

## Perbandingan Naive Bayes Classifier dan Support Vector Machine untuk Analisis Sentimen terhadap Penyebaran Nyamuk Wolbachia di Indonesia

Nurul Hijriani<sup>1</sup>, Ermatita<sup>\*2</sup>

<sup>1,2</sup>Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia  
Email: <sup>1</sup>[09012682327009@student.unsri.ac.id](mailto:09012682327009@student.unsri.ac.id), <sup>2</sup>[ermatita@unsri.ac.id](mailto:ermatita@unsri.ac.id)

### Abstrak

Demam Berdarah *Dengue* (DBD) merupakan penyakit endemik yang terus menjadi ancaman kesehatan masyarakat di Indonesia. Salah satu solusi inovatif yang diterapkan adalah pelepasan nyamuk *Aedes aegypti* yang terinfeksi *Wolbachia* untuk mengurangi penularan virus DBD. Program ini mendapat beragam tanggapan publik di media sosial, seperti Instagram, yang dapat memengaruhi keberhasilannya. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait program tersebut menggunakan algoritma *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM). Sebanyak 3.619 komentar Instagram dikumpulkan melalui scraping, kemudian diproses menggunakan *text preprocessing* dengan metode N-Gram dan TF-IDF untuk ekstraksi fitur. Untuk mengatasi ketidakseimbangan data, diterapkan Synthetic Minority Oversampling Technique (SMOTE), dan data divalidasi menggunakan K-Fold Cross Validation dengan k adalah 10. Hasil menunjukkan bahwa SVM memiliki *accuracy* lebih tinggi sebesar 83,24% dibandingkan NBC yang mencapai 82,15%. SVM unggul dalam mendeteksi sentimen positif dengan *recall* sebesar 85,39%, sementara NBC menunjukkan *precision* lebih tinggi untuk sentimen negatif sebesar 86,36%. Hasil ini menunjukkan bahwa SVM lebih efektif untuk mengukur dukungan publik, sedangkan NBC lebih sesuai untuk mengidentifikasi kritik atau kekhawatiran. Temuan ini memberikan wawasan praktis bagi pemerintah dan pemangku kepentingan dalam meningkatkan strategi komunikasi kesehatan berbasis data untuk mendukung keberhasilan program kesehatan masyarakat.

**Kata kunci:** Analisis Sentiment, Demam Berdarah Dengue, Instagram, Naive Bayes Classifier, Support Vector Machine, Wolbachia

### *Comparison of Naive Bayes Classifier and Support Vector Machine for Sentiment Analysis on the Release of Wolbachia Mosquitoes in Indonesia*

#### Abstract

*Dengue Hemorrhagic Fever (DHF) remains an endemic disease and a significant public health concern in Indonesia. One innovative solution implemented is the release of Aedes aegypti mosquitoes infected with Wolbachia bacteria to reduce dengue virus transmission. This program has received mixed responses from the public on social media platforms such as Instagram, which can influence its success. This study aims to analyze public sentiment regarding the program using Naive Bayes Classifier (NBC) and Support Vector Machine (SVM). A total of 3.619 Instagram comments were collected through scraping, processed using text preprocessing with N-Gram and TF-IDF methods for feature extraction. To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, and the data was validated using K-Fold Cross Validation with k set to 10. Results show that SVM achieved higher accuracy at 83.24% compared to NBC at 82.15%. SVM excelled in detecting positive sentiment with a recall of 85.39%, while NBC demonstrated higher precision for negative sentiment at 86.36%. These findings indicate that SVM is more effective in measuring public support, while NBC is better suited for identifying criticism or concerns. This study provides practical insights for the government and stakeholders to improve data-driven health communication strategies, supporting the success of public health programs.*

**Keywords:** dengue fever, Instagram, Naive Bayes Classifier, sentiment analysis, Support Vector Machine, Wolbachia

## 1. PENDAHULUAN

Demam Berdarah Dengue (DBD) tetap menjadi masalah kesehatan di Indonesia dengan angka kejadian tinggi, yaitu 52 per 100.000 penduduk pada 2022 [1]. Salah satu inovasi untuk mengendalikan penyebaran

adalah pelepasan nyamuk *Wolbachia* [2]. Implementasi program ini memunculkan beragam opini masyarakat, yang diekspresikan melalui platform media sosial seperti Instagram. Sebagai platform dengan karakteristik interaksi visual dan naratif, Instagram memungkinkan pengguna untuk mengungkapkan pendapat, kekhawatiran, dan dukungan mereka. Hal ini menjadikan Instagram sebagai sumber data sentimen publik yang unik dan kaya dibandingkan media sosial lainnya [3].

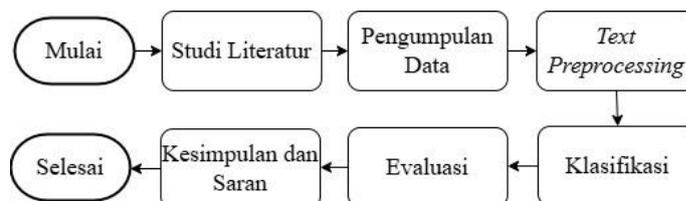
Analisis sentimen di media sosial menjadi alat penting untuk memahami persepsi masyarakat terhadap program kesehatan. Dua algoritma klasifikasi yang sering digunakan dalam analisis sentimen adalah *Naive Bayes Classifier* (NBC) [4] dan *Support Vector Machine* (SVM) [5]. Penelitian sebelumnya menunjukkan bahwa NBC unggul dalam efisiensi waktu proses, sementara SVM cenderung lebih akurat dalam klasifikasi [6]. Misalnya, penelitian Alvin Irwanto (2023) menunjukkan bahwa NBC lebih efisien dalam analisis sentimen vaksinasi Covid-19, tetapi SVM memiliki presisi lebih tinggi [6]. Penelitian Staphord Bengesi (2023) juga mengungkapkan bahwa SVM, bersama dengan TextBlob dan CountVectorizer, mencapai akurasi tertinggi dalam analisis sentimen terkait wabah monkeypox [7]. Penelitian lain oleh R. Kusumawati (2019) menunjukkan bahwa SVM unggul dalam mengklasifikasikan opini pengguna Twitter tentang layanan Tokopedia [8], sementara E. Prem Vardhan Reddy (2024) menyimpulkan bahwa SVM lebih akurat daripada NBC dalam memprediksi tingkat kepuasan penumpang maskapai [9].

Sebagian besar penelitian tersebut menggunakan platform Twitter sebagai sumber data utama. Namun, Instagram, yang memiliki karakteristik interaksi visual dan naratif yang unik, belum banyak dieksplorasi. Instagram memungkinkan pengguna mengekspresikan sentimen mereka melalui teks dan visual, memberikan perspektif yang lebih mendalam dalam memahami reaksi publik terhadap program kesehatan seperti pelepasan nyamuk *Wolbachia*. Studi terhadap virus Zika sebelumnya menunjukkan bahwa platform ini dapat mengungkap ketakutan, misinformasi, dan dukungan publik yang signifikan [10].

Penelitian ini bertujuan untuk mengisi celah tersebut dengan membandingkan efektivitas NBC dan SVM dalam menganalisis sentimen publik di Instagram terkait program pelepasan nyamuk *Wolbachia* di Indonesia. Pemahaman mendalam tentang sentimen publik ini penting bagi pemangku kepentingan untuk merancang strategi komunikasi yang lebih responsif, mengetahui serta meningkatkan penerimaan Masyarakat terhadap program tersebut.

## 2. METODE PENELITIAN

Langkah-langkah penelitian ini mencakup beberapa tahapan, mulai dari pengumpulan data hingga evaluasi model klasifikasi. Gambar 1 menggambarkan alur yang menjelaskan proses dari penelitian ini.



Gambar 1. Alur Penelitian

### 2.1. Studi Literatur

Pada tahap ini, dilakukan studi literatur dengan tujuan untuk membangun landasan teoritis yang kuat mengenai analisis sentimen, algoritma NBC, SVM, serta teknik text preprocessing dan ekstraksi fitur. Studi ini juga mencakup penelitian-penelitian terdahulu yang menggunakan algoritma NBC dan SVM dalam konteks analisis sentimen di media sosial.

### 2.2. Pengumpulan Data

Tahap ini adalah melakukan pengumpulan data yang akan diproses lebih lanjut. Proses pengumpulan data dimulai dengan membuka situs Apify di apify.com dan mendaftar akun untuk mendapatkan akses ke layanan tersebut. Setelah berhasil melakukan registrasi, peneliti masuk ke *console* Apify untuk memulai proses pengumpulan data. Peneliti mengumpulkan data dari beberapa postingan dengan kata kunci '*Wolbachia*', '*Nyamuk Wolbachia*', dan '*Wolbachia Nyamuk Baik*'. Data komentar tersebut diambil dari tanggal 28 Juni 2021 hingga 19 Desember 2023, rentang waktu tersebut relevan untuk mencerminkan respons publik terhadap implementasi program ini.

### 2.3. Text Preprocessing

Setelah data dikumpulkan, dilakukan serangkaian proses *text preprocessing* untuk membersihkan dan menyiapkan data untuk analisis. Tahapan *text preprocessing* meliputi *case folding*, *cleansing*, *tokenization*, *stopword removal*, *normalization*, dan *stemming* [11]. Proses ini bertujuan untuk menyederhanakan data teks sehingga dapat digunakan dalam langkah klasifikasi.

*Case folding* merupakan tahap untuk mengubah semua teks menjadi huruf kecil, *cleansing* menghapus karakter non-alfabet yang meliputi angka, tanda baca dan *emoticon*, *tokenization* memisahkan teks menjadi per-kata atau token, *normalization* bertujuan untuk mengubah kata-kata tidak baku, singkatan, atau slang yang ditemukan dalam teks menjadi bentuk yang standar atau baku, *stopword removal* yaitu menghapus kata-kata penghubung, dan *stemming* untuk mengubah kata-kata berimbuhan menjadi bentuk dasar atau kata dasar [11].

Setelah *text preprocessing*, selanjutnya dilakukan *labelling* menggunakan *Valance Aware Dictionary and Sentiment Reasoner (VADER)* yang merupakan *tools* untuk melakukan analisis sentimen yang berbasis leksikon dan aturan yang dirancang khusus untuk teks. Fungsinya adalah menentukan apakah teks bersifat negatif, positif atau netral, dan memberikan skor pada teks tersebut antara -1 sampai +1 [12].

*VADER* menghasilkan dua skor utama pada setiap teks yaitu *sentiment score* dan *compound score*. Pada *sentiment score*, mencakup tiga metrik yaitu positif dan negatif untuk mengukur intensitas polaritas teks. Sedangkan *compound score* berupa nilai dengan rentang -1 dan +1, hal tersebut merupakan skor keseluruhan yang dihitung sebagai agregasi dari sentimen positif dan negatif.

Pada proses *labelling*, komentar yang diberi label positif merupakan komentar yang memiliki *compound score* yang lebih besar dari 0 begitu pun sebaliknya untuk label negatif.

Setelah *labelling* dilakukan, maka fitur-fitur diekstraksi menggunakan N-Gram untuk memberikan hasil yang lebih akurat [13]. Model ini untuk menganalisis sentimen teks maupun dokumen serta membantu untuk melakukan klasifikasi sentimen sebagai positif atau negatif. Keuntungan dalam menggunakan kombinasi kata yaitu bigram dan trigram dapat memberikan hasil yang lebih akurat dibandingkan dengan unigram [13]. Unigram memecah kalimat menjadi satu kata per unit, bigram berfokus pada kombinasi dua kata yang berurutan pada sebuah kalimat, sedangkan trigram berfokus pada kombinasi tiga kata yang berurutan pada sebuah kalimat.

Metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. TF-IDF menghitung bobot setiap kata dalam dokumen, memberikan nilai yang lebih tinggi pada kata-kata yang jarang muncul di seluruh dokumen, namun sering muncul di dokumen tertentu [14]. Proses ini membantu meningkatkan akurasi dalam analisis sentimen.

Persamaan (1) untuk menghitung nilai TF sedangkan Persamaan (2) untuk menghitung nilai IDF [14]. Pada Persamaan (1)  $t$  adalah kata tertentu yang dianalisis,  $d$  merupakan dokumen tertentu pada korpus,  $f(t, d)$  merupakan jumlah kemunculan kata  $t$  dan kata  $d$ , kemudian  $N_d$  merupakan total jumlah kata dalam dokumen  $d$ . Pada Persamaan (2)  $D$  adalah korpus atau kumpulan dokumen,  $N$  adalah total jumlah dokumen yang terdapat pada korpus  $D$ , sedangkan  $n_t$  merupakan jumlah dokumen dalam korpus  $D$  yang mengandung kata  $t$ .

$$TF(t, d) = \frac{f(t, d)}{N_d} \quad (1)$$

$$DF(t, D) = \log\left(\frac{N}{n_t}\right) \quad (2)$$

Setelah persamaan tersebut telah dijalankan, kemudian hasil kali antara TF dan IDF memberikan skor yang lebih tinggi untuk kata-kata yang penting atau unik dalam dokumen. Di bawah ini Persamaan (3) dari TF-IDF [14].

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

### 2.4. Klasifikasi

Setelah melalui tahap *text preprocessing*, dataset dibagi menjadi dua bagian yaitu data latih dan data uji. Algoritma NBC dan SVM digunakan untuk membangun model klasifikasi berdasarkan data latih.

Namun sebelum melakukan klasifikasi, kumpulan data tersebut harus menggunakan *Synthetic Minority Oversampling Technique (SMOTE)* terlebih dahulu jika terjadi ketidakseimbangan data. SMOTE merupakan bagian dari teknik *over-sampling* untuk mengatasi ketidakseimbangan data dengan meningkatkan kelas yang minoritas [15]. Pada NBC, klasifikasi dilakukan berdasarkan Teorema Bayes, dengan persamaan dasar yang tertulis pada Persamaan (4) [13].

Pada Persamaan (4),  $P(H|X)$  adalah probabilitas posterior,  $P(X|H)$  adalah *likelihood*,  $P(H)$  adalah *prior probability*, dan  $P(X)$  adalah probabilitas bukti. Sementara itu, SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas positif dan negatif dengan margin terbesar.

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \tag{4}$$

Sementara itu, SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas positif dan negatif dengan margin terbesar. Persamaan dasar dari SVM ini dapat dilihat pada Persamaan (5)[16]. Pada persamaan (5),  $w$  adalah vektor bobot,  $x$  adalah vektor data, dan  $b$  adalah nilai bias.

$$f(x) = w \cdot x + b \tag{5}$$

Alternatifnya, fungsi keputusan dapat ditulis dalam persamaan (6) [14]. Pada persamaan (6),  $a_i$  adalah bobot pada setiap titik data,  $y_i$  adalah label kelas dari setiap titik data, dan  $K(x, x_i)$  adalah fungsi kernel yang mengukur kesamaan antara data masukan  $x$  dan data pendukung  $x_i$ . Jumlah titik data yang digunakan dilambangkan dengan  $m$ .

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \tag{6}$$

Proses pencarian *hyperplane* melibatkan perhitungan margin, yang merupakan jarak antara data terdekat (*support vectors*) dengan *hyperplane* [17]. *Hyperplane* dengan margin terbesar dipilih untuk memastikan model memiliki generalisasi yang baik pada data baru. Untuk menemukan nilai  $w$  dan  $b$ , terlebih dahulu dicari nilai *Lagrange* atau koefisien  $\alpha$  dengan menggunakan persamaan (7) [18]. Pada persamaan (7),  $\alpha_j$  adalah koefisien untuk setiap data latih,  $y_j$  adalah label kelas sedangkan  $x_j \cdot x_i$  adalah hasil perkalian antara data latih dan data masukan.

$$L(\alpha) = 1 - \sum_{j=1}^n \alpha_j \cdot y_j (x_j \cdot x_i) \tag{7}$$

Setelah nilai  $\alpha$  diperoleh, parameter  $w$  dapat dihitung pada persamaan (8) [19] dan bias  $b$  dapat dihitung dengan persamaan (9) [16].

$$w \cdot x_i + b = 0 \tag{8}$$

$$b = y_i - w \cdot x_i \tag{9}$$

Setelah model dilatih menggunakan data latih, baik NBC maupun SVM diterapkan pada data uji untuk memprediksi sentimen dari teks yang belum pernah dilihat sebelumnya.

### 2.5. Evaluasi

Evaluasi model dilakukan menggunakan *confusion matrix*, yang mencakup empat parameter utama yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Parameter-parameter ini akan mengukur performa model dalam mengklasifikasikan sentimen positif dan negatif, serta memberikan gambaran tentang efektivitas kedua algoritma dalam tugas klasifikasi [20]. Tabel 1 menunjukkan elemen-elemen dari *confusion matrix*.

Tabel 1. *Confusion Matrix*

Data Aktual	Positive	Neutral	Negative
Positive	True Positive (TP)	False Neutral (FNe)	False Negative (FN)
Neutral	False Positive (FP)	True Neutral (TNe)	False Negative (FN)
Negative	False Positive (FP)	False Neutral (FNe)	True Negative (TN)

Berdasarkan Tabel 1, *accuracy* akan menghitung seberapa sering model membuat prediksi yang benar, dihitung dengan membagi jumlah prediksi benar dengan total jumlah data, *precision* digunakan untuk mengukur ketepatan prediksi positif model, yaitu seberapa banyak prediksi positif yang benar dari semua prediksi positif, *recall* mengukur kemampuan model untuk menemukan semua contoh positif yang sebenarnya, dan *f1-score* menggabungkan nilai *precision* dan *recall* menjadi satu metrik tunggal dengan menggunakan rata-rata harmonis [21].

Persamaan (10) merupakan cara menghitung *accuracy*, Persamaan (11) untuk menghitung *precision*, Persamaan (12) untuk menghitung *recall* dan Persamaan (13) untuk menghitung *f1-score*.

$$accuracy = \frac{TP+TN+TNe}{TP+FP+FN+FNe+TN+TNe} \tag{10}$$

$$precision = \frac{TP}{TP+FP+FPe} \tag{11}$$

$$recall = \frac{TP}{TP+FN+FNe} \tag{12}$$

$$f1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Proses pengumpulan data ini, terdapat 3.619 komentar berhasil dikumpulkan. Dari jumlah ini, data menunjukkan ketidakseimbangan dalam distribusi sentimen, dengan 1.437 komentar positif yang artinya sebesar 39.72% komentar positif dan 2.231 komentar negatif yang artinya 60.28% komentar negatif.

Distribusi awal ini mencerminkan bahwa sentimen publik terhadap program cenderung lebih banyak bersifat negatif. Statistik data awal ini ditampilkan dalam Tabel 2.

Tabel 2. Statistik Data Awal

Sentimen	Jumlah Komentar	Presentase (%)
Positif	1.437	39.72%
Negatif	2.231	60.28%

Berdasarkan Tabel 2, data komentar mencakup berbagai opini, mulai dari dukungan terhadap program hingga kekhawatiran terkait dampaknya. Data yang terkumpul disimpan dalam format CSV untuk dianalisis lebih lanjut. Tabel 3 berikut ini merupakan tampilan dari sampel kumpulan data.

Tabel 3. Sampel Kumpulan Data

<i>no</i>	<i>text</i>
1	Kenapa harus banget di bandung ??? Wilayah indonesia kan luas
2	Kak @wirangbirawa bagaimana cara mencegahnya ini..
3	@soeryanieyanie dapat info dr mana kak, jelas2 Singapura menolak
4	tapi gatal tidak waktu digigit?
5	kalo menurut anda aman, saya tantang di rumah anda dilepasin nyamuk wolbachianya
6	Sekarang orang pada jahat, rakus banget sama dunia
...	...
(Total: 3619 komentar)	Data dilanjutkan...

Contoh data yang dikumpulkan dan disajikan pada Tabel 3 menunjukkan berbagai opini publik mengenai program tersebut. Setelah dikumpulkan, data mentah ini diproses lebih lanjut melalui tahapan *text preprocessing* untuk membersihkan dan menyederhanakan teks.

#### 3.2. Text Preprocessing

Setelah data dikumpulkan, langkah selanjutnya adalah melakukan *text preprocessing* untuk membersihkan data dari karakter yang tidak relevan. Hasil dari *text preprocessing* dapat dilihat pada Tabel 4.

Pada Tabel 4 menggunakan salah satu komentar yang diambil dari Kumpulan data yaitu “Negeri kita sudah aman gk perlu nyamuk baik” kemudian dapat dilihat hasilnya setiap tahap pada kolom ‘Sesudah.’

Tabel 4. Hasil *Text Preprocessing*

Alur Text Preprocessing	Sebelum	Sesudah
Casefolding	Negeri kita, sudah aman gk perlu nyamuk baik	negeri kita, sudah aman gk perlu nyamuk baik
Cleansing	negeri kita, sudah aman gk perlu nyamuk baik	negeri kita sudah aman gk perlu nyamuk baik
Tokenization	negeri kita sudah aman gk perlu nyamuk baik	['negeri', 'kita', 'sudah', 'aman', 'gk', 'perlu', 'nyamuk', 'baik']
Normalization	['negeri', 'kita', 'sudah', 'aman', 'gk', 'perlu', 'nyamuk', 'baik']	['negeri', 'kita', 'sudah', 'aman', 'tidak', 'perlu', 'nyamuk', 'baik']
Stopword Removal	['negeri', 'kita', 'sudah', 'aman', 'tidak', 'perlu', 'nyamuk', 'baik']	['negeri', 'aman', 'tidak', 'nyamuk']
Stemming	['negeri', 'aman', 'tidak', 'nyamuk']	negeri aman tidak nyamuk
Translated Text	negeri aman tidak nyamuk	safe country without mosquitoes

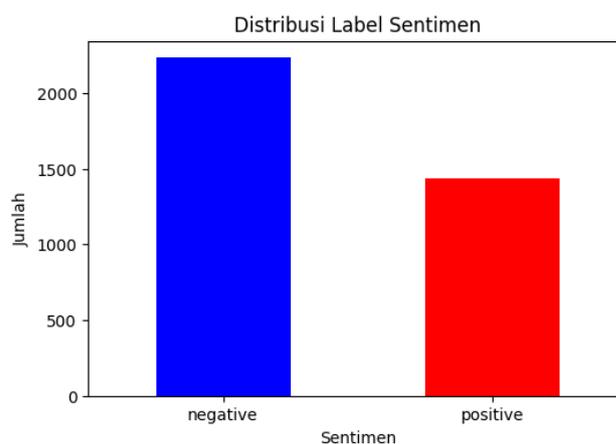
Selanjutnya, dilakukan pelabelan atau *labelling* sebagai bagian dari *text preprocessing*. Data diterjemahkan dan disimpan kemudian dianalisis menggunakan *Vader Sentiment Analyzer* untuk menentukan skor sentimen (positif, negatif, dan *compound*). Hasil pelabelan ini membantu model dalam mempelajari dan menentukan sentimen dengan akurat. Pada Tabel 5 adalah beberapa contoh dari hasil *labelling* ini.

Tabel 5. Hasil *Labelling*

Cleaned Text	Sentiment Score	Compound Score	Final Label
<i>Mosquitoes bring illness</i>	{'negative': 0.574, 'positive': 0.0, 'compound': -0.4019}	-4019	<i>Negative</i>
<i>The main thing is to pray don t think negatively</i>	{'negative': 0.0, 'positive': 0.204, 'compound': 0.3182}	3182	<i>Positive</i>

Pada Tabel 5 menunjukkan *sentiment score* dari data untuk menentukan apakah kalimat tersebut negatif atau positif. Jika nilai Tahap *labelling* menghasilkan nilai negatif pada kolom *compound score* maka diberikan label negatif, begitupun sebaliknya akan diberi label positif jika nilai positif.

Dari proses *labelling* tersebut menghasilkan 2231 komentar negatif dan 1437 komentar positif, menunjukkan bahwa program penyebaran nyamuk *Wolbachia* di Indonesia menerima lebih banyak respon negatif. Pada Gambar 2 merupakan grafik yang menunjukkan hasil dari *labelling* ini.



Gambar 2. Hasil *Labelling*

Hasil dari *text preprocessing* ini memberikan data yang lebih siap untuk diolah lebih lanjut dengan metode klasifikasi NBC dan SVM. Namun sebelumnya harus dilakukan ekstraksi fitur terlebih dahulu yaitu menggunakan TF-IDF dan N-Gram. TF-IDF akan dijalankan ke dalam Python dengan memanfaatkan pustaka *TfidfVectorizer* dari *Scikit-Learn* untuk memberikan bobot yang lebih akurat pada kata-kata yang penting dalam analisis teks. Pada Tabel 6 merupakan hasil dari pembobotan kata menggunakan TF-IDF.

Tabel 6. Hasil TF-IDF

Kata	TF	IDF	TF-IDF
nyamuk	0.366685	2.002782	0.071597
dbd	0.112323	3.184227	0.021605
wolbachia	0.118593	3.130032	0.021442
tidak	0.19084	2.655167	0.03486
hero	0.026445	4.622707	0.016518
eksperimen	0.002181	7.01045	0.000779
eksptert	0.000273	8.514527	0.000116
....	....	....	....
Total	....	....	4668 Data

Pada Tabel 6 terdapat kata dengan nilai TF-IDF yang tinggi, kata tersebut berarti lebih relevan dalam teks dan berfungsi sebagai kata kunci penting dalam klasifikasi. Gambar 3 merupakan *wordcloud* yang menunjukkan frekuensi kemunculan sebuah kata dalam dokumen.



Gambar 3. Wordcloud TF-IDF

Wordcloud yang ditunjukkan pada Gambar 3 memvisualisasikan kata-kata dari hasil perhitungan TF-IDF, kata dengan ukuran besar merupakan kata yang memiliki frekuensi yang lebih tinggi kemunculannya pada suatu dokumen. Terlihat bahwa kata “nyamuk”, “tidak”, “dbd” dan “wolbachia” menunjukkan bahwa kata-kata tersebut sering muncul dan dianggap penting dalam dokumen untuk dianalisis.

Setelah melewati proses TF-IDF, selanjutnya melakukan proses N-Gram untuk menghasilkan data yang lebih akurat atau seimbang. Pada bagian ini akan dilakukan pemecahan kata pada kalimat dapat dilihat pada Tabel 7 menunjukkan beberapa contoh kalimat dari hasil proses N-Gram merujuk pada data yang sudah di labelling.

Tabel 7. Hasil N-Gram

N-gram	Kalimat
Unigram	'area', 'bandung', 'indonesia', 'information', 'is', 'large', 'of', 'prevent', 'rejects', 'singapore'
Bigram	'area of', 'bandung is', 'information singapore', 'is large', 'large area', 'of indonesia', 'singapore rejects'
Trigram	'area of indonesia', 'bandung is large', 'information singapore rejects', 'is large area', 'large area of'

Setelah itu dilakukan evaluasi pada fitur N-Gram untuk melihat kinerja dari fitur ini yang mempengaruhi hasil dari penelitian analisis sentimen. Tabel *confusion matrix* gabungan antara unigram, bigram dan trigram dapat dilihat pada Tabel 8.

Tabel 8. Confusion Matrix N-Gram

	Unigram		Bigram		Trigram	
	Pred_Positive	Pred_Negative	Pred_Positive	Pred_Negative	Pred_Positive	Pred_Negative
TP	228	60	221	67	210	78
TN	37	409	26	420	24	422

Berdasarkan Tabel 8 Unigram terlihat memiliki kinerja yang cukup seimbang dalam mendeteksi kelas positif dan negatif meskipun masih terdapat kesalahan dalam prediksi positif dengan hasil 60. Bigram jumlah kesalahannya dalam mendeteksi kelas negatif dengan nilai 26, lebih sedikit dibandingkan dengan Unigram.

Sedangkan Trigram paling baik dalam mendeteksi kelas negatif yaitu 422, namun memiliki performa yang paling buruk dalam mendeteksi kelas positif karena FN terlalu tinggi hasilnya yaitu 78.

### 3.3. Klasifikasi

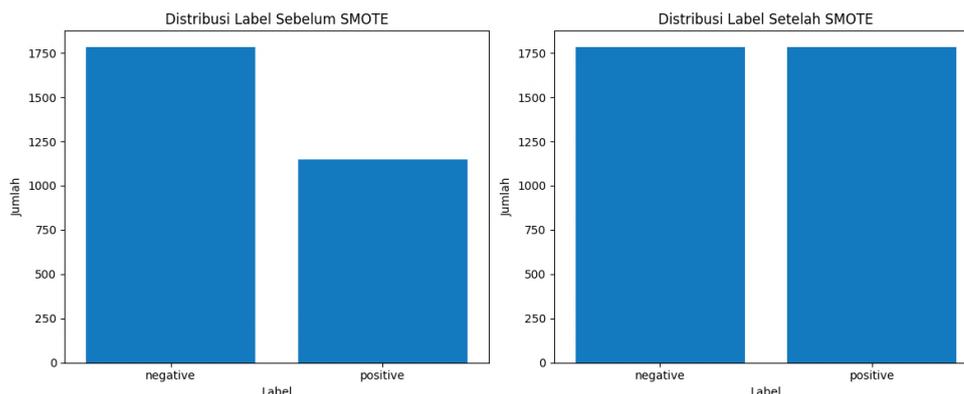
Pada tahap ini kumpulan data komentar akan dibagi menjadi dua bagian berupa data latih dan data uji. Proses klasifikasi dilakukan menggunakan algoritma NBC dan SVM. Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20, 70:30, dan 60:40 dapat dilihat pada Tabel 9.

Tabel 9. Perbandingan Data Latih dan Data Uji

Presentase		Jumlah Data	
<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
80%	20%	2895	724
70%	30%	2533	1086
60%	40%	2171	1448

Pada Tabel 9, dapat dilihat bahwa jumlah data berbeda pada setiap rasionya, seperti pada rasio 80:20 memiliki jumlah data latih yaitu 2895 sedangkan data uji 724.

Berdasarkan Gambar 2 mengenai hasil *labelling* yang terjadi ketidakseimbangan data, maka dari itu sebelum dilakukan klasifikasi menggunakan NBC dan SVM, ketidakseimbangan dalam data diatasi dengan menerapkan SMOTE untuk memastikan jumlah komentar positif dan negatif menjadi seimbang. Gambar 4 menunjukkan grafik sebelum dan sesudah dilakukan SMOTE.



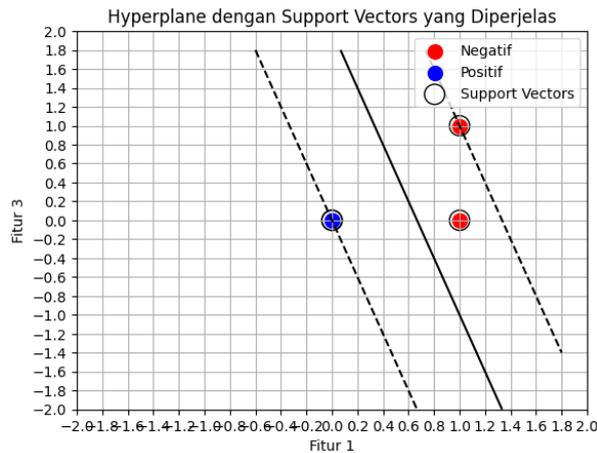
Gambar 4. Hasil SMOTE

Pada Gambar 4 menunjukkan bahwa grafik sebelah kiri terlihat bahwa dataset tidak seimbang dengan kelas negatif lebih tinggi dan mendominasi kelas positif. Sedangkan pada grafik di sebelah kanan menunjukkan bahwa kelas negatif dan positif telah dilakukan *re-sample* dengan membuat sampel sintetis yang membuat jumlahnya jadi sama dengan kelas negatif. Secara keseluruhan, SMOTE membantu memberikan hasil yang lebih seimbang sehingga tidak bias terhadap kelas mayoritas.

NBC dan SVM dilatih menggunakan data latih yang sudah dilakukan *labelling* sebelumnya untuk mempermudah tahap ini. NBC menghitung probabilitas kelas berdasarkan distribusi fitur dalam data latih. Kemudian dilakukan pengujian dengan data uji setelah NBC dilatih untuk melihat prediksi dari NBC. Pendekatan yang dilakukan oleh model NBC dengan menggunakan Persamaan (4) akan diterapkan untuk seluruh dataset dalam penelitian ini untuk memprediksi apakah sebuah kalimat memiliki sentimen positif atau negatif.

Setelah melakukan klasifikasi data menggunakan NBC, selanjutnya dilakukan klasifikasi menggunakan SVM. SVM menggunakan hyperplane optimal untuk memisahkan kelas positif dan negatif dengan margin terbesar. Pada Gambar 5 menunjukkan hyperplane SVM, di mana support vectors memainkan peran penting dalam menentukan margin.

Berdasarkan Gambar 5, terdapat titik-titik yang terlihat. Titik-titik biru dan merah adalah proyeksi dari data latih, di mana titik merah mewakili kelas negatif dan titik biru mewakili kelas positif. Garis hitam adalah *hyperplane* yang ditemukan oleh SVM yang merupakan pemisah dari kelas positif dan negatif. Titik yang memiliki lingkaran hitam di sekelilingnya dikenal sebagai *support vectors*, yakni titik data terdekat dengan *hyperplane* yang berperan penting dalam menentukan letak dan orientasi *hyperplane*.



Gambar 5. Contoh Hasil *Hyperplane*

**3.4. Evaluasi**

Evaluasi model penelitian merupakan tahap terakhir pada penelitian ini. Hasil evaluasi ini akan menggunakan *confusion matrix* yang berisi nilai *accuracy*, *precision*, *recall* dan *F1-score* dari data uji dengan menggunakan *K-Fold Cross Validation* dengan nilai k adalah 10 untuk memaksimalkan nilai.

*Confusion matrix* ditampilkan dalam bentuk tabel dengan setiap rasio pembagian data, hal tersebut dapat dilihat pada Tabel 10 untuk *confusion matrix* dari NBC.

Tabel 10. *Confusion Matrix* NBC

Rasio	Aktual	Prediksi Negatif	Prediksi Positif
80:20	Negatif	363	83
	Positif	48	240
70:30	Negatif	522	148
	Positif	66	365
60:40	Negatif	676	217
	Positif	74	501

Berdasarkan hasil dari *confusion matrix* pada Tabel 10 maka dapat ditarik kesimpulan bahwa model NBC secara konsisten memprediksi sentimen negatif dengan benar karena dapat dilihat tingginya nilai TN. Namun untuk nilai FP meningkat seiring dengan rasio yang semakin besar. Sebaliknya, untuk prediksi sentimen positif atau TP meningkat secara bertahap dari rasio 80:20 sampai rasio 60:40. Hal tersebut menunjukkan bahwa rasio yang lebih tinggi dapat membuat model lebih fokus untuk memprediksi sentimen positif.

Setelah didapatkan hasil dari *confusion matrix* pada tiap rasio, selanjutnya akan dilakukan evaluasi penghitungan metrik *accuracy*, *precision*, *recall* dan *F1-score*nya yang dapat dilihat pada Tabel 11.

Tabel 11. *Accuracy, precision, recall and f1-score* NBC

<i>Sentiment</i>	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-Score</i> (%)	Rasio
<i>positive</i>	82.15	74.30	83.33	78.56	80:20
<i>negative</i>	82.15	88.32	81.39	84.71	
<i>Average</i>	82.15	81.31	82.36	81.64	
<i>positive</i>	80.56	71.15	84.69	77.33	70:30
<i>negative</i>	80.56	88.78	77.91	82.99	
<i>Average</i>	80.56	79.96	81.30	80.16	
<i>positive</i>	80.18	69.78	87.13	77.49	60:40
<i>negative</i>	80.18	90.13	75.70	82.29	
<i>Average</i>	80.18	79.96	81.42	79.89	

Dari hasil evaluasi yang ditampilkan pada Tabel 11, terlihat bahwa pembagian data 80:20 memiliki *accuracy* terbaik yaitu 82.15% dan keseimbangan antara metrik *precision*, *recall* dan *F1-score* untuk kelas positif dan negatif.

Meskipun metrik precision pada kelas positif lebih rendah pada rasio 80:20 ini, namun model NBC cukup bekerja dengan konsisten pada kedua kelas tersebut dengan hasil F1-score yang tinggi antara positif (77.27%) dan negatif (82.96%). Selanjutnya adalah hasil dari *K-Fold Cross Validation* NBC yang ditampilkan pada Tabel 12.

Tabel 12. *K-Fold Cross Validation* NBC

Fold	Positive				Negative			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	79.29	69.82	82.52	75.64	79.29	69.82	82.52	75.64
2	81.74	73.91	82.64	78.03	81.74	73.91	82.64	78.03
3	80.93	74.34	78.47	76.35	80.93	74.34	78.47	76.35
4	83.38	73.71	89.58	80.88	83.38	73.71	89.58	80.88
5	85.56	77.91	88.19	82.74	85.56	77.91	88.19	82.74
6	86.65	79.50	88.89	83.93	86.65	79.50	88.89	83.93
7	84.47	76.36	87.50	81.55	84.47	76.36	87.50	81.55
8	80.65	72.12	82.64	77.02	80.65	72.12	82.64	77.02
9	84.70	76.36	88.11	81.82	84.70	76.36	88.11	81.82
10	81.69	72.62	85.31	78.46	81.69	72.62	85.31	78.46

Berdasarkan Tabel 12, metrik evaluasi yang digunakan yaitu *accuracy*, *precision*, *recall* dan juga *F1-score* pada sentimen positif dan negatif. Pada fold ke-6 memberikan menunjukkan performa yang lebih baik dibandingkan fold yang lainnya pada semua metrik. Secara keseluruhan, NBC menunjukkan kinerja yang cukup baik dan konsisten dengan akurasi rata-rata yang cukup tinggi terhadap kelas positif maupun negatif.

Setelah dilakukan evaluasi pada NBC, selanjutnya melakukan evaluasi terhadap SVM. Hasil *confusion matrix* dari SVM ditampilkan pada Tabel 13.

Tabel 13 *Confusion Matrix* SVM

Rasio	Aktual	Prediksi Negatif	Prediksi Positif
80:20	Negatif	423	23
	Positif	100	188
70:30	Negatif	629	41
	Positif	162	269
60:40	Negatif	833	60
	Positif	231	344

Pada Tabel 13, seiring dengan meningkatnya ukuran data latih dari rasio 80:20 sampai 60:40, nilai TN semakin naik sama halnya dengan *True Positive (TP)*. Sedangkan untuk *False Negative (FN)* menunjukkan bahwa dengan data latih yang lebih sedikit, maka lebih banyak contoh positif yang salah prediksi sebagai negatif. Namun secara keseluruhan, dapat dilihat bahwa model semakin baik seiring dengan meningkatnya data yang digunakan dan *confusion matrix* rasio 60:40 memiliki kinerja yang lebih baik dibandingkan dengan rasio lainnya. Hasil evaluasi dari tiap rasio menggunakan SVM data ditampilkan pada Tabel 14.

Tabel 14. *Accuracy, precision, recall and f1-score* SVM

Sentiment	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Rasio
positive	83.26	89.10	65.28	75.29	80:20
negative	83.26	80.84	94.85	87.25	
Average	83.26	84.92	80.07	81.27	
positive	81.56	86.77	62.41	72.60	70:30
negative	81.56	79.52	93.88	86.11	
Average	81.56	83.15	78.15	79.35	
positive	80.18	85.15	59.83	70.28	60:40
negative	80.18	78.29	93.28	85.13	
Average	80.18	81.72	76.55	77.70	

Pada Tabel 14 terlihat bahwa pembagian data 80:20 memiliki rata-rata akurasi tertinggi yaitu 83.26% dibandingkan dengan rasio lainnya. Meskipun rasio *recall* memiliki nilai yang sangat rendah pada kelas positif yaitu 65.28% dibandingkan dengan negatif yaitu 94,85% namun nilai pada *F1-score* lebih baik secara keseluruhan dibandingkan dengan rasio lainnya, hal ini menunjukkan keseimbangan yang baik pada rasio 80:20. Selanjutnya adalah hasil dari *K-Fold Cross Validation SVM* yang ditampilkan pada Tabel 15.

Tabel 15. *K-Fold Cross Validation SVM*

Fold	Positive				Negative			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	39.46	86.96	68.97	76.92	60.54	82.18	93.26	87.37
2	39.46	93.10	69.83	79.80	60.54	83.09	96.63	89.35
3	39.12	87.80	62.61	73.10	60.88	79.72	94.41	86.45
4	39.12	89.33	58.26	70.53	60.88	78.08	95.53	85.93
5	39.25	80.72	58.26	67.68	60.75	77.14	91.01	83.51
6	39.25	83.13	60.00	69.70	60.75	78.10	92.13	84.54
7	39.25	87.65	61.74	72.45	60.75	79.25	94.38	86.15
8	39.25	88.04	70.43	78.26	60.75	83.08	93.82	88.13
9	39.25	79.78	61.74	69.61	60.75	78.43	89.89	83.77
10	39.25	87.10	70.43	77.88	60.75	83.00	93.26	87.83

Berdasarkan Tabel 15 hasil dari *K-Fold cross validation* ini menunjukkan bahwa *fold* ke-2 adalah yang terbaik secara keseluruhan. Pada sentimen positif *fold* ke-2 memiliki nilai *precision* yaitu 93.10%, *recall* yaitu 69.83% dan *F1-score* yaitu 79.80%. Sedangkan pada *fold* ke-2 sentimen negatif, nilai *precision* yaitu 83.09%, *recall* 96.63% dan *F1-score* 89.35%. Model SVM secara keseluruhan memberikan kinerja yang cukup baik pada kelas negatif dibandingkan kelas positif. Perbandingan dari hasil perhitungan metrik berdasarkan *confusion matrix* dari NBC maupun SVM dapat dilihat pada Tabel 16.

Tabel 16. Perhitungan Metrik dari *Confusion Matrix* NBC dan SVM

Metric	SVM (%)	NBC (%)
Accuracy	83,26	82,16
Precision	84,99	81,31
Recall	80,06	82,36
F1-Score	81,33	81,64

Berdasarkan Tabel 16, SVM merupakan model yang lebih cocok jika ingin meminimalkan kesalahan dalam prediksi hasil positif (*higher precision*). Dalam artian jika ingin menghindari FP, maka SVM adalah pilihan yang tepat. Sedangkan NBC lebih cocok digunakan ketika tujuannya untuk menangkap sebanyak mungkin hasil yang benar (*higher recall*). NBC lebih cocok untuk menemukan semua hasil positif yang relevan.

Pada hasil metrik *F1-score* menunjukkan hasil yang tidak begitu jauh perbedaannya, SVM 81.33% sedangkan NBC 81.64% sedikit lebih unggul. Jika kondisinya menginginkan model yang seimbang, maka NBC merupakan model yang cocok untuk menangkap hasil yang benar dan menghindari kesalahan. Setelah melakukan klasifikasi data Setelah data dikumpulkan, langkah selanjutnya adalah melakukan *text preprocessing*.

#### 4. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma NBC dan SVM memiliki keunggulan spesifik dalam menganalisis sentimen publik terhadap program pelepasan nyamuk *Wolbachia* di Indonesia. Data yang dikumpulkan dari komentar Instagram diproses melalui *text preprocessing*, yang mencakup TF-IDF dan N-Gram untuk ekstraksi fitur, serta divalidasi menggunakan *k-Fold Cross Validation*. Proses *text preprocessing* dan validasi ini memastikan data yang lebih bersih, representasi fitur yang akurat, dan evaluasi model yang stabil.

Hasil penelitian menunjukkan bahwa SVM unggul dengan akurasi tertinggi sebesar 83,24% pada rasio data latih dan uji 80:20, serta *recall* sebesar 85,39% untuk sentimen positif. Keunggulan ini didukung oleh kemampuan SVM dalam memanfaatkan margin optimal yang memaksimalkan generalisasi model pada data baru. Dengan demikian, SVM lebih efektif dalam mengukur dukungan publik terhadap program kesehatan inovatif. Di sisi lain, NBC menunjukkan keunggulan dalam *precision* untuk sentimen negatif sebesar 86,36%, menjadikannya lebih cocok untuk mendeteksi kritik atau kekhawatiran publik. Efektivitas NBC ini didukung

oleh pendekatannya yang berbasis probabilitas, terutama pada data yang telah diseimbangkan menggunakan SMOTE.

Hasil penelitian ini memiliki implikasi praktis yang signifikan. Pemerintah dan pemangku kepentingan dapat menggunakan SVM untuk memantau tingkat dukungan masyarakat terhadap program kesehatan, sehingga dapat menentukan strategi komunikasi yang lebih tepat sasaran. NBC dapat digunakan untuk mendeteksi kritik publik secara lebih cepat, memungkinkan pengambilan tindakan yang responsif terhadap kekhawatiran masyarakat. Dengan demikian, kedua algoritma ini memberikan pendekatan yang saling melengkapi dalam memahami sentimen masyarakat secara komprehensif.

Namun, penelitian ini memiliki keterbatasan, yaitu penggunaan data yang hanya berasal dari platform Instagram. Data dari media sosial lain, seperti TikTok dan Facebook, yang memiliki karakteristik interaksi yang berbeda, dapat dieksplorasi untuk memberikan perspektif yang lebih luas. Penelitian selanjutnya juga disarankan untuk mengeksplorasi algoritma modern, seperti *Random Forest* atau *Neural Network*, yang dapat memberikan tingkat akurasi lebih tinggi dan pemahaman yang lebih mendalam tentang persepsi masyarakat. Selain itu, pendekatan multimodal yang menggabungkan analisis teks dengan data visual atau video dapat memperkaya wawasan mengenai opini publik terhadap program kesehatan.

Dengan hasil dan rekomendasi yang disampaikan, penelitian ini diharapkan dapat mendukung pengambilan keputusan berbasis data yang lebih strategis dan meningkatkan efektivitas implementasi program kesehatan di Indonesia. Penelitian ini juga memberikan landasan untuk pengembangan analisis sentimen berbasis media sosial sebagai alat untuk memahami dan merespons opini publik secara lebih efisien.

#### DAFTAR PUSTAKA

- [1] Kemenkes RI, "Membuka Lembaran Baru," *Lap. Tah. 2022 Demam Berdarah Dengue*, pp. 17–19, 2022.
- [2] Kemenkes RI, "Inovasi Wolbachia, Cara Ampuh dan Hemat Kendalikan Demam Berdarah," *Kementeri. Kesehat. RI*, pp. 1–2, 2023, [Online]. Available: <https://www.kemkes.go.id/id/rilis-kesehatan/inovasi-wolbachia-cara-ampuh-dan-hemat-kendalikan-demam-berdarah>
- [3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," *Proc. 2002 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2002*, pp. 79–86, 2002.
- [4] A. Irwanto and L. Goeirmanto, "Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia using Naïve Bayes and XGboost Classifier Algorithm," vol. 27, no. 2, pp. 145–152, 2023.
- [5] Ermatita, A. Sanmorino, Samsuryadi, and D. P. Rini, "Analyzing Factors Contributing to Research Performance using Backpropagation Neural Network and Support Vector Machine," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 1, pp. 153–172, 2022, doi: 10.3837/tiis.2022.01.009.
- [6] A. Irwanto and L. Goeirmanto, "Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia using Naive Bayes and XGboost Classifier Algorithm," pp. 145–152, 2023, doi: <http://doi.org/10.22441/sinergi.2023.2.001>.
- [7] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, no. January, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.
- [8] R. Kusumawati, A. Arofah, D, and A. Pramana, P, "Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter 's Classification of Tokopedia Services," *J. Phys. Conf. Ser.*, vol. 1320, pp. 1–11, 2019, doi: 10.1088/1742-6596/1320/1/012016.
- [9] E. P. V. Reddy and S. Ramesh, "Comparative analysis of support vector machine and Naive Bayes classifier for the prediction of airline passenger satisfaction," *AIP Conf. Proc.*, 2024.
- [10] E. K. Seltzer, M. Lu, and R. M. Merchant, "Public sentiment and discourse about Zika virus on Instagram," vol. 0, no. 215, pp. 0–5, 2017.
- [11] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, pp. 509–553, 2023, doi: 10.1017/S1351324922000213.
- [12] A. Mardjo and C. Choksuchat, "HyVADRF: Hybrid VADER-Random Forest and GWO for Bitcoin Tweet Sentiment Analysis," *IEEE Access*, vol. 10, no. August, pp. 101889–101897, 2022, doi: 10.1109/ACCESS.2022.3209662.
- [13] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Syst. Appl.*, vol. 57, pp. 117–126, 2016, doi: 10.1016/j.eswa.2016.03.028.
- [14] S. Aathira, B. Reena, and D. S. Uma, "Mental Stress Detection Using TF-IDF with Multinomial Naive

- 
- Bayes,” *Int. J. Res. Sci. Eng. Technol.*, vol. 9, no. 6, pp. 1–5, 2023.
- [15] F. D. Astuti and F. N. Lenti, “Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution Menggunakan KNN,” *JUPITER*, vol. 13, pp. 89–98, 2021.
- [16] S. Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, “Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm,” *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 1–9, 2023, doi: 10.57152/precedecs.v1i1.814.
- [17] S. Hilda Kusumahadi, H. Junaedi, and J. Santoso, “Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 54–60, 2019, doi: 10.30591/jpit.v4i1.1125.
- [18] L. F. Hu, W. Gong, L. X. Qi, and P. Wang, “A Method for Feature Selection Based On the Optimal Hyperplane of SVM and Independent Analysis,” pp. 14–17, 2013, doi: 10.1109/ICMLC.2013.6890454.
- [19] 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, 2021, doi: 10.34148/teknika.v10i1.311.
- [20] A. Saepudin, A. Faqih, and G. Dwilestari, “Perbandingan Algoritma Klasifikasi Support Vector Machine , Random Forest dan Logistic Regression Pada Ulasan Shopee,” vol. 18, no. 1, pp. 178–192, 2024.
- [21] A. Muhaddisi, B. N. Prastowo, and D. U. Kusumaning Putri, “Sentiment Analysis With Sarcasm Detection On Politician’s Instagram,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 4, p. 349, 2021, doi: 10.22146/ijccs.66375.