

ANALISIS SENTIMEN TERKAIT OPINI MASYARAKAT TERHADAP PERKEMBANGAN E-SPORT MOBILE DI INDONESIA MENGGUNAKAN K NEAREST NEIGHBOR

Muhammad Kega Alamsyah¹⁾, Nunik Pratiwi^{*2)}

1. Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia
2. Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia

Article Info

Kata Kunci: E-sport; Game Online; K-Nearest neighbor; Analisis Sentimen; Twitter

Keywords: E-sport; Game Online; K-Nearest neighbor; *Sentimen Analysis*; Twitter

Article history:

Received 2 December 2023

Revised 16 December 2023

Accepted 30 December 2023

Available online 1 March 2024

DOI :

<https://doi.org/10.29100/jipi.v9i1.4927>

* Corresponding author.

Nunik Pratiwi

E-mail address:

npratiwi@uhamka.ac.id

ABSTRAK

Fenomena *E-sport* mobile belakangan ini menjadi perbincangan pada kalangan masyarakat Indonesia. Banyak masyarakat berpendapat bahwa *E-sport* mobile ini merupakan sebuah bentuk hiburan baru dan menarik serta juga dapat membuka peluang bagi anak-anak muda yang ingin berkembang ke dunia profesional. Tetapi selain pendapat positif pasti ada juga pendapat negatif tentang *E-sport* mobile ini, banyak pendapat ini juga berasal dari para orang tua. Analisis sentimen berguna untuk mengolah data dari opini masyarakat luas dan mengklasifikasi opini atau pendapat tersebut ke dalam bentuk positif dan negatif. Penelitian ini bertujuan untuk mengetahui bagaimana sentimen masyarakat terhadap perkembangan *E-sport mobile* di Indonesia, apakah bidang ini mengarah ke hal yang positif atau negatif. Pada penelitian ini data didapat dari tweet pada media sosial twitter, dan menggunakan metode klasifikasi *K-Nearest Neighbor*. Hasil sentimen yang didapat dari penelitian adalah didominasi oleh sentimen positif. Pada proses Klasifikasi dibagi dua skenario yaitu pembagian data 70:30 dan 80:20, didapat hasil yang lebih optimal pada skenario dua yaitu pembagian data dengan rasio 80:20 dengan akurasi sebesar 84,66% dan hasil presisi, recall, dan f1-score yang lebih optimal dibandingkan dengan rasio 70:30 pada nilai uji $K=5$.

ABSTRACT

The mobile E-sports phenomenon has recently become a topic of conversation among Indonesian people. Many people think that mobile E-sports is a new and interesting form of entertainment and can also open up opportunities for young people who want to develop into the professional world. But apart from positive opinions, there are also negative opinions about mobile E-sports, many of these opinions also come from parents. Sentiment analysis is useful for processing data from public opinion and classifying these opinions into positive and negative forms. This research aims to find out how public sentiment is towards the development of mobile E-sports in Indonesia, whether this field is leading to something positive or negative. In this study, data was obtained from tweets on Twitter social media, and used the K-Nearest Neighbor classification method. The sentiment results obtained from the research are dominated by positive sentiment. In the classification process for part two of the scenarios, namely data sharing 70:30 and 80:20, more optimal results were obtained in scenario two, namely data sharing with a ratio of 80:20 with an accuracy of 84.66% and precision, recall and f1-score results. which is more optimal compared to the ratio of 70:30 at the test value $K=5$.

I. PENDAHULUAN

FENOMENA *E-sport* belakangan ini sedang ramai di masyarakat. Banyak developer yang berlomba-lomba menggunakan *E-sport* untuk meramaikan *game* buatan mereka, mulai dari *game* olahraga, *FPS*, sampai *fighting* [1]. *E-sport* adalah olahraga dengan peralatan elektronik seperti konsol *game*, komputer, dan handphone. Bidang ini memberikan gambaran bahwa sebuah *game* juga bisa menjadi kompetitif dan terorganisir [2]. Jenis olahraga ini sangat populer tidak hanya dikalangan anak-anak dan remaja, tetapi juga orang dewasa bahkan orang tua [3]. Banyak masyarakat kita yang juga memainkan *game online*, hal ini yang membuat *E-sport* muncul dan berkembang pesat hingga seperti sekarang [4]. Perkembangan teknologi yang sudah pesat ini membuat

olahraga tidak lagi tentang fisik tetapi juga bisa menggunakan teknologi atau bisa disebut dengan *E-sport*. Fenomena *E-sport mobile* juga tidak dapat diabaikan pada masa kini, *games* bukan hanya permainan belaka lagi, tetapi sekarang bisa dipertandingkan secara resmi dan sudah masuk dalam kategori *E-sport* [5]

Banyak pihak yang mendukung berkembangnya *E-sport mobile* ini karena dianggap sebagai hiburan yang menarik dan menyenangkan. Tapi ada juga yang beranggapan bahwa *E-sport mobile* ini menyebabkan sifat kecanduan dan merugikan bagi yang memainkannya, mulai dari segi kesehatan hingga efek sosial dan emosional [6]. Berdasarkan [7], Indonesia berada di peringkat ketiga dalam jumlah pemain video *game* di dunia. Berdasarkan [8] Indonesia mempunyai 2.375 pemain aktif *E-sport* yang sudah mendapat hadiah uang pada sebanyak 1.074 turnamen dengan nilai 16.242.356,25 USD. Dengan *game* yang menjadi penghargaan tertinggi adalah *game mobile* yaitu Free Fire sebesar 3,712,043.57 USD dengan disusul dengan *game mobile* lainnya seperti MLBB, dan PUBGM. Ini membuktikan bahwa *E-sport mobile* di Indonesia memiliki banyak peminat dan juga berdampak kepada perekonomian di Indonesia. Masalah kesehatan terkait penikmat internet yang sudah berlebihan juga permainan *online* sudah menjadi perhatian di dunia. Kecanduan *game* disebut sebagai kecanduan internet dalam beberapa literatur [9].

Analisis sentimen merupakan proses analisis opini seseorang berdasar peristiwa yang sedang terjadi, baik itu positif atau negatif. Biasanya pada analisis sentimen, nilai pada analisis sentimen dibagi menjadi tiga yaitu positif, negatif, dan netral. Sehingga dalam proses pembagian data dapat ditemukan sentimen yang positif dan sentimen negatif [10]. Tujuan analisis sentimen untuk menilai emosi dan pendapat terhadap suatu produk, atau layanan. Analisis sentimen banyak digunakan karena memiliki keuntungan dalam kecepatan pengolahan data karena analisis sentimen ini dapat dilakukan secara otomatis sehingga tidak memakan waktu yang lama [11].

Algoritma *K-Nearest Neighbor* adalah algoritma yang mengklasifikasi data berdasarkan tetangga yang terdekat [12]. Metode *K-Nearest Neighbor* melakukan klasifikasi dengan cara menghitung jarak terdekat atau kemiripan objek berdasarkan data training terhadap objek tersebut. *K-Nearest Neighbor* memiliki kelebihan dalam hal akurasi yang baik dalam perhitungan yang sudah diterapkan. Pada perbandingan antara *K-Nearest Neighbor* dan *naïve bayes* dalam penelitian ini, hasil akurasi yang didapat algoritma *naïve bayes* sebesar 75,30% dan *K-Nearest Neighbor* sebesar 86,09%. Ini memberi hasil bahwa metode *K-Nearest Neighbor* lebih akurat dalam hal klasifikasi pada penelitian ini [13]. Algoritma *K-Nearest Neighbor* ini adalah salah satu algoritma yang sederhana karena dalam hal klasifikasinya mampu memberikan hasil yang optimal dan juga signifikan. Algoritma ini mudah dipahami dan mudah dipelajari, dan juga efektif pada data pelatihan yang besar [13].

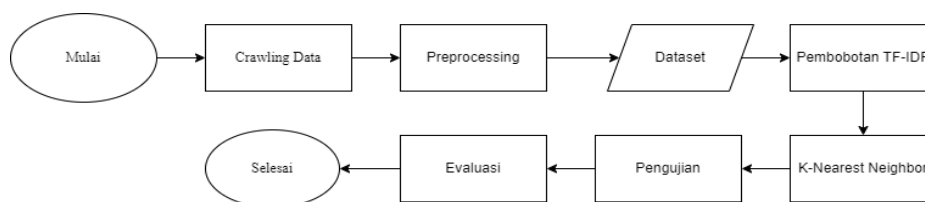
Dalam penelitian [11], peneliti mencari nilai sentimen dan akurasi dari *feedback* layanan gojek melalui X(Twitter). setelah dilakukan klasifikasi menggunakan algoritma *K-Nearest Neighbor*, didapat akurasi sebesar 79,43% dengan nilai K yang diuji yaitu K= 15. Dalam penelitian selanjutnya [15] mengenai analisis sentimen *E-sport* sebagai kurikulum pendidikan dengan bantuan module smote hasil akurasi klasifikasi *naïve bayes* sebesar 70.32%, sedangkan untuk support vector machine sebesar 66.92%, dari hasil akurasi kedua klasifikasi tersebut dapat disimpulkan bahwa kurikulum *E-sports* dalam pendidikan belajar siswa bisa membawa siswa ke arah yang lebih baik. Pada penelitian [10], menggunakan algoritma *K-Nearest Neighbor*, yang bertujuan untuk mencari pendapat positif atau negatif pada media sosial facebook. Dari 200 data didapat sentimen negatif lebih dominan, dan saat klasifikasi data menggunakan RapidMiner didapat akurasi sebesar 87%. Selanjutnya dalam penelitian [16], di penelitian ini peneliti menggunakan metode latent dirichlet allocation, tujuan penelitian ini adalah untuk mengetahui ulasan video *game* lokal yang ada pada steam. Menggunakan rapidminer dan RStudio didapat hasil akurasi sebesar 75.45% dan di dominan oleh sentimen positif sebesar 69.8%. Disebutkan dalam penelitian [17] tentang pandangan masyarakat tentang vaksin Covid-19. Dengan menggunakan algoritma *K-Nearest Neighbor*, klasifikasi KNN dilakukan dan didapat nilai akurasi sebesar 79,25% dengan *K-fold* 10 dan KNN 5, dengan didominasi oleh sentimen positif.

Berdasarkan penjelasan tentang penelitian terkait diatas, alasan peneliti menggunakan algoritma KNN karena dari segi akurasi lebih optimal dibanding jenis algoritma yang lain seperti *naïve bayes*, *Latent Dirichlet Allocation*, *Support Vector Machine* dan algoritma *K-Nearest Neighbor* dapat mudah dipelajari dan dipahami. Tujuan dari penelitian ini adalah untuk mengetahui bagaimana tanggapan masyarakat Indonesia terhadap perkembangan *E-sport mobile* di Indonesia apakah positif atau negatif berdasarkan dari *tweet* dari X(Twitter) dan dianalisis dengan algoritmat *K-Nearest Neighbor*. Penelitian ini diharapkan dapat membantu peneliti lain yang ingin membahas tentang *E-sport* sebagai referensi dan juga bisa melengkapi literatur terkait yang sudah ada sebelumnya.

II. METODE PENELITIAN

A. Alur Penelitian

Alur penelitian akan menjelaskan bagaimana tahapan yang akan dilakukan pada penelitian ini. Alur penelitian dapat dilihat dengan gambar seperti pada gambar 1:



Gambar. 1. Alur Penelitian

B. Crawling Data

Tahap pertama adalah *crawling data*, tahap ini merupakan teknik mengumpulkan data dari suatu *database* seperti contohnya adalah dari X(Twitter) yang hasilnya adalah user dan *tweet* dari *user* tersebut. [18]. Data *tweet* diambil menggunakan *python* dengan *module snsrape*, module ini memudahkan proses *crawling data* karena tidak lagi membutuhkan API dari X(Twitter). Data *tweet* adalah data dunia nyata yang dapat digunakan untuk melacak tren dan perilaku konsumen secara cepat dan akurat. Data ini dapat memberikan gambaran yang lebih akurat tentang perilaku konsumen daripada data yang dikumpulkan melalui survei atau penelitian lain. [19].

C. Preprocessing

Di tahapan ini adalah tahapan yang biasanya dilakukan setelah mendapatkan data dari proses *crawling*, tahapan *preprocessing* dilakukan untuk mengolah teks yang sudah diambil dari *tweet user*. Karena data mentah yang didapat dari proses sebelumnya itu terkadang masih ada penggunaan kata yang kurang dan karakter yang tidak diperlukan dalam proses klasifikasi [20]. Tahapan *preprocessing* teks:

1. *Cleansing*, tahap menghapus karakter yang tidak diperlukan pada analisis, karakter yang dihilangkan adalah karakter yang tidak diperlukan yang menyebabkan kekeliruan [21].
2. *Case folding*, tahapan ini menghilangkan huruf kapital yang ada pada teks *tweet* menjadi huruf kecil sehingga pada *tweet* semua teks menjadi huruf kecil [22].
3. *Tokenizing*, tahap ini membagi teks kalimat menjadi beberapa bagian yang dipisahkan oleh spasi dan koma menjadi seperti token, tahapan ini dilakukan untuk analisis teks [23].
4. *Stopword removal*, menghapus kata umum yang muncul dalam jumlah besar dalam teks dan biasanya tidak memiliki makna [24].
5. *Stemming*, proses ini menghilangkan *sufiks*, *prefix*, *konfiks* sehingga nanti yang didapat adalah kata dasar [25].

D. Pembobotan TF-IDF

Term frequency-inverse document frequency atau *tf-idf* adalah metode yang menghitung bobot kata berdasarkan frekuensi kemunculan kata tersebut dalam suatu dokumen dan frekuensi kemunculan kata tersebut dalam seluruh dokumen. Metode ini ada dari gabungan *term frequency* dan *inverse document frequency*. Dimana *term frequency* adalah banyaknya jumlah *term* yang muncul pada dokumen, jumlah munculnya juga berbanding lurus dengan pembobotan. Lalu *inverse document frequency* digunakan untuk mengukur relevansi kata terhadap suatu dokumen [26].

$$w_{i,j} = t f_{i,j} \times \log\left(\frac{N}{d f_i + 1}\right) \quad (1)$$

$Tf_{i,j}$ = banyak kata-i pada dokumen-j

N = total dokumen

df_i = banyak dokumen yang mengandung kata ke-i

E. K-Nearest Neighbor

K-Nearest Neighbor adalah suatu metode klasifikasi yang dalam penentuan hasil klasifikasinya, *K-Nearest Neighbor* ini melihat jarak terdekat objek dengan masing-masing kelompok. Metode ini juga terbilang mempunyai kelebihan dalam hal tingkat akurasi yang tinggi [27]. Dengan menggunakan dua skenario pembagian data dan nilai uji $K = 3$, $K = 4$, $K = 5$, Evaluasi menggunakan dua pembagian data yaitu 70:30 dan 80:20, penggunaan dua pembagian data untuk mengetahui model rasio yang kinerja hasil klasifikasi yang paling baik [28] Pada penelitian ini perhitungan jarak menggunakan metode *Euclidean Distance*. metode ini mempunyai penerapan yang luas,

karena dapat digunakan dalam berbagai aplikasi juga *clustering*, klasifikasi, dan pengelompokan data. berikut rumus menghitung *Euclidean Distance*.

$$d_i = \sqrt{\sum_{i=1}^n (X_{2i} - X_{1i})^2} \quad (2)$$

X_{2i} = jarak data training ke-i

X_{1i} = jarak data test ke-i

n = banyak data training

i = baris ke-i

F. Evaluasi

Pada tahap evaluasi ini menggunakan *confusion matrix* dalam mengukur nilai seperti *accuracy*, *recall*, dan *precision*. *Accuracy* adalah nilai antara nilai prediksi yang benar dengan nilai aktual. *Precision* adalah nilai prediksi

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) Type I Error
	0 (Negative)	FN (False Negative) Type II Error	TN (True Negative)

Gambar. 2. Confussion Matrix

dan data yang diinginkan. *Recall* merupakan nilai prediksi yang benar dengan nilai benar yang aktual [29]. Setelah *confusion matrix* juga ada visualisasi yang berguna untuk melihat data yang sudah di olah dalam bentuk grafik [30]. Berikut pada gambar 2 contoh *confusion matrix*.

TP (*True Positive*) = Data positif yang diprediksi positif.

TN (*True Negatif*) = Data negatif yang diprediksi negatif.

FP (*False Positif*) = Data negatif yang diprediksi positif.

FN (*False Negatif*) = Data positif yang diprediksi negatif.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1\ score = 2 \frac{Recall * precision}{Recall + precision} = \frac{2TP}{2TP+FN+FP} \quad (6)$$

III. HASIL DAN PEMBAHASAN

A. Pembahasan

Pada penelitian ini menggunakan algoritma *K-Nearest Neighbor*, penelitian tentang opini masyarakat tentang *E-sport* mendapat hasil akurasi sebesar 84,66 % pada rsasio 80:20 di pengujian $K=5$. Berdasarkan pembahasan sebelumnya, ada beberapa penelitian yang sejenis tentang analisis sentimen. Kelebihan *K-Nearest Neighbor* adalah dapat mencapai tingkat akurasi yang tinggi [31]. Pada penelitian [13] yang menggunakan algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* menunjukkan hasil yang lebih baik dan akurat dibandingkan dengan *Naïve Bayes Classifier*. Pada penelitian lainnya [31], akurasi yang didapat dari klasifikasi algoritma *K-Nearest Neighbor* dengan menggunakan *RapidMiner* sebesar 87%. Berdasarkan pembahasan sebelumnya, algoritma *K-Nearest Neighbor* adalah algoritma yang dapat memberikan hasil yang optimal dalam klasifikasi.

B. Hasil

1) Pengumpulan Data

Data dikumpulkan dari media sosial X(Twitter) dengan mengambil *tweet* dari pengguna. Data diambil dengan cara *crawling data* menggunakan *python* dengan *module sncrape*, dengan *module* ini pengambilan data di X(Twitter) tidak perlu menggunakan API. Kata kunci yang digunakan dalam pengambilan data adalah “*e-sport mobile*”, dan “*e-sport Indonesia*”. X(Twitter) digunakan sebagai sumber pengambilan data karena proses pengambilan datanya yang mudah. Data *tweet* di ambil pada tahun 2022 dari bulan Januari sampai Desember dengan total jumlah data sebesar 1.868 data *tweet*. Jumlah data tersebut dipilih dikarenakan untuk menimalisir banyaknya pengurangan data pada proses *preprocessing*. Hasil *crawling data* dapat dilihat pada Tabel 1.

TABEL 1
HASIL CRAWLING DATA

Username	Text
multiakgae	b'Atlet e-sport Indonesia kalo mau terkenal di kancah e-sport yang lebih global, mesti coba fighting game, karena lemahnya olahraga di Indo in general aja tuh kerjasama team. Masalah akarnya selain mesti ngepush fighting game jadi populer ya mesti dapet suntikan bansos dari menpora'
LiZhiquan23	b'Anak muda Indonesia sangat menyukai e-sport. https://t.co/QwJicCfEQi '
Awanmalangit	b'Buat yang senang bermain game kini asah terus aja hobi kalian tersebut supaya bisa menjadi e-sport profesional. Namun seperti apa perkembangan e-sport di Indonesia? Baca pada artikel Kak @april_hamsa berikut ini. https://t.co/yjyUX5uRHH #bwbplus'

2) Preprocessing

Setelah proses *crawling data* dilakukan dan data diperoleh, selanjutnya dilakukan proses *preprocessing* untuk mengolah data mentah yang sebelumnya didapatkan melalui proses seperti *cleaning*, *case folding*, *tokenizing*, *stop-word*, dan *stemming*.

Cleaning dilakukan untuk mengurangi atau menghilangkan kata-kata yang tidak diperlukan dalam klasifikasi nanti seperti simbol, tanda baca, tautan terkait, ataupun emoticon pada teks *tweet*. Berikut pada Tabel 2 adalah contoh *cleaning*.

TABEL 2
CLEANING

Sebelum	Sesudah
b'Atlet e-sport Indonesia kalo mau terkenal di kancah e-sport yang lebih global, mesti coba fighting game, karena lemahnya olahraga di Indo in general aja tuh kerjasama team. Masalah akarnya selain mesti ngepush fighting game jadi populer ya mesti dapet suntikan bansos dari menpora'	Atlet sport Indonesia kalo mau terkenal kancah sport yang lebih global mesti coba fighting game karena lemahnya olahraga Indo general aja tuh kerjasama team Masalah akarnya selain mesti ngepush fighting game jadi populer mesti dapet suntikan bansos dari menpora

Setelah melakukan *cleaning*, proses selanjutnya adalah *case folding*. *Case folding* dilakukan untuk merubah semua teks menjadi *lowercase*, karena saat awal *tweet* diambil menggunakan *crawling* dalam *tweet* masih terdapat huruf kapital diantara *tweet* yang diambil. Berikut contoh pada Tabel 3 untuk hasil dari *case folding*.

TABEL 3
CASE FOLDING

Sebelum	Sesudah
Atlet sport Indonesia kalo mau terkenal kancah sport yang lebih global mesti coba fighting game karena lemahnya olahraga Indo general aja tuh kerjasama team Masalah akarnya selain mesti ngepush fighting game jadi populer mesti dapet suntikan bansos dari menpora	atlet sport indonesia kalo mau terkenal kancah sport yang lebih global mesti coba fighting game karena lemahnya olahraga indo general aja tuh kerjasama team masalah akarnya selain mesti ngepush fighting game jadi populer mesti dapet suntikan bansos dari menpora

Pada tahap *tokenizing* ini membagi atau memisahkan kalimat menjadi per kata. Dengan menggunakan *library nltk* Pada tahap *tokenizing* ini membagi atau memisahkan kalimat menjadi per kata. Dengan menggunakan *library nltk tokenize* dan mengimport *word tokenize* pada python. Contoh *tokenizing* dapat dilihat pada Tabel 4.

TABEL 4
 TOKENIZING

Sebelum	Sesudah
atlet sport indonesia kalo mau terkenal kancah sport yang lebih global mesti coba fighting game karena lemahnya olahraga indo general aja tuh kerjasama team masalah akarnya selain mesti ngepush fighting game jadi populer mesti dapet suntikan bansos dari men- pora	['atlet', 'sport', 'indonesia', 'kalo', 'mau', 'terkenal', 'kancah', 'sport', 'yang', 'lebih', 'global', 'mesti', 'coba', 'fighting', 'game', 'karena', 'lemahnya', 'olahraga', 'indo', 'general', 'aja', 'tuh', 'kerjasama', 'team', 'masalah', 'akarnya', 'selain', 'mesti', 'ngepush', 'fighting', 'game', 'jadi', 'populer', 'mesti', 'dapet', 'suntikan', 'bansos', 'dari', 'menpora']

Berikutnya adalah *stopword*, *Stopword* berfungsi untuk menghilangkan kata yang tidak sesuai dengan kamus indonesia ata menghilangkan kata yang tidak memiliki arti seperti ‘yang’, ‘krm’, ‘nya’, ‘sdh’, dan lain-lain. Kata-kata tersebut dihapus menggunakan *library nltk corpus* dengan mengimport *stopwords*. Berikut adalah contoh *stopword* yang ada pada Tabel 5.

TABEL 5
 STOPWORD

Sebelum	Sesudah
['atlet', 'sport', 'indonesia', 'kalo', 'mau', 'terkenal', 'kancah', 'sport', 'yang', 'lebih', 'global', 'mesti', 'coba', 'fighting', 'game', 'karena', 'lemahnya', 'olahraga', 'indo', 'general', 'aja', 'tuh', 'kerjasama', 'team', 'masalah', 'akarnya', 'selain', 'mesti', 'ngepush', 'fighting', 'game', 'jadi', 'populer', 'mesti', 'dapet', 'suntikan', 'bansos', 'dari', 'menpora']	['atlet', 'sport', 'indonesia', 'terkenal', 'kancah', 'sport', 'global', 'mesti', 'co- ba', 'fighting', 'game', 'lemahnya', 'olahraga', 'indo', 'general', 'kerjasa- ma', 'team', 'akarnya', 'mesti', 'ngepush', 'fighting', 'game', 'populer', 'mesti', 'dapet', 'suntikan', 'bansos', 'menpora']

Tahap selanjutnya setelah *stopword* adalah normalisasi, tahap ini berguna untuk menghilangkan kata alay atau salah eja dengan kata yang benar. Contoh seperti ‘indo’ menjadi ‘indonesia’ dengan menggunakan kamus *colloquial indonesian lexicon* pada *python* sebagai acuan dalam merubah kata. Berikut adalah contoh *tweet* yang di normalisasi pada Tabel 6.

TABEL 6
 NORMALISASI

Sebelum	Sesudah
['atlet', 'sport', 'indonesia', 'terkenal', 'kancah', 'sport', 'global', 'mesti', 'coba', 'fighting', 'game', 'lemahnya', 'olahraga', 'indo', 'general', 'kerjasama', 'team', 'akarnya', 'mesti', 'ngepush', 'fighting', 'game', 'populer', 'mesti', 'dapet', 'suntikan', 'bansos', 'menpora']	['atlet', 'sport', 'indonesia', 'terkenal', 'kancah', 'sport', 'global', 'mesti', 'co- ba', 'fighting', 'game', 'lemahnya', 'olahraga', 'indonesia', 'general', 'ker- jasama', 'team', 'akarnya', 'mesti', 'ngepush', 'fighting', 'game', 'populer', 'mesti', 'dapet', 'suntikan', 'bansos', 'menpora']

Terakhir ada tahap *stemming*, dimana tahap ini berfungsi untuk menghilangkan imbuhan pada kata dan merubahnya menjadi kata dasar seperti ‘per’, ‘ber’, ‘men’. Imbuhan dihilangkan dengan menggunakan *module* sastrawi. Kemudian setelah imbuhan dihilangkan proses selanjutnya adalah menghilangkan kurung siku. Pada Tabel 7 berikut adalah contoh dari hasil *stemming*.

TABEL 7
STEMMING

Sebelum	Sesudah
['atlet', 'sport', 'indonesia', 'terkenal', 'kancah', 'sport', 'global', 'mesti', 'coba', 'fighting', 'game', 'lemahnya', 'olahraga', 'indonesia', 'general', 'kerjasama', 'team', 'akarnya', 'mesti', 'ngepush', 'fighting', 'game', 'populer', 'mesti', 'dapat', 'suntikan', 'bansos', 'menpora']	atlet sport indonesia kenal kancah sport global mesti coba fighting game lemah olahraga indonesia general kerjasama team akar mesti ngepush fighting game populer mesti dapat sunti bansos menpora

3) Labelling Data

Setelah *preprocessing* selesai, tahap selanjutnya adalah *labelling* data. Pelabelan dilakukan secara auto dengan *module vader lexicon*, jadi data harus di *translate* terlebih dahulu ke bahasa inggris. Setelah data diubah kedalam bahasa inggris, akan dilakukan pemberian skor dan penjumlahan skor untuk melihat apakah data bernilai positif atau negatif. Setelah data selesai dinilai positif atau negatif, kolom *translate* bahasa inggris akan dihapus lagi. *Vader lexicon* dapat memberikan hasil penilaian yang cukup baik pada teks bahasa inggris dan juga pada teks sosial media yang singkat seperti pada X(Twitter). Berikut adalah contoh *labelling* pada Tabel 8.

TABEL 8
LABELLING DATA

Teks	Sentimen
atlet sport indonesia kenal kancah sport global mesti coba fighting game lemah olahraga indonesia general kerjasama team akar mesti ngepush fighting game populer mesti dapat sunti bansos menpora	Negatif
timnas sport indonesia tekad raih gelar juara sea games kamboja kirim atlet turun total nomor tanding timnas esport indonesia target medali emas kamboja	Positif
siapa bilang main game tuh cuma buat hiburan sekarang main game itu bisa jadi atlet sport lho gak cuma berprestasi tapi menghasilkan materi seperti apa perkembangan sport indonesia	Positif

4) Pembobotan TF-IDF

Pada tahap ini, pembobotan dilakukan dengan menggunakan TF-IDF. Perhitungan ini dilakukan dengan *module* dari *sklearn.tfidfvectorizer*. berikut pada Gambar 3 adalah contoh dari pembobotan TF-IDF.

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(use_idf=True, smooth_idf=False)
X_train_vect = vectorizer.fit_transform(X_train)
X_test_vect = vectorizer.transform(X_test)
dense_num = X_train_vect.todense()

print(X_train_vect)
```

```
(0, 3845) 0.0852440019323559
(0, 356) 0.6295435256189861
(0, 2183) 0.4923485971014468
(0, 4160) 0.5949800571599109
(1, 3625) 0.4426677627893245
(1, 3145) 0.3360379943166571
(1, 4353) 0.2641342988061658
(1, 230) 0.18923376918365947
(1, 676) 0.35447561678643985
(1, 1086) 0.27328833830726734
(1, 4420) 0.2907386032308104
(1, 1507) 0.3112706281968959
(1, 1709) 0.21392444269290964
(1, 3093) 0.3444849677331145
```

Gambar. 3. Pembobotan TF-IDF

Dalam TF-IDF ini bertujuan untuk mengekstrak kata ke dalam bentuk numerik agar dapat lanjut ketahap selanjutnya. Nilai dari term(kata) sering muncul dalam dokumen, semakin banyak kemunculannya, maka semakin berbobot nilai yang akan diperoleh [32]

Contoh perhitungan dari TD-IDF menggunakan persamaan 1, menghitung kata machine pada dokumen 1:

D1: "Machine learning is fascinating. Machine learning is a subfield of artificial intelligence."

D2: "Artificial intelligence is evolving rapidly."

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_{i+1}}\right) = \frac{2}{13} * \log\left(\frac{2}{2+1}\right) = \frac{2}{13} * 0 = 0$$

Hasil dari TF-IDF kata "machine" adalah 0

5) K-Nearest Neighbor

Selanjutnya ada tahap klasifikasi yang pada penelitian ini menggunakan *K-Nearest Neighbor*. Menggunakan *module sklearn.neighbors*, nilai tetangga atau neighbor yang di uji adalah K=3, 4, 5, dengan dua skenario pembagian data yaitu 70:30 dan 80:20. kemudian perhitungan jarak menggunakan *euclidean distance*. Pada penelitian [33] menggunakan metode *K-Nearest Neighbor* dengan nilai uji validasi yang digunakan yaitu K = 3 sampai 9, dapat memberi hasil *accuracy* yang memuaskan. Dengan K=3 sebesar 80,66% dan K=5 sebesar 75%. Pada penelitian [10], diketahui bahwa K=4 memberikan hasil yang cukup bagus dengan *accuracy* sebesar 82.00%. Pada Gambar 4 adalah contoh dari pengklasifikasian KNN.

```
# Pembangunan Model Klasifikasi Teks dengan KNN
from sklearn.neighbors import KNeighborsClassifier
n_neighbors=5
knn_tfidf = KNeighborsClassifier(n_neighbors, weights='distance', metric='euclidean')

knn_tfidf.fit(X_train_vect, y_train) #model
#Melakukan prediksi nilai y pada dataset testing
y_predict = knn_tfidf.predict(X_test_vect)

print(classification_report(y_test, y_predict))
```

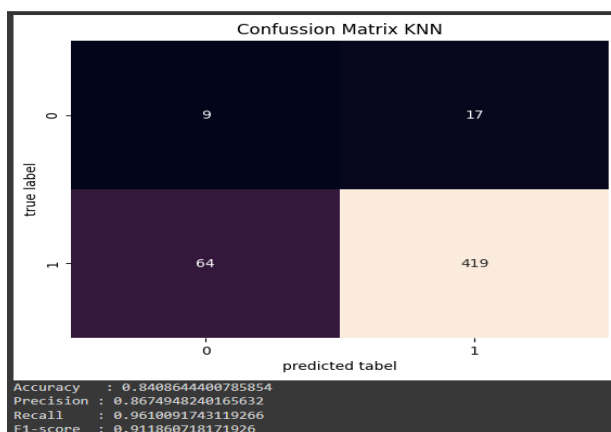
Gambar. 4. K-Nearest Neighbor

6) Evaluasi

Evaluasi menggunakan metode *confusion matrix* dengan menggunakan *library scikit-learn* dengan *module sklearn.metrics*. *confusion matrix* akan menghitung nilai dari *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Dengan skenario pembagian rasio pertama 70:30, dan nilai K yang digunakan untuk di uji adalah K= 3, K= 4, K= 5. Pada penelitian ini didapatkan hasil terbaik berada pada K= 5, dengan *Accuracy* sebesar 84,08%, *Precision* 86,74%, *Recall* 96,10%, dan *F1-score* 91,18%, dibanding dengan nilai K = 3, dan K = 4 yang mendapat hasil dibawah K=5. Untuk hasil klasifikasi dan *confusion matrix* dapat dilihat pada Gambar 5 dan Gambar 6.

	precision	recall	f1-score	support
-1	0.35	0.12	0.18	73
1	0.87	0.96	0.91	436
accuracy			0.84	509
macro avg	0.61	0.54	0.55	509
weighted avg	0.79	0.84	0.81	509

Gambar. 5. Hasil Klasifikasi rasio 70:30



Gambar. 6. Confussion Matrix rasio 70:30 81

Adapun perhitungan *accuracy* dengan persamaan 3, *precision* dengan persamaan 4, *recall* dengan persamaan 5, *F1-score* dengan persamaan 6.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{419+9}{419+17+64+9} = c = 0.840$$

$$Precision = \frac{TP}{TP+FP} = \frac{419}{419+64} = \frac{419}{483} = 0,867$$

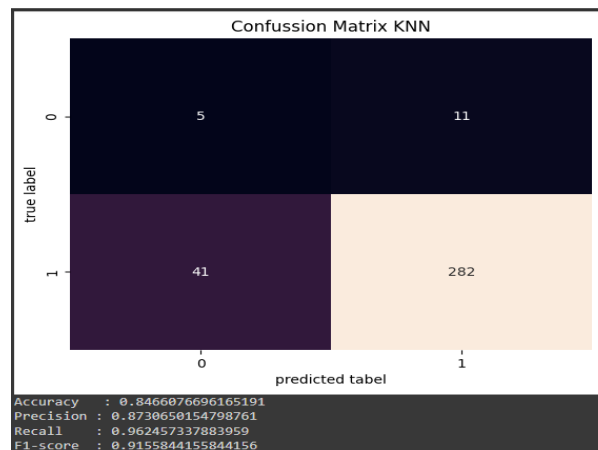
$$Recall = \frac{TP}{TP+FN} = \frac{419}{419+17} = \frac{419}{436} = 0.961$$

$$F1\ score = \frac{2TP}{2TP+FN+FP} = \frac{838}{919} = 0.911$$

Selanjutnya di skenario kedua, pembagian rasio 80:20 dengan nilai K yang akan diuji sama seperti sebelumnya yaitu K= 3, K= 4, K= 5. Didapatkan hasil yang terbaik pada nilai uji K= 5 dengan *Accuracy* sebesar 84,66%, *Precision* 87,30%, *Recall* 96,24%, dan *F1-score* 91,55%. Hasil klasifikasi dan *confussion matrix* dapat dilihat pada Gambar 7 dan Gambar 8.

	precision	recall	f1-score	support
-1	0.31	0.11	0.16	46
1	0.87	0.96	0.92	293
accuracy			0.85	339
macro avg	0.59	0.54	0.54	339
weighted avg	0.80	0.85	0.81	339

Gambar. 7. Hasil Klasifikasi rasio 80:20



Gambar. 8. Confussion Matrix rasio 80:20

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{282+5}{282+11+41+5} = \frac{287}{339} = 0.846$$

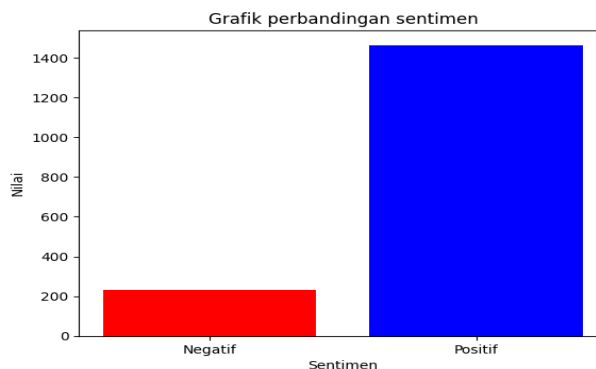
$$Precision = \frac{TP}{TP+FP} = \frac{282}{282+41} = \frac{282}{323} = 0,873$$

$$Recall = \frac{TP}{TP+FN} = \frac{282}{282+11} = \frac{282}{293} = 0.962$$

$$F1\ score = \frac{2TP}{2TP+FN+FP} = \frac{564}{616} = 0.915$$

Dari hasil kedua skenario, didapat bahwa skenario kedua yaitu pembagian rasio 80:20 lebih optimal dibanding 70:30 (skenario pertama). karena pada pembagian rasio 80:20, didapat *Accuracy*, *Precision*, *Recall*, dan *F1-score* yang terbaik dibanding rasio 70:30.

Kemudian selanjutnya dilakukan visualisasi untuk dapat menampilkan data yang sudah diolah dengan bentuk grafik ataupun bagan. Pada grafik ini menampilkan hasil labelling sentimen dari data yang didapat. Grafik perbandingan dapat dilihat pada Gambar 9.



Gambar. 9. Grafik Perbandingan Sentimen

Kemudian pada Gambar 10, ditunjukkan *wordcloud* yang menampilkan kata-kata apa yang sering muncul pada opini masyarakat mengenai *E-sport mobile*. Frekuensi jumlah kata yang diambil dalam *wordcloud* ini sebesar 1000 kata, dengan kata “*sport*” yang menempati urutan pertama dengan jumlah kemunculan sebesar 1.892 kali dan kata “*indonesia*” sebanyak 927 kali.



Gambar. 10. Wordcloud

Berdasarkan hasil yang sudah didapat, diketahui bahwa sentimen masyarakat mengenai perkembangan *E-sport mobile* di Indonesia mengarah kepada sentimen positif. Untuk dua skenario yang dilakukan diketahui bahwa rasio 80:20 lebih optimal dibanding dengan 70:30. Dikarenakan ruang lingkup yang hanya terbatas pada *mobile* dan Indonesia saja, diharapkan peneliti selanjutnya agar bisa menambah ruang lingkup tidak hanya terikat di Indonesia saja dan juga bisa menambah ruang lingkup tidak hanya pada *E-sport mobile*.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan, data yang terkumpul dari hasil *crawling* sebesar 1.868 data, setelah dilakukan proses *preprocessing* data menjadi 1.696 data. Setelah melakukan proses *labelling*, berikutnya adalah tahap klasifikasi dengan menggunakan dua skenario split data yaitu 70:30 dan 80:20. Diketahui bahwa hasil dari banyaknya sentimen positif berjumlah 1464 dan negatif berjumlah 231. Dalam penelitian ini metode yang digunakan adalah *K-Nearest Neighbor* dan perhitungan jarak menggunakan *euclidean distance* dengan nilai uji tetangga sebesar $K=3$, $K=4$, $K=5$. Dari ketiga nilai K yang diuji dengan ke dua skenario, didapat hasil yang optimal pada skenario kedua yaitu 80:20 dengan nilai uji $K=5$ hasil performa yang didapat *Accuracy* sebesar 84,66%, *Precision* 87,30%, *Recall* 96,24%, dan *F1-score* 91,55%. Jika dilihat dari hasil sentimen yang sudah didapat, diketahui bahwa perkembangan *E-sport mobile* di Indonesia dominan kepada sentimen positif. Kekurangan dalam penelitian ini adalah lingkup yang diambil dalam pembahasan hanya terbatas pada *E-sport mobile*, jadi harap pada peneliti yang akan melanjutkan penelitian ini bisa mengambil lingkup yang lebih luas lagi, dan bisa menambah data antara positif dan negatif yang lebih seimbang agar proses klasifikasi menjadi lebih optimal.

DAFTAR PUSTAKA

- [1] F. Kurniawan, "E-Sport dalam Fenomena Olahraga Kekinian," *Jorpres (Jurnal Olahraga Prestasi)*, vol. 15, no. 2, pp. 61–66, Jan. 2020, doi: 10.21831/jorpres.v15i2.29509.
- [2] R. Qomarrullah, S. A. Kesumawati, S. Suratni, U. Pakasi, T. S. Guntoro, and L. W. S., "Sosialisasi tentang E-sport pada Mahasiswa S1 Penjaskresrek Universitas Binadarma Palembang sebagai Peluang Bisnis Digital Masa Depan," *Jurnal Abdi Masyarakat Indonesia*, vol. 2, no. 4, pp. 1319–1324, Jul. 2022, doi: 10.54082/jamsi.428.
- [3] S. Nugroho, S. Sumarjo, A. Nasrulloh, and K. W. Pratama, "Impact of e-sport games on the character building and sports culture," *Jurnal Keolahragaan*, vol. 10, no. 1, pp. 91–100, Apr. 2022, doi: 10.21831/jk.v10i1.48310.
- [4] Indri, "Perkembangan Esports di Indonesia Sampai Tahun 2020," INDOESPORTS. Accessed: Oct. 08, 2023. [Online]. Available: <https://www.indoesports.com/news/esports-bisnis/perkembangan-esports-indonesia-sampai-tahun-2020>
- [5] K. Prawajasena, "Elektronik Sports Jurnalistik dalam Acara" eSports Star Indonesia", *Prosiding Konferensi Nasional Komunikasi*, vol. 3, no. 1, pp. 29–39, 2020.
- [6] K. Azwar and M. Mailindawati, "DAMPAK ESPORT GAMES TERHADAP TINGKAT EMOSIONAL DAN PRESTASI BELAJAR REMAJA DI KOTA LHOKEUMAWE PROVINSI ACEH TAHUN 2020," *Visipena*, vol. 11, no. 2, pp. 255–265, Dec. 2020, doi: 10.46244/vi-sipena.v11i2.1208.
- [7] Vika Azkiya Dhini, "Jumlah Gamers Indonesia Terbanyak Ketiga di Dunia." Accessed: Oct. 08, 2023. [Online]. Available: <https://data-boks.katadata.co.id/datapublish/2022/02/16/jumlah-gamers-indonesia-terbanyak-ketiga-di-dunia>
- [8] "Indonesia." Esport Earnings. Accessed: Nov. 15, 2023. [Online]. Available: <https://www.esportearnings.com/countries/id>
- [9] E. F. Mumtaz, S. K. Ragamustari, and F. B. Hirawan, "The Impact of the E-Sport Curriculum Toward Online Game Addiction," *TAZKIYA: Journal of Psychology*, vol. 9, no. 1, pp. 29–39, May 2021, doi: 10.15408/tazkiya.v9i1.19986.
- [10] A. Tunggu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - Journal of Informatics Engineering*, vol. 2, no. 01, pp. 24–31, Jun. 2021, doi: 10.53682/jointer.v2i01.30.
- [11] F. Rizqi Irawan, "ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEARSET NEIGHBORS," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 1, pp. 62–68, Apr. 2022, doi: 10.33387/jiko.v5i1.4267.
- [12] Yunitasari, H. S. Hopipah, and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritma k-nearest neighbor (k-NN) dan Naive Bayes," *Technomedia Journal*, vol. 6, no. 1, pp. 99–110, Jul. 2021, doi: 10.33050/tmj.v6i1.1531.
- [13] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, pp. 1329–1336, 2022.
- [14] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naive Bayes dan K-Nearest Neighbor," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 3, no. 4, pp. 203–208, Jan. 2022, doi: 10.35746/jtim.v3i4.178.
- [15] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE," *Jurnal Ilmu Komputer dan Informasi*, vol. 13, no. 2, pp. 109–122, Jul. 2020, doi: 10.21609/jiki.v13i2.885.
- [16] M. Y. Febrianta, S. Widyanesti, and S. R. Ramadhan, "Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *Journal of Animation and Games Studies*, vol. 7, no. 2, pp. 117–144, Oct. 2021, doi: 10.24821/jags.v7i2.5162.
- [17] D. Apriliani, A. Susanto, M. F. Hidayattullah, and G. W. Sasmito, "Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 1, pp. 34–37, Jan. 2023, doi: 10.30591/jpit.v8i1.4759.
- [18] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, vol. 5, no. 2, pp. 697–711, 2021.
- [19] F. F. Mailoa, "Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia," *Journal of Information Systems for Public Health*, vol. 6, no. 1, p. 44, Oct. 2021, doi: 10.22146/jisph.44455.
- [20] D. Alita, Y. Fernando, and H. Sulistiani, "IMPLEMENTASI ALGORITMA MULTICLASS SVM PADA OPINI PUBLIK BERBAHASA INDONESIA DI TWITTER," *Jurnal Tekno Kompak*, vol. 14, no. 2, p. 86, Aug. 2020, doi: 10.33365/jtk.v14i2.792.
- [21] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *Jurnal Linguistik Komputasional (JLK)*, vol. 5, no. 1, p. 34, Apr. 2022, doi: 10.26418/jlk.v5i1.79.
- [22] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 2, p. 1, Aug. 2021, doi: 10.33365/jdmsi.v2i2.1344.
- [23] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 1, pp. 1–10, Apr. 2020, doi: 10.35957/jatisi.v7i1.196.
- [24] A. Santosa, I. Purnamasari, and R. Mayasari, "Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 6, no. 1, pp. 81–93, 2022.
- [25] E. A. Lisangan, A. Gormantara, and R. Y. Carolus, "Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, Apr. 2022, doi: 10.24002/konstelasi.v2i1.5609.
- [26] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311.
- [27] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, Aug. 2020, doi: 10.28932/jutisi.v6i2.2658.
- [28] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *Jurnal Sains dan Informatika*, vol. 8, no. 1, pp. 88–96, Jul. 2022, doi: 10.34128/jsi.v8i1.419.
- [29] S. D. Fritama, Y. R. Ramadhan, and M. A. Komara, "Analisis Sentimen Review Produk Acne Spot Treatment di Female Daily Menggunakan Algoritma K-Nearest Neighbor," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 134–143, 2023.
- [30] J. F. Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 360–367, 2023.
- [31] A. Tunggu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - Journal of Informatics Engineering*, vol. 2, no. 01, pp. 24–31, Jun. 2021, doi: 10.53682/jointer.v2i01.30.
- [32] F. Novianti and K. R. N. Wardani, "ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET TRAVELOKA SELAMA RAPID TEST ANTIGEN MENGGUNAKAN ALGORITMA NAIVE BAYES," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 922–933, Aug. 2023, doi: 10.29100/jipi.v8i3.3973.
- [33] A. Halimi, K. Kusri, and M. R. Arief, "ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP PEMBELAJARAN ONLINE DARI DI MEDIA SOSIAL TWITTER MENGGUNAKAN LEXICON DAN K-NEAREST NEIGHBOR," *COREAI: Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, vol. 2, no. 1, pp. 18–28, Aug. 2021, doi: 10.33650/coreai.v2i1.2283.