

[8] B. Andrian, T. Simanungkalit, I. Budi, and AF Wicaksono, "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," *Int. J. Adv. Comput. Sci. Appl.* , vol. 13, no. 3, pp. 466–473, 2022, doi: 10.14569/IJACSA.2022.0130356.

[9] X. Li, Z. Zhang, and K. Stefanidis, "Sentiment-aware Analysis of Mobile Apps User Reviews Regarding Particular Updates," *J. Icsea* , no. January 2020, p. 9, 2018, [Online]. Available: https://people.uta.fi/~kostas.stefanidis/docs/icsea18.pdf

[10] NT Romadloni, I. Santoso, and S. Budilaksono, "Comparison of Naïve Bayes, KNN and Decision Tree Methods on KRL Commuter Line Transportation Sentiment Analysis," *IKRA-ITH Inform. J. Comput. and Inform.* , vol. 3, no. 2, pp. 1–9, 2019.

[11] H. Wisnu, M. Afif, and Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," *J. Phys. Conf. Ser.*, vol. 1444, no. 1, 2020, doi: 10.1088/1742-6596/1444/1/012034.

[12] D. A. Nurdeni, I. Budi, and A. B. Santoso, "Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 122–127, 2021, doi: 10.1109/EIConCIT50028.2021.9431852.

[13] G. J. Krishna *et al.* , "Sentiment Classification of Indian Banks' Customer Complaints," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON* , vol. 2019-Octob, pp. 429–434, 2019, doi: 10.1109/TENCON.2019.8929703.

[14] AD Adhi Putra, "Sentiment Analysis on User Reviews of Bibit and Bareksa Applications with the KNN Algorithm," *JATISI (Journal of Information Technology and Information Systems)* , vol. 8, no. 2, pp. 636–646, 2021 , doi: 10.35957/jatisi.v8i2.962.

[15] A. Oktian Permana and Sudin Saepudin, "Comparison of k-nearst neighbor and naïve bayes algorithms on the shopee application," *J. CoSciTech (Computer Sci. Inf. Technol.* , vol. 4, no. 1, pp. 25– 32, 2023, doi: 10.37859/coscitech.v4i1.4474.

[16] AI Tanggraeni and MNN Sitokdana, "Sentiment Analysis of E-Government Applications on Google Play Using the Naïve Bayes Algorithm," *JATISI (Journal of Information Technology and Information Systems)* , vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[17] MS Yupi Kuspandi Putra, "Comparison of Naive Bayes and Naive Bayes Algorithms," *J. Inform. and Technol.* , vol. 2, no. 3, pp. 61–69, 2019.

[18] DT Ksatria, Y. Yunefri, and LL Van FC, "Sentiment Analysis of Mypertamina Application User Reviews on Google Playstore Using K-Nearest Neighbor and Naïve Bayes," *Teknol. Inf.* , vol. 2, no. 1, pp. 213–227, 2023, [Online]. Available: https://journal.unilak.ac.id/index.php/Semaster/article/view/18526

[19] S. Mulyani and R. Novita, "Implementation of the Naive Bayes Classifier Algorithm for Classification of Community Sentiment About Depression on Youtube," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1355–1361, 2022, doi: 10.20884/1.jutif.2022.3.5.374.

[20] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis on COVID-19 Public Reviews," *3rd Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2021*, pp. 444–449, 2021, doi: 10.1109/ICAIIC51459.2021.9415191.

[21] M. Arbane, R. Benlamri, Y. Brik, and A. D. Alahmar, "Social media-based COVID-19 sentiment classification model using Bi-LSTM," *Expert Syst. Appl.*, vol. 212, no. June 2022, p. 118710, 2023, doi: 10.1016/j.eswa.2022.118710.

[22] U. B. Mahadevaswamy and P. Swathi, "Sentiment Analysis using Bidirectional LSTM Network," *Procedia Comput. Sci.*, vol. 218, pp. 45–56, 2022, doi: 10.1016/j.procs.2022.12.400.

[23] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: 10.34010/jamika.v12i2.7764.

[24] Y. A. Wubet and K. Lian, "International Journal of Information How can we detect news surrounding community safety crisis incidents in the internet ? Experiments using attention-based Bi-LSTM models," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 1, p. 100227, 2024, doi: 10.1016/j.jjimei.2024.100227.

[25] A. Alzhrani, A. Alatawi, B. Alsharari, and U. Albalawi, "Towards Security Awareness of Mobile Applications using Semantic-based Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp. 800–809, 2022, doi: 10.14569/IJACSA.2022.0130493.

[26] U. Uslu, Ö. D. Incel, and G. I. Alptekin, "Evaluation of Deep Learning Models for Continuous Authentication Using Behavioral Biometrics," *Procedia Comput. Sci.*, vol. 225, pp. 1272–1281, 2023, doi: 10.1016/j.procs.2023.10.115.

[27] M. J. Islam, R. Datta, and A. Iqbal, "Actual rating calculation of the zoom cloud meetings app using user reviews on google play store with sentiment annotation of BERT and hybridization of RNN and LSTM," *Expert Syst. Appl.*, vol. 223, no. March, p. 119919, 2023, doi: 10.1016/j.eswa.2023.119919.

[28] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 728–735, 2021, doi: 10.1016/j.procs.2021.01.061.

[29] M. S. Islam and E. Hossain, "Foreign exchange currency rate prediction using a GRU-LSTM hybrid network," *Soft Comput. Lett.*, vol. 3, no. June 2020, p. 100009, 2021, doi: 10.1016/j.socl.2020.100009.

[30] B. Dangsawang and S. Nuchitprasitchai, "A machine learning approach for detecting customs fraud through unstructured data analysis in social media," *Decis. Anal. J.*, vol. 10, no. February, p. 100408, 2024, doi: 10.1016/j.dajour.2024.100408.

[31] Z. H. Kilimci, H. Yoruk, and S. Akyokus, "Sentiment Analysis Based Churn Prediction in Mobile Games using Word Embedding Models and Deep Learning Algorithms," *INISTA 2020 - 2020 Int. Conf. Innov. Intell. Syst. Appl. Proc.*, 2020, doi: 10.1109/INISTA49547.2020.9194624.