

ANALISIS SENTIMEN REVIEW APLIKASI WETV PADA PLATFORM TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE

¹Vina Alviani, ²Syariful Alam, ³Imay Kurniawan

^{1,2,3} Sekolah Tinggi Teknologi (Program Studi, Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana Purwakarta)

¹vinaalviani02@wastukencana.ac.id, ²syarifulalam@wastukencana.ac.id,

³imaykurniawan@wastukencana.ac.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 28 Juli 2023

Disetujui : 31 Juli 2023

Kata Kunci :

Twitter, Media Sosial, Support Vector Machine, Streaming

ABSTRAK

Twitter salah satu media sosial, yang memiliki fungsi untuk send pesan, sering disebut dengan kicauan atau (*tweet*). Penggunaan *twitter* di Indonesia mencapai 59%, dengan peringkat ke-5 media sosial yang sering digunakan tahun 2020. Dari data tersebut, maka dapat dipastikan bahwa *platform twitter* menjadi media sosial yang memiliki pengaruh bagi para pengguna di Indonesia. Selain menggunakan *twitter*, masyarakat indonesia juga menyukai hiburan di waktu senggang mereka, dengan berkegiatan seperti membaca, bermain, bahkan menonton bersama dengan keluarga secara streaming pada aplikasi. Semakin banyak pengguna dari WeTV maka banyak juga ulasan pada aplikasi tersebut dan dapat memberikan pengaruh untuk WeTV dalam perbaikan terhadap aplikasi, untuk mengatasi permasalahan yang ada maka dilakukannya analisis sentimen terhadap data *review* pada aplikasi WeTV dengan metode *Support Vector Machine*. Dengan data baru yang terdapat di situs *Google Play Store*. Berdasarkan hasil penelitian mengenai analisis sentimen pengguna aplikasi WeTV, dengan jumlah 4024 data yang sudah melewati proses *cleaning, labelling, preprocess text, filtering, tokenization, transformation* dan klasifikasi menggunakan algoritma *Support Vector Machine*, dan dilanjutkan dengan evaluasi data, dengan *confusion matrix* menggunakan *tools Google Collab* dapat disimpulkan bahwa hasil dari ulasan *tweet* memiliki tingkat persentase sebesar 89% pada akurasi (*accuration*), dan memiliki nilai 87% pada hasil *precision*, dan tingkat keberhasilan (*recall*) sebesar 83%. Dapat disimpulkan bahwa algoritma *SVM* memiliki kinerja yang cukup baik. Analisis sentimen pengguna *twitter* terhadap aplikasi WeTV menghasilkan 1624 sentimen positif dan 2400 sentimen negatif. Berdasarkan nilai tersebut dapat diketahui bahwa sentimen masyarakat mengenai aplikasi WeTV di media sosial *twitter*, tergolong negatif dan pengguna masih kurang puas dengan aplikasi karena masih banyaknya iklan, aplikasi yang sering mengalami *error*, hingga *translate* yang masih belum terupdate secara otomatis.

ARTICLE INFO

Article History :

Received : July 28, 2023

Accepted : July 31, 2023

Keywords:

Twitter, Media Social, Support
Vector Machine, Streaming.

ABSTRACT

Twitter is one of the social media, which has a function to send messages, often referred to as tweets or (tweets). The use of Twitter in Indonesia has reached 59%, with the 5th most frequently used social media ranking in 2020. From this data, it can be ascertained that the Twitter platform is a social media that has influence on users in Indonesia. Apart from using Twitter, Indonesian people also like entertainment in their spare time, with activities such as reading, playing, and even watching with family by streaming on the application. The more users from WeTV, the more reviews on the application and can influence WeTV in improving the application, to overcome existing problems, sentiment analysis is carried out on data reviews on the WeTV application using the Support Vector Machine method. With new data found on the Google Play Store site. Based on the results of research on sentiment analysis of WeTv application users, with a total of 4024 data that has gone through the process of cleaning, labeling, preprocessing text, filtering, tokenization, transformation and classification using the Support Vector Machine algorithm, and followed by data evaluation, with a confusion matrix using Google Collab tools, it can be concluded that the results of tweet reviews have a percentage level of 89% in accuracy, and has a value of 87% in precision results, and a success rate (recall) of 83%. It can be concluded that the SVM algorithm has a pretty good performance. Twitter user sentiment analysis for the WeTv application resulted in 1624 positive sentiments and 2400 negative sentiments. Based on this value, it can be seen that public sentiment regarding the WeTv application on Twitter social media is classified as negative and users are still not satisfied with the application because there are still many advertisements, applications that often experience errors, and translations that are still not updated automatically.

1. PENDAHULUAN

Twitter merupakan media sosial yang didirikan oleh Jack Dorsey, yang biasanya digunakan untuk *send* pesan yang disebut kicauan (*tweet*). *User Twitter* secara umum digunakan oleh pengguna Indonesia sebesar 59%, sebagai media sosial yang paling banyak digunakan ke-5 (lima) pada tahun 2020. Berdasarkan data, *Twitter* merupakan media sosial yang berdampak besar bagi pengguna Indonesia (Krisdiyanto, 2021).

Streaming merupakan salah satu aktivitas hiburan yang melelahkan yang bisa dilakukan atau ditonton bersama keluarga dan teman dekat. Banyak aplikasi tontonan meningkatkan minat orang untuk menonton film di rumah melalui *smartphone* karena menghemat waktu dan mudah tersedia di mana saja. Ini meningkatkan kemungkinan bagi pengguna *program streaming* seperti WeTV. WeTV adalah *platform video-on-demand* Cina yang memungkinkan pengguna menonton *online*. Sejak peluncuran internasional pertamanya pada Juni 2019, sukses diluncurkan di Indonesia, dan beberapa negara di Asia Tenggara. WeTV adalah salah satu *platform streaming* pertama yang menghadirkan konten dari China ke pengguna lokal di luar negeri, dan sejak itu terus berkembang untuk menawarkan konten Asia berkualitas lebih tinggi di semua genre (Al et al., 2022).

Berdasarkan penelitian sebelumnya dengan menggunakan metode *SVM*, sebaiknya digunakan sebagai *method* klasifikasi dalam proses analisis sentimen ulasan teks bahasa Indonesia di *Play Store*. Hal ini dapat dilihat dengan tingkat akurasi yang diperoleh metode *SVM* lebih tinggi 81,46%, berbanding dengan metode klasifikasi *Naive Bayes* dengan akurasi sebesar 51,4%. Pada penelitian lain, metode *SVM* menunjukkan tingkat akurasi yang lebih besar dibandingkan *method Naive Bayes* (*NB*). Nilai yang diperoleh algoritma *SVM* adalah 81,22 dan akurasi algoritma *NB* adalah 74,37%. Dengan hasil evaluasi analisis opini *Twitter*, berbanding dengan *method SVM* dan metode *JST* memiliki akurasi lebih tinggi yaitu 89,7%. Dapat diartikan bahwa *method SVM* lebih tinggi akurasinya berbanding dengan *method* lain. Dengan demikian, penelitian ini digunakanlah algoritma *SVM* untuk

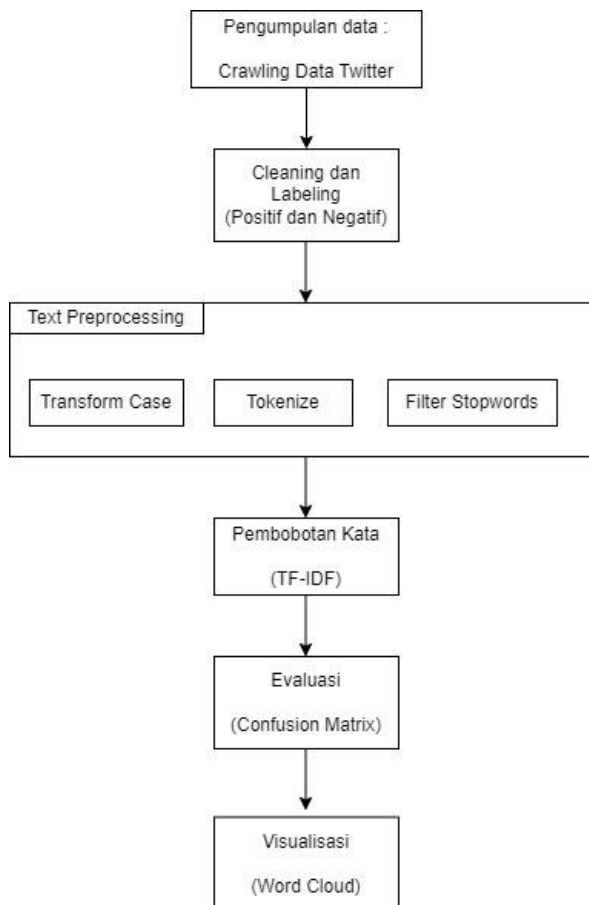
mengklasifikasikan data audit (Kulsum, Jajuli and Sulistiyowati, 2022).

Penelitian lainnya yang serupa, algoritma *SVM* memiliki hasil analisis sentimen *tweet* TIX-ID disimpulkan dengan tingkat akurasi yang dicapai oleh algoritma *support vector machine* dengan kernel *point* dan *scenario* 80 data *sharing*: 20 memberikan nilai sebanyak 74,17%. Dapat disadari bahwa algoritma *SVM* mendapatkan nilai yang cukup baik. Analisis pendapat user *Twitt* TIX ID mendapat 146 pendapat baik dan 231 pendapat kurang baik. Nilai tersebut menunjukkan bahwa pendapat umum aplikasi TIX-ID di jejaring sosial *twitter* dinilai kurang baik, dan pengguna belum puas akan aplikasi tersebut di karenakan jadwal *update* yang tidak sesuai, aplikasi selalu melakukan kesalahan & kembali. Dalam studi selanjutnya, proses pengumpulan data mungkin membutuhkan waktu tambahan. Dengan harapan dapat menganalisis menggunakan *method* lain atau membandingkan materi penilaian dengan algoritma lain agar memperbaikinya (Nabillah, Alam and Resmi, 2022).

Berdasarkan penerapan lima fungsi TF-IDF yang berbeda dengan metode pembobotan menggunakan metode *multi-class support vector machine* @KAI121 untuk mengklasifikasikan data akun *tweet* untuk analisis sentimen, peneliti dapat menyimpulkan bahwa akurasi tertinggi yang diperoleh dengan metode *SVM* multi-kelas OAA dalam analisis sentimen diperoleh pada rasio 90:10 menggunakan plot unigram, bobot TF-IDF dan nilai parameter gamma 0.7 sebesar 80.59 (Dhina Nur Fitriana and Yuliant Sibaroni, 2020).

2. METODE

Penelitian ini di lakukan dengan menggunakan beberapa tahapan, mulai dari data yang terkumpul, *Cleaning*, *Labeling*, *Text Preprocessing*, Pembobotan Kata, Evaluasi, dan Visualisasi yang dapat dilihat pada gambar 1 dibawah ini.



Gambar 1 Kerangka Kerja Penelitian

3. HASIL DAN PEMBAHASAN

Hasil yang didapatkan oleh penulis ini, pengambilan data yang dilakukan dengan memakai *windows* dan *python* (Bahasa pemrograman). Data di *cleaning* dan *labelling* secara manual menggunakan *microsoft excel* berjumlah 4024 data yang dimana menghasilkan 1624 sentimen positif dan 2400 sentimen negatif. Selanjutnya dilakukan tahap *text preprocessing* sehingga dapat menghitung proses *confusion matrix* dan dapat menghasilkan *wourd cloud*.

3.1 Pengumpulan Data

Pengambilan data dilakukan dengan *windows* dan di program bahasa *python*, dengan menggunakan *Auth Token Twitter* yang sudah ada di halaman inspect setelah membuka *twitter* di *Google Chrome*. Dataset yang didapatkan sebanyak 5000 data, yang ditemukan pada gambar 2.

```

    C:\SemuelPras> python C:\Users\SemuelPras\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\.../twitter_crawler.py
    Welcome to the Twitter Crawler
    This script uses Chromium Browser to crawl data from Twitter with your Twitter auth token.
    Please enter your Twitter auth token when prompted.
    What's your Twitter auth token? *****
    What's the search keyword? ...
    How many tweets do you want to crawl? 1000
    [Enter any key to start] | IdealTree-User: 1110 IdealTree buildDev
    Changed 2 packages, and audited 4 packages in 35s
    Found 0 vulnerabilities
    Running Uninstall Browser at C:\Users\SemuelPras\AppData\Local\Temp\playwright\chromium-1806
    Downloading Chromium 115.0.5790.76 (playwright build v1077) from https://storage.googleapis.com/chromium-build-logs/115.0.5790.76
    134.3 Mb [=====] 100% 8.25
    Chromium 115.0.5790.76 (playwright build v1077) downloaded to C:\Users\SemuelPras\AppData\Local\Temp\playwright\chromium-1871
    
```

Gambar 2 Source Code Proses Crawling

3.2 Cleaning dan Labelling

Dalam tahap ini menghilangkan beberapa komponen atau fitur yang terdapat pada *twitter* seperti (*created_at*), (*id*), (*id_str*), (*full_text*), (*quote_count*), (*reply_count*), (*retweet_count*), (*favorite_count*), (*geo*), (*lang*), (*user_id_str*), (*conversation_id*), (*conversation_id_str*), (*media_url_https*), (*media_type*), (*username*). Kemudian di ambil informasi yang terkandung di dalam nya untuk mengelola data, diberikan analisa sentiment secara manual dengan cara membaca, sehingga menentukan nilai yang bernilai positif, negatif, baik berupa saran maupun keluhan. Sehingga menghasilkan 1624 sentimen positif dan 2400 sentimen negatif.

3.3 Text Preprocessing

Untuk tahap ini akan dilakukannya *Transform Case*, *Tokenize*, dan *Filter Stopwords*.

1. Transform Case

Dilakukannya perubahan huruf besar menjadi huruf kecil yang dapat di lihat pada tabel 1.

Tabel 1 Hasil Transform Case

Tweet	Transform Case
Kebanyakan iklan2 gabisa skip, jadi gak enak nonton	kebanyakan iklan2 gabisa skip, jadi gak enak nonton

2. Tokenize

proses ini dilakukan untuk pemecahan teks menjadi kata tunggal, menghilangkan tanda baca, dan angka, dapat dilihat pada tabel 2.

Tabel 2 Hasil *Tokenize*

Kebanyakan	iklan	gabisa	Skip
jadi	gak	enak	nonton

3. *Filter Stopwords*

Proses selanjutnya dilakukan dengan tujuan menghilangkan kata yang tidak dibutuhkan pada proses berikutnya yang dapat dilihat pada tabel 3.

Tabel 3 Hasil *Stopwords*

<i>Tweet</i>	<i>Filter Stopwords</i>
Kebanyakan iklan2 gabisa skip, jadi gak enak nonton	banyak iklan gabisa skip gak enak nonton

3.4 *Pembobotan Kata*

TF-IDF memiliki fungsi membilang frekuensi kemunculan kata di dokumen, pada tahap ini merubah kata dalam bentuk 0 (numerik). *Term Frequency (TF)* merupakan perhitungan besar kemunculan kata pada sebuah dokumen, makin sering nilai bermunculan maka akan menghasilkan nilai besar untuk nilai bobot kata tersebut. *IDF (Inverse Document Frequency)* yaitu semakin >frekuensi (besar) kata hadir maka nilai bobot kata akan semakin< (kecil). Persamaan *TF-IDF* yang digunakan pada penelitian:

$$W_{i,j} = t_{fi,j} \log \left(\frac{N+1}{df_{i+1}} \right) + 1$$

Keterangan:

$W_{i,j}$ = sebanyak data ke-i terhadap kata ke-j

$T_{fi,j}$ = sebanyak kata i yang dicari pada sebuah data j N = total data

df_i = banyak data yang mengandung kata ke-i (Aziz, 2022).

Proses ini menggunakan data hasil tahap *preprocessing*. Pada tahap *term frequency* semakin besar jumlah kemunculan suatu kata, maka semakin besar bobotnya pada dokumen. Kata yang tidak muncul bernilai 0, dan jika ada bernilai 1 yang dapat dilihat pada tabel 4.

Tabel 4 Hasil *Term Frequency*

term	TF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
terimakasih	1	0	0	0	1	0	0	0	0	0
ih aplikasi	0	1	1	1	1	1	1	1	1	1
kualitas	0	1	1	0	0	0	0	0	0	0
bagus	0	1	0	0	1	0	0	0	0	0
keluar	0	0	0	0	0	1	0	1	0	0
vip	0	0	0	0	1	0	0	0	1	1
logout	0	0	0	0	0	0	0	0	1	1
lama	0	0	0	0	0	0	1	1	0	0

Langkah selanjutnya adalah menghitung berapa banyak kata yang sering muncul dalam sebuah dokumen, yaitu. menghitung *Kepadatan. Dokumen*. Seluruh dokumen dibagi dengan beberapa nilai untuk mendapatkan kerapatan dokumen terbalik yang ditunjukkan pada tabel 5.

Tabel 5 Hasil *Inverse Document*

term	TF										df	D/DF	idf (log(d/df))	1+Df
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10				
terimakasih	1	0	0	0	1	0	0	0	0	0	2	5	0.698970004	1.69897
aplikasi	0	1	1	1	1	1	1	1	1	1	9	1.1111111	0.045757491	1.045757
kualitas	0	1	1	0	0	0	0	0	0	0	2	5	0.698970004	1.69897
bagus	0	1	0	0	1	0	0	0	0	0	2	5	0.698970004	1.69897
keluar	0	0	0	0	0	1	0	1	0	0	2	5	0.698970004	1.69897
vip	0	0	0	0	1	0	0	0	1	1	3	3.3333333	0.522878745	1.522879

Selanjutnya hitung nilai dari *term frequency inverse document frequency (TF-IDF)* dengan mengalikan tiap nilai dari kata pada dokumen *TF (term frequency)* dengan nilai *IDF*, hasil tersebut yang dapat dilihat pada gambar 3.

term	W=rf (idf=1)									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
terimakasih	1.69897	0	0	0	1.69897	0	0	0	0	0
aplikasi	0	1.045757	1.045757	1.045757	1.045757	1.045757	1.045757	1.045757	1.045757	1.045757
kuantitas	0	1.69897	1.69897	0	0	0	0	0	0	0
bagus	0	1.69897	0	0	1.69897	0	0	0	0	0
kebur	0	0	0	0	0	1.69897	0	1.69897	0	0
vip	0	0	0	0	1.522879	0	0	0	1.522879	1.522879
lagu	0	0	0	0	0	0	0	0	1.69897	1.69897
lama	0	0	0	0	0	0	1.69897	1.69897	0	0
	1.69897	4.443897	2.744727	1.045757	5.966576	2.744727	2.744727	4.443897	4.267966	4.267966

Gambar 3 Hasil Term Frequency - Inverse Document Frequency (TF-IDF)

3.5 Klasifikasi Algoritma Support Vector Machine

SVM adalah *method* klasifikasi yang menggunakan pembelajaran *machine* dan memprediksi kelas menurut model dari hasil pelatihan yang dihasilkan. Keakuratan model yang dihasilkan oleh proses sangat bergantung pada karakteristik dan parameter kernel. Karena sifatnya, ada SVM linier dan SVM non-linier. Linear SVM memisahkan 2 (dua) kelas di *hyperplane* dengan batas *soft*. Sebaliknya, SVM non linier memiliki fungsi trik kernel ke ruang dimensi tinggi ini (Tineges, Triayudi and Sholihati, 2020).

Pada tahap ini dilakukan proses training model menggunakan *google colaborartory* dan bahasa pemrograman *python*. Membutuhkan Kernel dan nilai *c* sebagai parameter, dimana kernel di isi dengan *rbf* dan *c* diisi dengan 1. Setelah itu SVM membutuhkan data *x train* dan *y train* untuk melatih model. Dapat dilihat pada gambar 4.

```
[ ] #Training Model

from datetime import datetime
start_time = datetime.now()
from joblib import dump
#algoritme fitting

text_algorithm = svm.SVC(kernel="rbf", C=1.0)

model = text_algorithm.fit(x_train, y_train)

# save the model to disk
dump(model, filename="model_sentiment_naive.joblib")

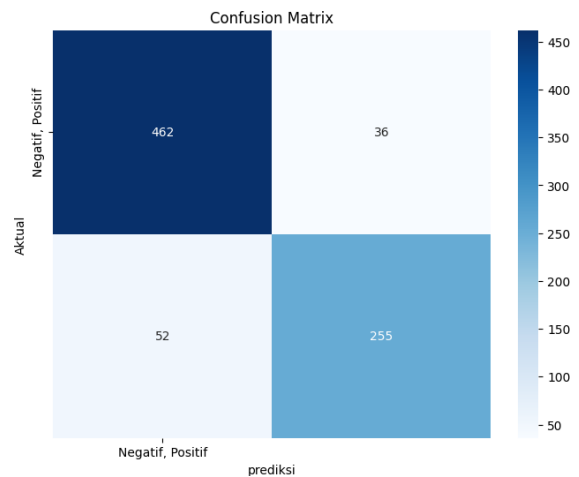
end_time = datetime.now()
result_time = end_time-start_time
print("Duration:",result_time)

Duration: 0:00:17.434908
```

Gambar 4 Training Model SVM

3.6 Evaluasi

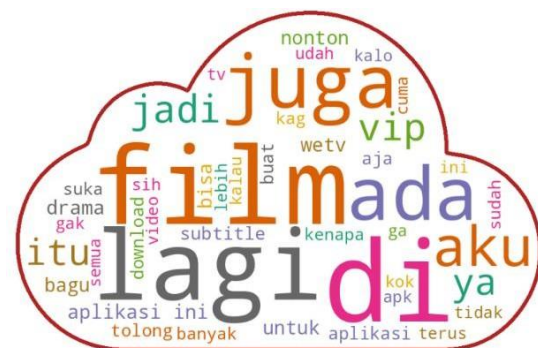
Hasil evaluasi pada penelitian ini menggunakan metode *cross validation* yang akan diukur dengan *Confusion Matrix* agar mengetahui hasil akurasi, presisi, *recall*. Hasil pengujian keseluruhan data *tweet* dengan menggunakan *tools Google Colab*, dapat ditelaah pada gambar 6.



Gambar 5 Confusion Matrix

3.7 Visualisasi

Visualisasi atau tampilan dalam pembuatan gambar, diagram atau animasi untuk menunjukkan suatu informasi yang digambarkan dengan *word cloud* pada sentimen seluruh dokumen positif dan negatif yang dapat dilihat pada gambar 6 dan gambar 7.



Gambar 6 Word Cloud Sentimen Positif



Gambar 7 Word Cloud Sentimen Negatif

4. PENUTUP

4.1. Kesimpulan

Hasil penelitian mengenai analisis sentimen pengguna aplikasi WeTv sebanyak 4024 data, yang telah melalui proses *cleaning* dan *labelling*, *preprocess text* yang terdiri dari *filtering*, *tokenization*, *transformation* dan klasifikasi menggunakan algoritma *Support Vector Machine* serta evaluasi data dengan *confusion matrix* menggunakan tools *Google Colab* dapat disimpulkan, bahwa hasil dari ulasan *tweet* dengan hasil persentase 89% pada akurasi (*accuration*), 87% pada nilai *precision*, dan tingkat keberhasilan (*recall*) 83%. Kesimpulannya yaitu, algoritma *Support Vector Machine* memiliki kinerja yang cukup baik. Sehingga analisis sentimen pengguna *twitter* terhadap aplikasi WeTv mendapatkan 1624 sentimen baik (*positif*) dan 2400 sentimen tidak baik (*negative*). Berdasarkan tingkat akurasi yang di dapat membuktikan bahwa sentimen masyarakat mengenai aplikasi WeTv di *twitter* termasuk kurang baik dan user kurang puas, dikarenakan adanya iklan yang sering muncul, kesalahan aplikasi (*error*), dan *translate* belum terupdate.

4.2. Saran

Pada penelitian selanjutnya diharapkan dapat melakukan pengambilan data dengan menambah waktu dalam pengumpulan data agar mendapat hasil yang lebih bagus. Bagi penelitian selanjutnya, sebaiknya menggunakan algoritma-algoritma klasifikasi lain dengan melakukan perbandingan antar klasifikasi tersebut, sehingga dapat mengetahui algoritma dengan kinerja terbaik. Dapat juga menggunakan algoritma *support vector machine* yang analisis sentimen pada penelitian ini, baik berupa sentimen positif maupun sentimen negatif dapat dijadikan sebagai bahan evaluasi untuk dipertahankan

ataupun lebih ditingkatkan di masa yang akan datang.

5. DAFTAR PUSTAKA

Al, A. *et al.* (2022) 'Pendahuluan Metode Penelitian', 21, pp. 579–586.

Aziz, A. (2022) 'Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength', *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), p. 115.

Dhina Nur Fitriana and Yuliant Sibaroni (2020) 'Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM)', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(5), pp. 846–853. Available at: <https://doi.org/10.29207/resti.v4i5.2231>.

Krisdiyanto, T. (2021) 'Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers', *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 7(1), p. 32. Available at: <https://doi.org/10.24014/coreit.v7i1.12945>.

Kulsum, U., Jajuli, M. and Sulistiyowati, N. (2022) 'Analisis Sentimen Aplikasi WETV di Google Play Store Menggunakan Algoritma Support Vector Machine', *Journal of Applied Informatics and Computing*, 6(2), pp. 205–212. Available at: <https://doi.org/10.30871/jaic.v6i2.4802>.

Nabillah, A., Alam, S. and Resmi, M.G. (2022) 'Twitter User Sentiment Analysis Of TIX ID Applications Using Support Vector Machine Algorithm', *RISTEC : Research in Information Systems and Technology*, 3(1), pp. 14–27. Available at: <https://doi.org/10.31980/ristec.v3i1.1898>.

Tineges, R., Triayudi, A. and Sholihati, I.D. (2020) 'Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)', *Jurnal Media Informatika Budidarma*, 4(3), p. 650. Available at: <https://doi.org/10.30865/mib.v4i3.2181>.