

Analisis Sentimen Performansi Operator Telekomunikasi di Indonesia Menggunakan Metode *Text Mining*

Ersha Aisyah Elfaiz¹, Riza Akhsani Setyo Prayoga², Monica Cinthya³, Muhammad Sonhaji Akbar⁴, Rizky Basatha⁵

^{1,2,4,5}Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia

³Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia

Email: ¹ershaelfaiz@unesa.ac.id, ²rizaprayoga@unesa.ac.id, ³monicacinthya@unesa.ac.id,

⁴muhammadakbar@unesa.ac.id, ⁵rizkybasatha@unesa.ac.id

ABSTRACT

The telecommunications sector in Indonesia has experienced rapid development in recent years, characterized by the increasing number of telecommunications operators offering various services and products. Therefore, there is competitive rivalry among operators. The right strategy is needed to survive and compete effectively. One of the efforts that can be made by telecommunication companies is to evaluate operational performance. This research aims to analyze the sentiment of X users towards telecommunication operational performance, at Telkomsel and Tri operators using text mining methods, namely Naïve Bayes, Support Vector Machine (SVM) and Decision Tree Learning (DTL). The research data is obtained by crawling from the X application, then the data is processed to remove unnecessary words or affixes. Then data modeling and validation is carried out using split validation and cross validation techniques. In the split validation technique, the data is divided into 70% training data and 30% testing data, while in the cross-validation technique the fold parameter is set to determine which fold has the highest accuracy. The results of the study show that the SVM method has the highest accuracy, where in split validation the accuracy is 84.29% for Telkomsel data and 75.70% for Tri data. Similarly, in cross validation, the accuracy is 82.15% on fold 4, 7 for Telkomsel data and 61.41% on fold 9 for Tri data. In addition, it is known that Telkomsel data has 18.64% positive sentiment and 81.36% negative sentiment. While Tri data has 61.11% positive sentiment and 38.89% negative sentiment.

Keywords: Sentiment Analysis, Telecommunications, Text Mining, VADER.

ABSTRAK

Sektor telekomunikasi di Indonesia telah mengalami perkembangan yang pesat dalam beberapa tahun terakhir, ditandai dengan semakin banyaknya operator telekomunikasi yang menawarkan berbagai layanan dan produk. Oleh karena itu, terjadi persaingan yang kompetitif diantara operator. Diperlukan strategi yang tepat untuk dapat bertahan dan bersaing secara efektif. Salah satu upaya yang dapat dilakukan oleh perusahaan telekomunikasi adalah dengan melakukan evaluasi terhadap performansi operasional. Penelitian ini bertujuan untuk menganalisis sentimen pengguna X terhadap performansi operasional telekomunikasi, pada operator Telkomsel dan Tri dengan menggunakan metode text mining, yaitu Naïve Bayes, Support Vector Machine (SVM) dan Decision Tree Learning (DTL). Data penelitian didapatkan dengan cara *crawling* dari aplikasi X, selanjutnya data diolah untuk menghilangkan kata atau imbuhan yang tidak perlu. Kemudian dilakukan *data modelling* dan validasi dengan menggunakan teknik *split validation* dan *cross validation*. Pada teknik *split validation*, data dibagi menjadi 70% *data training* dan 30% *data testing*, sedangkan pada teknik *cross validation* diatur parameter *fold* untuk menentukan *fold* berapa yang paling tinggi akurasi. Hasil dari penelitian menunjukkan metode SVM memiliki akurasi paling tinggi, dimana pada *split validation* akurasi 84.29% untuk data Telkomsel dan 75.70% untuk data Tri. Sama halnya pada *cross validation* akurasi 82.15% pada *fold* 4, 7 untuk data Telkomsel dan 61.41% pada *fold* 9 untuk data Tri. Selain itu, diketahui data Telkomsel memiliki sentimen positif 18.64% dan 81.36% sentimen negatif. Sedangkan data Tri memiliki 61.11% sentimen positif dan 38.89% sentimen negatif.

Kata Kunci: Analisis Sentimen, Telekomunikasi, Text Mining, VADER.

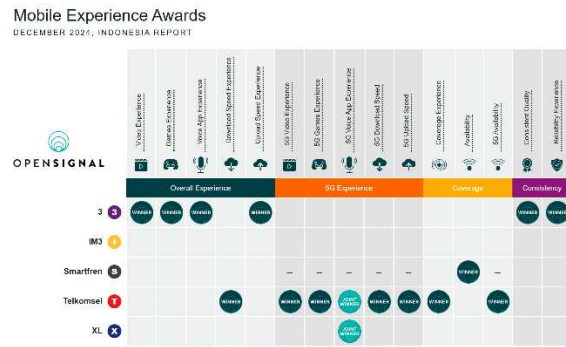
1. Pendahuluan

Berdasarkan Global Digital Report 2024 tercatat sebanyak 66.5% dari total populasi di Indonesia atau sekitar 185.3 juta pengguna internet di Indonesia. Selain itu, 39 juta pengguna sosial media sudah aktif menggunakan sosial media, baik facebook, X, Instagram, youtube, tiktok, dan lain sebagainya [1]. Pada Januari 2024, tercatat bahwa X memiliki 24.69 juta pengguna dari Indonesia atau setara dengan 8.9% dari total populasi [1]. Pengguna melakukan akses ke aplikasi media sosial dengan menggunakan jaringan internet [2], [3]. Jaringan internet merupakan salah satu layanan dari operator telekomunikasi. Hal itu berarti kualitas layanan operator telekomunikasi penting untuk memenuhi kebutuhan pengguna dalam mengakses media sosial [2].

Sektor telekomunikasi di Indonesia sendiri telah mengalami perkembangan yang pesat dalam beberapa tahun terakhir, ditandai dengan semakin banyaknya operator telekomunikasi yang menawarkan berbagai layanan dan produk. Pada era telekomunikasi di Indonesia saat ini, terdapat indikasi degradasi pada kreativitas dan inovasi operator telekomunikasi yang tertinggal dari penawaran di luar negeri [4]. Akibatnya, beberapa perusahaan telah mulai menghadapi kondisi yang semakin kompetitif dan menantang, sehingga diperlukan strategi yang tepat untuk dapat bertahan dan bersaing secara efektif. Salah satu upaya yang dapat dilakukan oleh perusahaan telekomunikasi adalah dengan melakukan evaluasi terhadap performansi operasional.

Terdapat lima operator telekomunikasi besar yang menyediakan layanan telekomunikasi dan internet di Indonesia, yaitu Telkomsel, Tri, Smartfren, XL dan Indosat [5]. Pada Gambar 1 dijelaskan bahwa terdapat empat operator yang memberikan pengalaman baik kepada pelanggan selama 2024. Menurut laporan tersebut, terdapat empat operator yang memiliki performa baik selama tahun 2024, yaitu Telkomsel, Tri XL, dan Smartfren. Laporan tersebut disusun berdasarkan pengumpulan data dari 1 Agustus sampai 29 Oktober 2024 dengan mengumpulkan miliaran data setiap hari dari berbagai demografi, perangkat, dan pengguna [6].

Masing-masing operator telekomunikasi memiliki akun resmi di media sosial, khususnya X. Hal ini sekaligus memperlihatkan kemungkinan pengguna untuk memberikan komentar positif maupun negatif terkait layanan operator melalui akun resmi tersebut. Beberapa penelitian telah melakukan analisis sentiment performansi suatu aplikasi atau layanan berdasarkan komentar yang diberikan pelanggan melalui X [7], [8]. Berdasarkan beberapa penelitian tersebut dapat diketahui bahwa performansi suatu layanan dapat dilihat dari komentar yang diunggah oleh pengguna di X.



Gambar 1. Laporan pengalaman pengguna pada operator telekomunikasi Indonesia [5]

Text mining adalah salah satu pendekatan yang dapat digunakan untuk menganalisis sentimen dari teks yang tersedia, seperti pada media sosial [9]. Analisis sentimen atau yang dapat disebut dengan *opinion mining* adalah salah satu bagian dari *text mining*. Analisis sentimen dapat dimanfaatkan untuk mengevaluasi persepsi dan opini pelanggan terhadap kebijakan, layanan atau produk tertentu [10], [11], seperti halnya layanan operator telekomunikasi. Analisis sentimen dilakukan untuk melihat kecenderungan opini yang diberikan oleh seseorang terhadap sebuah permasalahan atau objek sehingga dapat diketahui apakah kecenderungan orang tersebut berpandangan positif atau negatif [12], [13], [14]. Analisis sentimen seringkali digunakan untuk menganalisis opini pelanggan terhadap produk dan layanan suatu bisnis [12].

Terdapat beberapa metode yang biasanya digunakan untuk analisis sentimen. Salah satu metode yang dapat digunakan adalah metode Naive Bayes. metode naive bayes memiliki keunggulan dalam hal kemudahan implementasi, kecepatan komputasi, dan kemampuan untuk beradaptasi dengan data yang tidak seimbang [13]. Metode Naive Bayes Classifier merupakan salah satu metode klasifikasi yang dikemukakan oleh Thomas Bayes. Berdasarkan penelitian sebelumnya, penelitian ini mencapai akurasi 94% [15]. Penelitian lain yang menggunakan metode ini untuk mengetahui sentiment opini publik [10].

Metode ini menggunakan konsep dasar Teorema Bayes. Teorema Bayes diekspresikan dengan Persamaan 1 [16]:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Dimana :

A = sampel data yang label kelasnya tidak diketahui

B = kelas-kelas hasil klasifikasi

P(A|B) = probabilitas terjadinya A jika B diketahui

P(B|A) = probabilitas terjadinya B jika A diketahui

P(A) = probabilitas prior A yang mendahului terjadinya B

P(B) = probabilitas prior B

Metode Decision Tree Learning (DTL) merupakan metode klasifikasi yang mudah untuk diinterpretasi oleh manusia. Metode ini mengadopsi struktur pohon yang memiliki hirarki untuk memprediksi suatu data. Oleh karena itu, manfaat dari penggunaan metode ini adalah dapat menurunkan proses pengambilan keputusan yang kompleks menjadi lebih simpel. Selain itu menurut penelitian yang dilakukan oleh [16] metode ini memiliki akurasi dan presisi yang cukup baik yaitu 91.94% dan 97.34%.

Metode Support Vector Machine (SVM) diperkenalkan oleh Vladimir Vapnik dan turunan standar saat ini diusulkan oleh Clorinna Cortes [17]. SVM adalah sistem pembelajaran mesin yang bekerja dengan prinsip Structural Risk Minimization. Prinsip tersebut bertujuan untuk menemukan hyperplane terbaik yang memisahkan dua buah kelas. Hyperplane terbaik antara dua kelas dapat didapatkan dengan mengukur margin hyperplane dan mencari titik maksimalnya. Margin merupakan jarak antara hyperplane dengan pattern terdekat dari masing-masing kelas. Pattern yang paling dekat disebut dengan support vector. Beberapa penelitian juga menganalisis sentimen dengan menggunakan metode SVM dan *lexicon based* [18], [19]. Menurut penelitian terdahulu, Support Vector Machine terbukti mampu menghasilkan kinerja yang cukup baik dengan hasil akurasi sebesar 98% dan f1-score sebesar 0.98 atau sebesar 98% [20].

Penelitian ini bertujuan untuk mengetahui sentimen mengenai performansi operator telekomunikasi di X berdasarkan tweet yang diunggah dengan menggunakan tiga metode, yaitu Naïve Bayes, Decision Tree dan SVM. Operator telekomunikasi yang diteliti adalah Tri, Smartfren dan Telkomsel. Kemudian, dirumuskan pertanyaan penelitian yang dijawab pada penelitian ini. Pertanyaan penelitian ini yaitu: Bagaimana sentimen masyarakat Indonesia mengenai performansi operator telekomunikasi berdasarkan tweet yang diunggah? Metode apa yang memiliki akurasi paling tinggi?

2. Metode Penelitian

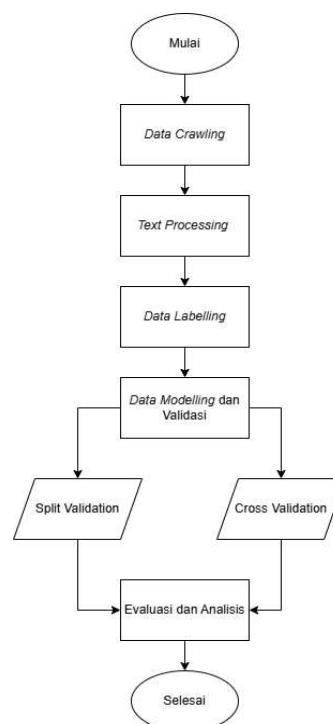
Pada bagian ini dijelaskan metodologi penelitian. Penelitian menggunakan data tweet dari X. Terdapat tiga operator telekomunikasi yang diteliti pada penelitian ini, yaitu Telkomsel, Tri, dan Smartfren. Kemudian dilakukan perhitungan dengan metode Naive Bayes Classifier, Decision Tree Learning dan Support Vector Machine untuk analisis sentimen. Pada Gambar 2 dijelaskan alur tahapan penelitian mulai awal sampai akhir.

2.1. Data Crawling

Data yang digunakan pada penelitian ini adalah data tweet yang didapatkan dengan menggunakan X API. Kode ditulis dengan bahasa Python dengan node.js. Kata kunci pencarian yang digunakan dalam pencarian data tweet dijelaskan pada Tabel 1.

Tabel 1. Kata kunci pencarian data

Operator Telekomunikasi	Kata kunci pencarian
Telkomsel	telkomsel AND mytelkomsel AND simpati
Tri	tri AND triindonesia
Smartfren	smartfren AND smartfrenworld



Gambar 2. Tahapan penelitian

2.2. Text Preprocessing

Sebelum melakukan pengolahan data teks, data tweet yang telah dilabeli kemudian diproses terlebih dahulu. Proses tersebut bertujuan untuk mengubah data dari teks yang tidak terstruktur menjadi data terstruktur. Tahapan yang digunakan dalam text preprocessing pada penelitian ini adalah:

- 1) *Tokenizing*: Tahap ini bertujuan untuk mengubah teks yang panjang menjadi beberapa bagian kecil atau token. Kalimat yang ada pada data tweet dipisahkan berdasarkan spasi.
- 2) *Cleansing*: Tahap ini bertujuan untuk menghilangkan karakter yang tidak diutuhkan untuk proses selanjutnya. Pada tahap ini akan dihilangkan angka, tanda baca, dan emoji.
- 3) *Stopword Removal*: Tahap ini bertujuan untuk menghapus kata yang dianggap tidak memiliki makna penting pada data tweet. Contoh kata yang dihapus adalah kata hubung atau kata depan seperti “di”, “ke”, “dan”, “dengan”, atau “ini”.
- 4) *Stemming*: Tahap ini dilakukan untuk menghapus imbuhan pada token sehingga menjadi kata dasar, misalkan kata “pembuatan” menjadi “buat”. Proses ini dilakukan dengan pemrograman python dan library Sastrawi untuk Bahasa Indonesia

2.3. Data Labelling

Data yang terkumpul dari proses sebelumnya dilabeli dengan label “Positif” atau “Negatif” secara otomatis dengan menggunakan bahasa pemrograman Python dan menggunakan VADER (Valence Aware Dictionary and sEntiment Reasoner) [21]. Lexicon VADER adalah sebuah kamus kata yang diciptakan oleh CJ Hutto dan Eric Gilbert dari Institute Technology Georgia. Lexicon VADER adalah satu alat analisis sentimen berbasis leksikon dan aturan yang dirancang khusus untuk teks yang mengandung bahasa informal seperti posting dan ulasan media sosial dengan menggunakan bahasa Inggris. Tahapan pelabelan menggunakan VADER adalah sebagai berikut:

- 1) Melakukan perubahan bahasa dari bahasa Indonesia ke bahasa Inggris. Proses ini diperlukan karena VADER leksikon dirancang dengan menggunakan bahasa Inggris.
- 2) Melakukan perhitungan bobot
Terdapat pembobotan pada mekanisme VADER. Jika bobot lebih dari 0.05 akan dikelompokkan ke sentimen positif, jika kurang dari -0.05 masuk dikelompokkan ke sentimen negatif, dan apabila bobot diantara -0.05 sampai 0.05 akan dikelompokkan ke sentimen netral [21].
- 3) Data dan hasil sentimen ditampilkan dalam bahasa Indonesia

2.4. Data Modelling dan Validasi

Setelah data tweet yang berupa token telah menjadi kata yang terstruktur, dilakukan validasi dengan menggunakan metode *split validation* dan *cross validation*.

1) Split Validation

Penelitian ini menggunakan pemisahan data menjadi dua bagian, yaitu *data training* atau data latih dan *data testing* atau data uji. Pemisahan data dilakukan dengan membagi 70% data latih dan 30% secara acak sebagai data uji [22]. Pada data Telkomsel, 165 data digunakan untuk data latih dan 71 data digunakan untuk data uji. Sedangkan untuk data Tri, 176 data digunakan untuk data latih dan 76 data digunakan untuk data uji.

Data latih akan dijalankan terlebih dahulu dengan ketiga algoritma. Selanjutnya, data uji dijalankan dengan cara yang sama. Kemudian hasil dari uji algoritma dibandingkan dengan hasil pelabelan, sehingga didapatkan persentase akurasi.

2) Cross Validation

Cross validation adalah metode yang digunakan untuk evaluasi akurasi suatu algoritma dengan memisahkan dataset ke dalam sejumlah kelompok k yang sama besar. Selanjutnya, model dilatih pada $k(1)$ dan data lainnya digunakan sebagai data uji. Proses tersebut diulang k kali, sehingga seluruh data digunakan sekali sebagai data uji [23]

Selain pada penelitian ini juga dilakukan *split validation*, untuk membandingkan akurasi dari ketiga algoritma, dilakukan fungsi cross validation yang bertujuan untuk meningkatkan akurasi dengan berbagai variabel k -fold yang bisa dipilih. K -fold adalah membagi data menjadi sejumlah kelompok, dimana jika k -fold sama dengan 10 maka data dibagi menjadi 1- bagian secara acak. Kemudian, 90% digunakan untuk data latih dan 10% digunakan untuk data uji. Proses ini diulang sebanyak jumlah k -fold sehingga setiap data pernah menjadi data uji [24]. Pada penelitian ini akan diuji nilai k -fold dari 2 sampai dengan 10, sehingga didapatkan variabel k -fold yang menghasilkan akurasi paling tinggi.

Pada tahap ini akan dibandingkan hasil dari kedua teknik validasi tersebut, sehingga akan didapatkan akurasi dari masing-masing metode yang digunakan. Proses menghitung akurasi dilakukan dengan formula (2) [24]

$$\text{akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Dimana

TP = True Positive
FP = False Positive
TN = True Negative
FN = False Negative

2.5. Evaluasi dan Analisis

Tahapan selanjutnya adalah melakukan evaluasi dari hasil data yang didapatkan pada tahapan sebelumnya. Pada tahap ini akan diketahui akurasi masing-masing metode. Selain itu, juga didapatkan karakteristik opini, baik sentimen negatif atau positif, dari masing-masing operator telekomunikasi yang diteliti berdasarkan pembacaan data akhir.

3. Hasil dan Pembahasan

Pada bagian ini dijelaskan hasil dari tahapan penelitian yang telah dilakukan.

3.1. Data Crawling dan Data Preprocessing

Proses pengumpulan data menghasilkan 1530 tweet. Selanjutnya, dilakukan penghapusan data yang duplikat, yang merupakan balasan dari admin media sosial operator dan data yang tidak sesuai dengan konteks performansi, seperti promosi. Berdasarkan proses tersebut, didapatkan 490 data dimana persebaran data dijelaskan pada Tabel 2.

Tabel 2. Hasil *data crawling* dan *data preprocessing*

Operator Telekomunikasi	Hasil Data Crawling	Hasil Data Preprocessing
Telkomsel	702	236
Tri	425	252
Smartfren	403	2

Tweet
@Telkomsetl Gatau kenapa semua pr
@anticurur Makasih udah mau nung
@Telkomsetl Oke baik
@Telkomsetl boleh minta tolong cek r
RT @10gku_ : Telkomsetl skrg udh miri
@daniel_randa Pagi, Kak Danil. Maaf
@Telkomsetl min ini gimana nih.. beli
@Telkomsetl #TelkomsetlNoGame pr
@AbuAlRayyan1 @Telkomsetl Buat op
@MazByan @Telkomsetl Parah ya ,,,,
@westenth sebelum provider telkor
@Pratama_Gaa Hai, Kak Pratama. M
@ismiafryanti @Telkomsetl Iya sist k
@Telkomsetl @sibegenggg @CNNInd
@Telkomsetl Saya udah melakukan b
@fuadoshi @Telkomsetl @TelkomCar
Galau banget mau ganti nomer ðŸ˜ˆ~”
@xyzzyy_Hai, Kak. Maaf ya jadi ga n
Min @Telkomsetl aku kan bbrp kali co
@rizkidwika Senasib kalo gitu. Di sini
@avocuddle_lrz Pokonya sebel kali is

Gambar 3. Hasil *data crawling* operator Telkomsetl

Pada Gambar 3 dan Gambar 4 dijelaskan contoh hasil *data crawling* dari data operator Telkomsetl dan Trii. Setelah dilakukan pengumpulan data, selanjutnya dilakukan tahap *tokenizing* untuk mengubah teks yang panjang menjadi beberapa bagian kecil atau token.

Tweet
@triindonesia Signyal Tri jelek. Gak kaya
@Andika15_ @triindonesia Bukan nya da
@triindonesia Kak, jika kartu tri hilang ak
@triindonesia Pgn tanya Saya pke tri dari
@fadlialiff @triindonesia @IndiHome wa
@fadlialiff @triindonesia @IndiHome tk k
@triindonesia Ini udh di undi blm ya?:(c
@triindonesia min ini tri lagi gangguan ya
@RockOn89306536 Maaf ya kak, untuk k
@triindonesia kartu tri saya tidak bisa m
@triindonesia Min.. saya pelanggan setia
Tri baik bgt deh, makin luv @triindonesia
@chaerdinal Hai Kak, maaf ya buat ngga
@triindonesia hallo min ini jaringan tri m
@triindonesia Masih sama aja i min. Apa
@triindonesia Sudah 8 tahun pakai tri #u
@triindonesia dari SMP sampai sudah ke
@triindonesia Min kenapa pagi ini jaringa
@triindonesia min, signalmu 4G di kota B
@triindonesia Aku udah lama pake tri, su
@triindonesia PakeTRI isi paket data mu

Gambar 4. Hasil *data crawling* operator Trii

```
# tokenize
def tokenize(text):
    text = word_tokenize(text)
    return text
```

Gambar 5. Proses tokenisasi

Pada Gambar 5 dijelaskan koding menggunakan bahasa python untuk proses tokenisasi. Pada Tabel 3 dijelaskan contoh data hasil tokenisasi dari sebuah kalimat menjadi kata.

Tabel 3. Hasil tokenisasi

Hasil data crawling	Hasil tokenisasi
RT @ariagaana: telkomsetl + indihome suatu kombinasi yg mantabðŸ˜ˆ• ðŸ˜ˆ• »	RT, @ariagaana:, telkomsetl, + indihome, suatu, kombinasi, yg mantab, ðŸ˜ˆ• ðŸ˜ˆ• »
hai @Telkomsetl lagi error ya? ngisi pulsa dari tadi pagi belum masuk	Hai, @Telkomsetl, lagi, error, ya?, ngisi, pulsa, dari, tadi, pagi, belum, masuk
@triindonesia Kuota bejibun harga merakyat Ga nyesel pake tri #UntungPakaiTri	@triindonesia, Kuota, bejibun, harga, merakyat, Ga, nyesel, pake, tri, #UntungPakaiTri
@triindonesia Hallo tri, ini kenapa jaringan tri ku, tak bisa ya pakai mobile data ? Cek pulsa juga gal bisa :(	@triindonesia, Hallo, tri, ini, kenapa, jaringan, tri, ku, tak, bisa, ya, pakai, mobile, data, ?, Cek, pulsa, juga, gal, bisa : (

Selanjutnya, dilakukan penghilangan karakter yang tidak dibutuhkan untuk proses selanjutnya. Gambar 6 menjelaskan koding proses pembersihan data. Pada tahap ini akan dihilangkan angka, tanda baca, dan emoji. Hasil pembersihan data dijelaskan pada Tabel 4.

```
# remove urls, handles, and the hashtag from hashtags
def remove_urls(text):
    new_text = ' '.join(re.sub("[A-Za-z0-9+]|(^0-9A-Za-z \t)]|(\w+:\/\/\S+)", "", text).split())
    return new_text
# make all text lowercase
def text_lowercase(text):
    return text.lower()
# remove numbers
def remove_numbers(text):
    result = re.sub(r'\d+', '', text)
    return result
# remove punctuation
def remove_punctuation(text):
    translator = str.maketrans('', '', string.punctuation)
    return text.translate(translator)
```

Gambar 6. Proses penghilangan pembersihan data

Tabel 4. Hasil pembersihan data

Hasil tokenisasi	Hasil pembersihan data
RT, @ariagaana:, telkomsetl, indihome, suatu, telkomsetl, + indihome, kombinasi, yg mantab suatu, kombinasi, yg mantab, ðŸ˜ˆ• ðŸ˜ˆ• »	RT, @ariagaana:, telkomsetl, indihome, suatu, kombinasi, yg mantab
Hai, @Telkomsetl, lagi, error, ya, ngisi, pulsa, dari, error, ya?, ngisi, pulsa, dari, tadi, pagi, belum, masuk	Hai, @Telkomsetl, lagi, error, ya, ngisi, pulsa, dari, tadi, pagi, belum, masuk
@triindonesia, Kuota, bejibun, harga, merakyat, bejibun, harga, merakyat, ga, nyesel, pake, tri, Ga, nyesel, pake, tri, untungpakaitri #UntungPakaiTri	@triindonesia, Kuota, bejibun, harga, merakyat, ga, nyesel, pake, tri, untungpakaitri
@triindonesia Hallo tri, ini kenapa jaringan tri ku, tak bisa ya pakai mobile data, cek, pulsa, juga, gal, bisa	@triindonesia Hallo tri, ini kenapa jaringan tri, ku, tak, bisa ya, pakai, mobile, data, cek, pulsa, juga, gal, bisa

Setelah data sudah bersih, dilakukan penghapusan kata yang dianggap tidak memiliki makna penting pada data tweet. Contoh kata yang dihapus adalah kata hubung atau kata depan seperti “di”, “ke”, “dan”,

“dengan”, atau “ini”. Proses penghapusan kata yang tidak penting dijelaskan pada Gambar 7. Hasil penghapusan kata ditampilkan pada Tabel 5.

```
# remove stopwords
stop_words = set(stopwords.words('indonesian'))
def remove_stopwords(text):
    text = [i for i in text if not i in stop_words]
    return text

# lemmatize
lemmatizer = WordNetLemmatizer()
def lemmatize(text):
    text = [lemmatizer.lemmatize(token) for token in text]
    return text
```

Gambar 7. Proses *stopword removal*

Tabel 5. Hasil proses *stopword removal*

Hasil pembersihan data	Hasil proses <i>stopword removal</i>
telkomsel, indihome, suatu, kombinasi, yg mantab	telkomsel, indihome, suatu, kombinasi, mantab
lagi, error, ya, ngisi, pulsa, dari, tadi, pagi, belum, masuk	lagi, error, isi, pulsa, dari, tadi, pagi, belum, masuk
kuota, bejibun, harga, merakyat, ga, nyesel, pake, tri, untungpakaitri	kuota, bejibun, harga, rakyat, ga, nyesel, tri, untungpakaitri
hallo, tri, ini kenapa, jaringan, tri, ku, tak, bisa ya, pakai, mobile, data, cek, pulsa, juga, gak, bisa	hallo, tri, ini kenapa, jaringan, tri, tak, bisa, pakai, mobile, data, cek, pulsa, juga, gak, bisa

Tahap terakhir pada *preprocessing* data adalah melakukan stemming. Tahap ini dilakukan untuk menghapus imbuhan pada token. Proses tersebut dijelaskan pada Gambar 8.

```
#stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()
def stemming(text):
    text = [stemmer.stem(token) for token in text ]
    return text
```

Gambar 8. Proses *stemming data*

Hasil keseluruhan *preprocessing*, berupa data bersih yang sudah siap untuk dilabeli dijelaskan pada Gambar 9 dan Gambar 10. Gambar 9 menjelaskan hasil akhir untuk data Telkomsel dan Gambar 10 menjelaskan hasil akhir untuk data Tri.

C	D	E
pp_text		
halo nelpon menit putus ya		
gak ambil voucher gamesmax si		
gatau provider tsel jaring suka i		
makasih udah nunggu kak kelu		
oke		
tolong cek nomor telepon ga di		
rt telkomsel skrg udh ya jaring t		
randa pagi kak daniil maaf ya ka		
min gimana nih beli paket kuota		
telkomselnogame provider not		
operational nombok ma		
parah ya		
provider telkomsel rumah aja		
gaa hai kak pratama maaf ya ka		
iya sist kesel banget telkomsel		
gw kek tai mentang mentang ca		
udah ngerefresh pindah mode j		
hayang ngeretweet gak		
galau banget ganti nomer pake		
hai kak maaf ya ga nyaman ken		
min bbrp kali coba pake kuota		

Gambar 9. Hasil data *preprocessing* operator Telkomsel

C	D	E
pp_text		
bikin malu aja postingan twitter t		
halo kartu tri mohon henti layan k		
sudahh mbakk matiin hp sudahh		
alah mulu solusi taik ledik emng u		
selamat mlm min nanya knp inter		
min jaringan tri daerah tangerang		
emang tri ganggu kah beli vouche		
aplikasi kak pakai tri kode otp lan		
min tri lelet banget si		
min gabisa login gojek gara nome		
min jaring kartu tri kemarin ilang		
jaring tri ngaco bgt sumpedahh		
parah bos sinyal parah afk mulu g		
min gua nelpon pake tri kenapa s		
selamat siang min min kendala k		
tri lot kak Palembang ganggu ya k		
minn tolongg lahh bantuuu sayaa		
signal tri jelek gak kayak		
nya daerah nyah udah lintasin tri		
kak kartu tri hilang aktif aktif		
pgn pke tri thn pgn ganti kartu no		

Gambar 10. Hasil data *preprocessing* operator Tri

Pada Tabel 2 diketahui bahwa terdapat banyak data yang perlu dihapus karena tidak sesuai dengan konteks. Operator telekomunikasi “Smartfren” memiliki data dengan ketidaksesuaian tertinggi. Hal ini dikarenakan pada rentang waktu pengumpulan data, operator tersebut sedang mengadakan promo dan kuis pertanyaan berhadiah sehingga pengguna memberikan jawaban atas pertanyaan tersebut berulang kali. Data tersebut dianggap tidak sesuai karena tidak termasuk ke dalam konteks performansi. Oleh karena itu, data operator telekomunikasi “Smartfren” tidak diolah lebih lanjut dikarenakan dianggap tidak mencukupi.

3.2. Data Labelling

Selanjutnya dari data yang telah bersih, dilakukan pelabelan dengan memberikan sentimen “Positive” atau “Negative” pada masing-masing tweet. Pelabelan dilakukan oleh satu orang. Hasil pelabelan dijelaskan pada. Pada tabel tersebut diketahui bahwa untuk operator telekomunikasi “Telkomsel” terdapat 44 opini positif dan 192 opini negatif. Sedangkan untuk operator telekomunikasi “Tri” terdapat 154 opini positif dan 98 opini negatif.

Selanjutnya dilakukan pelabelan data menggunakan VADER *lexicon*. Data diubah ke dalam bahasa inggris kemudian dilabeli, sehingga didapatkan sentiment dari masing-masing data. Pada Gambar 11 dijelaskan koding untuk proses *import* VADER *lexicon*.

```
import numpy as np
import nltk
nltk.download('vader_lexicon') #download the vader lexicon

from nltk.sentiment.vader import SentimentIntensityAnalyzer

#create a sentiment analyzer object
sid = SentimentIntensityAnalyzer()
```

Gambar 11. Proses *import* VADER *lexicon*

Setelah dilakukan proses *import library*, selanjutnya dilakukan pengubahan bahasa dari bahasa Indonesia ke bahasa Inggris untuk mempermudah pelabelan. Gambar 12 menjelaskan coding untuk proses pembobotan kata.

```
#define a function to map the compount scores to sentiment labels
def get_sentiment_label(score):
    if score >= 0.05:
        return 'Positive'
    elif score <= -0.05:
        return 'Negative'
    else:
        return 'Neutral'

#apply the function to the sentiment score column to create a new sentiment
#label column
sentiment_labels = df['sentiment_score'].apply(get_sentiment_label)

#add the sentiment labels as a new column
df['sentiment'] = sentiment_labels
```

Gambar 12. Proses pembobotan data

Gambar 13 menjelaskan hasil akhir pelabelan data dengan setiap data sudah diberi label sentimen positif maupun negatif.

Sentimen	pp_text
Negative	bikin malu aja postingan twitter tri keluh yg ...
Negative	halo kartu tri mohon henti layan konten sm pul...
Negative	sudahh mbakk matiin hp sudahh lelet suruh face...
Negative	alah mulu solusi taik ledik emng udah dasar je...
Negative	met mlm min nanya knp internet nya ga b dipake...
Negative	min jari tri daerah tangerang normal ya malem ...
Negative	emng tri ganggu kah beli voucher kuota ga masu...
Positive	aplikasi kak pakai tri kode otp langsung masuk...
Negative	min tri lelet banget si

Gambar 13. Hasil proses *labelling* data Tri

Selanjutnya, data yang telah dilabeli kemudian diolah untuk menjadi data terstruktur. Proses preprocessing dilakukan dengan menggunakan pemograman python. Hasil pelabelan disimpan ke dalam bentuk csv untuk mempermudah pemodelan data. Hasil pelabelan dijelaskan pada Tabel 6.

Tabel 6. Hasil data labelling

Operator	Opini Positif	Opini Negatif
Telekomunikasi		
Telkomsel	44	192
Tri	154	98

3.3. Data Modelling dan Validasi

Penelitian ini menggunakan tiga metode, yaitu Naïve Bayes Classifier, Support Vector Machine dan Decision Tree Learning. Untuk mengetahui metode yang memiliki akurasi paling tinggi, dilakukan validasi dengan dua teknik, yaitu *split validation* dan *cross validation*.

1. Split Validation

Pada bagian ini data dibagi menjadi dua bagian, yaitu data training dan data testing. Pembagian tersebut menggunakan rasio 70:30, dimana 70% adalah data training dan 30% data testing. Pembagian data dijelaskan pada Tabel 7.

Tabel 7. Pemisahan data dengan *split validation*

Tipe data	Telkomsel	Tri
Data training	165	176
Data testing	71	76

Selanjutnya dilakukan training data untuk setiap metode. Model yang sudah ditraining kemudian diuji dengan menggunakan data testing untuk mengetahui akurasi dari masing-masing metode. Hasil pengujian dijelaskan pada Tabel 8.

Tabel 8. Hasil pengujian dengan *split validation*

Operator	Metode		
	Naïve Bayes	DTL	SVM
Telkomsel	81.40%	81.43%	84.29%
Tri	74.67%	50.67%	75.70%

Berdasarkan pemodelan data dengan teknik *split validation* didapatkan hasil bahwa metode Support Vector Machine memiliki akurasi paling tinggi, baik untuk “Telkomsel” maupun “Tri”. Akurasi dengan metode SVM untuk “Telkomsel” adalah 84.28% dan untuk “Tri” adalah 75.70%. Pemodelan data dengan metode lain, yaitu Naïve Bayes Classifier menghasilkan akurasi 81.40% untuk “Telkomsel” dan 74.67% untuk “Tri”. Sedangkan dengan metode Decision Tree Learning akurasi “Telkomsel” adalah 81.43% dan “Tri” adalah 50.67%.

2. Cross Validation

Pada bagian ini pemodelan pada masing-masing metode menggunakan *cross validation*. Pada fungsi *cross validation* terdapat parameter *fold* yang dapat diatur nilainya agar memberikan hasil pemodelan yang paling baik. Pada Tabel 9 dan Tabel 10 dijelaskan pemodelan berdasarkan nilai *fold* yang ditentukan.

Tabel 9. Hasil pengujian data Telkomsel dengan *cross validation*

Telkomsel			
k	Naive Bayes	DTL	SVM
2	82.12%	81.29%	81.29%
3	82.14%	81.25%	81.25%
4	82.09%	81.31%	82.16%
5	82.13%	81.28%	81.28%
6	82.12%	81.27%	81.27%
7	82.12%	81.31%	82.16%
8	82.13%	81.32%	82.13%
9	82.15%	81.24%	82.10%
10	82.10%	81.33%	81.32%

Tabel 10. Hasil pengujian data Tri dengan *cross validation*

Tri			
k	Naive Bayes	DTL	SVM
2	61.33%	61.36%	61.37%
3	61.35%	61.35%	61.38%
4	61.32%	61.35%	61.31%
5	61.36%	61.35%	61.31%
6	61.33%	61.35%	61.34%
7	61.35%	61.35%	61.39%
8	61.32%	61.34%	61.27%
9	61.37%	61.37%	61.41%
10	61.34%	61.35%	61.31%

Dari perhitungan akurasi didapatkan hasil bahwa untuk data “Telkomsel” dengan metode Naive Bayes Classifier didapatkan akurasi tertinggi 82.15% pada k sama dengan 9. Sedangkan untuk metode Decision Tree Learning, akurasi tertinggi adalah 81.33% pada k sama dengan 10. Selanjutnya, untuk metode Support Vector Machine akurasi tertinggi adalah 82.16% pada k sama dengan 4 dan 7. Oleh karena itu, dapat diketahui untuk data “Telkomsel” metode yang memiliki akurasi paling tinggi adalah Support Vector Machine.

Pada data “Tri” didapatkan hasil akurasi tertinggi dengan metode Naive Bayes Classifier adalah 61.37% pada k sama dengan 9. Hasil tersebut sama dengan pemodelan metode Decision Tree Learning dengan hasil akurasi 61.37% pada k sama dengan 9. Sedangkan pada pemodelan metode Support Vector Machine akurasi tertinggi adalah 61.41% pada k sama dengan 9. Oleh karena itu, dapat diketahui pemodelan dengan metode Support Vector Machine memiliki akurasi paling tinggi untuk data “Tri”.

3.3. Evaluasi dan Analisis

Berdasarkan pengolahan data yang telah dilakukan sebelumnya, didapatkan komposisi opini negatif dan positif yang diberikan oleh pengguna Twitter pada rentang waktu pengumpulan data. Komposisi opini tersebut dijelaskan pada Tabel 11. Pada tabel tersebut dapat diketahui bahwa data “Telkomsel” memiliki lebih banyak opini negatif, sedangkan data “Tri” memiliki lebih banyak opini positif.

Tabel 11. Perbandingan persentase sentimen

Operator Telekomunikasi	Persentase Sentimen Positif	Persentase Sentimen Negatif
Telkomsel	18.64%	81.36%
Tri	61.11%	38.89%

Selain itu pada Tabel 12 diketahui hasil dari perhitungan akurasi yang dihitung dan divalidasi dengan dua teknik validasi didapatkan hasil bahwa akurasi tertinggi untuk teknik split validation adalah pemodelan dengan metode Support Vector Machine. Hasil tersebut berlaku untuk dua data, yaitu “Telkomsel” dengan 84.28% dan “Tri” dengan 75.70%. Validasi dengan teknik cross validation juga mendapatkan akurasi tertinggi pada pemodelan dengan metode Support Vector Machine. Data “Telkomsel” memiliki akurasi 82.16%, sedangkan data “Tri” memiliki akurasi 61.41%. Sehingga dapat diketahui bahwa teknik split validation menghasilkan akurasi yang lebih tinggi daripada teknik cross validation.

Tabel 12. Perbandingan hasil dari dua teknik validasi

Operator Telekomunikasi	Split Validation	Cross Validation
Telkomsel	84.28%	82.16%
Tri	75.70%	61.41%

Kemudian dilakukan analisis lebih lanjut dengan memerhatikan opini yang diuji serta dilakukan perhitungan terhadap kata-kata yang paling banyak keluar untuk masing-masing opini. Hasil yang didapatkan ditunjukkan pada Tabel 13.

Tabel 13. Karakteristik sentimen

Operator Telekomunikasi	Sentimen Positif	Sentimen Negatif
Telkomsel	Memiliki layanan <i>platform</i> video	Jaringan kurang stabil
	Kecepatan internet baik	Harga paket internet relatif mahal
	Sinyal menjangkau sampai daerah	Paket internet tidak dapat digunakan padahal sudah terbayar
Tri	Harga paket internet relatif terjangkau	Sinyal buruk
	Telah menggunakan operator sejak lama	Sering terjadi gangguan ketika hujan atau mati listrik
	Pelayanan pelanggan baik	Aplikasi Bima+ sering mengalami gangguan
	Masa aktif kartu lebih lama	Kecepatan internet kurang

4. Kesimpulan

Berdasarkan hasil pemodelan data didapatkan hasil bahwa metode Support Vector Machine memiliki akurasi yang paling baik dibandingkan metode Naive Bayes Classifier dan Decision Tree Learning pada data yang digunakan dan dua teknik validasi yang dilakukan. Pada penelitian ini ditemukan bahwa akurasi dengan teknik split validation lebih tinggi daripada cross validation. Selain itu, sebanyak 18.64% pengguna memberikan sentimen positif dengan pelayanan yang diberikan oleh Telkomsel, sedangkan sisanya sebanyak 81.36% memberikan sentiment negatif. Selain itu, sebanyak 61.11% pengguna memberi sentimen positif yang diberikan oleh Tri, sedangkan sisanya sebanyak 38.89% memberikan sentimen negatif. Hal ini menunjukkan operator telekomunikasi perlu melakukan perbaikan layanan khususnya untuk jaringan internet yang kurang stabil karena dapat mempengaruhi sentimen masyarakat terhadap layanan.

Penelitian ini baru meneliti mengenai dua operator telekomunikasi Indonesia. Oleh karena itu, rekomendasi yang diusulkan untuk penelitian selanjutnya adalah melakukan penelitian untuk operator telekomunikasi lain supaya data yang dikumpulkan lebih banyak dan dapat lebih memberikan kontribusi. Kemudian, melakukan labelling data kepada tiga orang sehingga hasil pelabelan tidak terlalu subyektif. Penelitian selanjutnya juga dapat melakukan analisis sentimen untuk operator telekomunikasi lain di luar negeri.

SUMBER RUJUKAN

Referensi

- [1] S. Kemp, "Digital 2024: Indonesia," Feb. 2024.
- [2] I. A. N. Gunawan, S., and I. Shalahuddin, "Dampak Penggunaan Media Sosial Terhadap Gangguan Psikososial Pada Remaja: A Narrative Review," *Jurnal Kesehatan*, vol. 15, no. 1, pp. 78–92, Jun. 2022, doi: 10.23917/jk.v15i1.17426.
- [3] H. P. Elisa, M. Fakhri, and M. Pradana, "The moderating effect of social media use in impulsive buying of personal protective equipments during the COVID-19 pandemic," *Cogent Soc Sci*, vol. 8, no. 1, Dec. 2022, doi: 10.1080/23311886.2022.2062094.
- [4] L. R. Megawati, "The Analysis of Effects of Operating Leverage, Financial Leverage, and Liquidity on Profitability in the Telecommunications Industry Listed in Indonesia Stock Exchange," in *Proceedings of the First ASEAN Business, Environment, and Technology Symposium (ABEATS 2019)*, Paris, France: Atlantis Press, 2020. doi: 10.2991/aebmr.k.200514.025.
- [5] R. Wyrzykowski, "Mobile Network Experience Report," Indonesia, Dec. 2024.
- [6] OpenSignal, "OpenSignal Methodology," <https://www.opensignal.com/our-approach/mobile-metrics>.
- [7] A. Novanto, D. Indra, and W. Astuti, "Analisis Pre-processing Sentimen Terhadap Komentar Layanan Indihome Pada Twitter," *Literatur Informatika & Komputer*, vol. 1, no. 2, pp. 145–152, 2024, doi: 10.33096/linier.vxix.xxxx.
- [8] D. Puspitasari, N. Nazhiifah, and T. Sutabri, "Analisis Sentimen Pada Ulasan Mobile Jkn Berdasarkan Pada Media Sosial Twitter (X) Menggunakan Metode Support Vector Machine (SVM)," *Nusantara Journal of Multidisciplinary Science*, vol. 2, no. 6, pp. 1273–1282.
- [9] D. Sarkar, "Sentiment Analysis," in *Text Analytics with Python*, Berkeley, CA: Apress, 2019, pp. 567–629. doi: 10.1007/978-1-4842-4354-1_9.
- [10] D. R. Lazuardi, T. A. Munandar, H. Harsiti, Z. Mutaqin, and R. N. Hays, "Sentiment analysis of public opinions on the welfare of honorary educators using Naive Bayes," *IOP Conf Ser Mater Sci Eng*, vol. 830, no. 3, p. 032018, Apr. 2020, doi: 10.1088/1757-899X/830/3/032018.
- [11] E. E. Pratama, Helen Sastypratiwi, and Yulianti, "Analisis Kecenderungan Informasi Terkait Covid-10 Berdasarkan Big Data Sosial Media dengan Menggunakan Metode Data Mining," *Jurnal Informatika Polinema*, vol. 7, no. 2, pp. 1–6, Feb. 2021, doi: 10.33795/jip.v7i2.453.
- [12] V. A. Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *Jurnal Buana Informatika*, vol. 11, no. 2, pp. 141–151, Oct. 2020, doi: 10.24002/jbi.v11i2.3769.
- [13] M. Rezki, D. N. Kholifah, M. Faisal, P. Priyono, and R. Suryadithia, "Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes," *Jurnal Infotech*, vol. 2, no. 2, pp. 264–270, Dec. 2020, doi: 10.31294/infotech.v2i2.9286.
- [14] F. Sudirjo, K. Diantoro, J. A. Al-Gasawneh, H. Khoonimah Azzaakiyyah, and A. M. Almaududi Ausat, "Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 3, pp. 283–288, Jul. 2023, doi: 10.47233/jteksis.v5i3.871.
- [15] F. Fridom Mailo *et al.*, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," 2019.
- [16] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *IJCSI International Journal of Computer Science Issues*, vol. 4, no. 1, 2009.
- [17] C. Cortes and V. Vapnik, "Support-vector networks," *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [18] S. Shalehanny, A. Triayudi, and E. T. E. Handayani, "PUBLIC'S SENTIMENT ANALYSIS ON SHOPEE-FOOD SERVICE USING LEXICON-BASED AND SUPPORT VECTOR MACHINE," *Jurnal Riset Informatika*, vol. 4, no. 1, pp. 1–8, Dec. 2021, doi: 10.34288/jri.v4i1.287.
- [19] R. Nooraeni, H. D. Sariyanti, A. F. F. Iskandar, S. F. Munawwaroh, S. Pertiwi, and Y. Ronaldias, "Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM)," *Paradigma - Jurnal Komputer dan Informatika*, vol. 22, no. 1, pp. 55–60, Mar. 2020, doi: 10.31294/p.v22i1.6869.
- [20] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjee.v5i1.16830.
- [21] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," 2014. [Online]. Available: <http://sentic.net/>
- [22] W. Silalahi and A. Hartanto, "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024," *JRST (Jurnal Riset Sains dan Teknologi)*, vol. 7, no. 2, p. 245, Sep. 2023, doi: 10.30595/jrst.v7i2.18133.
- [23] H. Hafid, "Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia." [Online]. Available: <http://www.ojs.unm.ac.id/jmathcos>
- [24] J. J. Purnama, Nina Kurnia Hikmawati, and Sri Rahayu, "Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko Kesehatan Ibu Hamil," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 120–127, Jun. 2024, doi: 10.52158/jacost.v5i1.809.