

Analisis Sentiment Terhadap Diabetes Menggunakan Algoritma Naïve Bayes, Random Forest, SVM Pada Media Sosial X

LindaApriani, Nirwana Hendrastuty*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹lindaa@teknokrat.ac.id, ^{2,*}nirwanahendrastuty@teknokrat.ac.id

Email Penulis Korespondensi: nirwanahendrastuty@teknokrat.ac.id

Submitted: 06/02/2025; Accepted: 05/03/2025; Published: 07/03/2025

Abstrak—*Diabetes* merupakan salah satu penyakit kronis yang mendapat perhatian luas di masyarakat, khususnya di media sosial X. Hal ini disebabkan oleh meningkatnya jumlah penderita setiap tahunnya. Berdasarkan data Organisasi Kesehatan Dunia (WHO), pada tahun 2021 diperkirakan sebanyak 537 juta orang berusia 20–79 tahun hidup dengan diabetes, meningkat dari estimasi tahun 2019 yang berjumlah 463 juta orang. Selain itu, sekitar 1,3 juta kematian disebabkan oleh diabetes, dengan 4 persen di antaranya terjadi sebelum usia 70 tahun. Kondisi ini terjadi akibat tingginya kadar gula darah yang mengganggu fungsi metabolisme tubuh, sehingga tubuh mengalami kesulitan dalam mengolah gula secara optimal. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM) dalam analisis sentimen terkait diabetes. Data penelitian diperoleh dari platform Twitter dengan total 8.401 tweet yang dikumpulkan menggunakan teknik *crawling* menggunakan *keyword* tertentu dalam rentang waktu 2024 hingga 2025. Data tersebut kemudian melalui tahap pra-pemrosesan untuk menghasilkan data yang bersih. Pengujian dilakukan guna mengevaluasi tingkat akurasi masing-masing model dalam memprediksi sentimen publik. Hasil pengujian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi sebesar 85%, diikuti oleh *Random Forest* dengan akurasi 82%, dan *Naïve Bayes* dengan akurasi 74% sebelum penerapan *Synthetic Minority Oversampling Technique* (SMOTE). Setelah optimasi menggunakan SMOTE, algoritma SVM tetap menunjukkan performa terbaik dengan akurasi 96%, diikuti oleh *Random Forest* dengan akurasi 95%, dan *Naïve Bayes* dengan akurasi 85%. Berdasarkan hasil tersebut, SVM terbukti menjadi algoritma paling efektif dalam mengklasifikasikan sentimen terkait diabetes. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam upaya penanganan diabetes melalui pemahaman yang lebih baik terhadap persepsi masyarakat.

Kata Kunci: Analisis Sentimen; Diabetes; Naïve Bayes; Random Forest; SVM; SMOTE.

Abstract—*Diabetes* is one of the chronic diseases that has received widespread attention in society, especially on social media X. This is due to the increasing number of sufferers every year. Based on data from the World Health Organization (WHO), in 2021 it is estimated that 537 million people aged 20–79 years are living with diabetes, an increase from the 2019 estimate of 463 million people. In addition, around 1.3 million deaths are caused by diabetes, with 4 percent of them occurring before the age of 70. This condition occurs due to high blood sugar levels that interfere with the body's metabolic functions, making it difficult for the body to process sugar optimally. This study aims to compare the performance of *Naïve Bayes*, *Random Forest*, and *Support Vector Machine* (SVM) algorithms in sentiment analysis related to diabetes. The research data was obtained from the Twitter platform with a total of 8,401 tweets collected using crawling techniques using certain keywords in the time span of 2024 to 2025. The data then went through a pre-processing stage to produce clean data. Tests were conducted to evaluate the accuracy of each model in predicting public sentiment. The test results show that the SVM algorithm provides the best performance with 85% accuracy, followed by *Random Forest* with 82% accuracy, and *Naïve Bayes* with 74% accuracy before the application of *Synthetic Minority Oversampling Technique* (SMOTE). After optimization using SMOTE, the SVM algorithm still showed the best performance with 96% accuracy, followed by *Random Forest* with 95% accuracy, and *Naïve Bayes* with 85% accuracy. Based on these results, SVM proved to be the most effective algorithm in classifying sentiment related to diabetes. It is hoped that the results of this research can contribute to efforts to manage diabetes through a better understanding of public perceptions.

Keywords: Sentiment Analysis; Diabetes; Naïve Bayes; Random Forest; SVM; SMOTE.

1. PENDAHULUAN

Diabetes adalah salah satu penyakit kronis yang mendapat perhatian besar secara global. Kondisi ini terjadi akibat kadar gula darah yang tinggi, yang mengganggu fungsi metabolisme tubuh sehingga tubuh kesulitan memproses gula dengan optimal. Penyakit ini tergolong berbahaya karena dapat menyebabkan dampak serius, terutama jika sudah memicu berbagai komplikasi [1]. *Diabetes* dapat menyebabkan berbagai komplikasi serius, seperti amputasi kaki, kebutaan, kerusakan saraf, gagal ginjal, stroke, serangan jantung, bahkan kematian. Mengingat tingginya risiko komplikasi yang dapat terjadi, diperlukan upaya pencegahan untuk mengendalikan penyakit ini. Selain memengaruhi kesehatan individu, diabetes juga menimbulkan beban ekonomi yang besar bagi keluarga serta sistem layanan kesehatan masyarakat [2].

Berdasarkan data dari Organisasi Kesehatan Dunia (WHO), pada tahun 2021 diperkirakan sebanyak 537 juta orang berusia 20–79 tahun hidup dengan *diabetes*. Angka ini meningkat dibandingkan estimasi tahun 2019, yaitu 463 juta orang. Selain itu, terdapat sekitar 1,3 juta kematian akibat *diabetes*, di mana 4 persen di antaranya terjadi sebelum usia 70 tahun. Sebagian besar kematian akibat *diabetes* ditemukan pada kelompok usia 45–54 tahun, terutama di daerah perkotaan dibandingkan pedesaan [3]. Masalah *diabetes* ini tidak hanya terbatas di Indonesia, tetapi prevalensinya sangat bervariasi di seluruh dunia. WHO melaporkan bahwa wilayah dengan prevalensi *diabetes* tertinggi adalah Timur Tengah dan Afrika Utara, dengan sekitar 10,3% populasi yang terkena dampak. Wilayah lain dengan prevalensi tinggi termasuk Amerika Tengah dan Selatan (9,2%), Asia Tenggara (8,5%), dan Pasifik Barat (8,2%). Di Eropa,

prevalensi *diabetes* diperkirakan sebesar 6,6%, sementara di Amerika Utara mencapai 6,2%. Prevalensi terendah tercatat di Afrika, yaitu 4,6% [4].

Dengan meningkatnya jumlah kasus *diabetes*, media sosial seperti Media X telah berkembang menjadi salah satu platform utama untuk menggali pandangan masyarakat mengenai penyakit ini. Media sosial X menyediakan akses ke berbagai opini, emosi, dan pengalaman yang dibagikan oleh pengguna, termasuk mengenai pengobatan, langkah pencegahan, serta dampaknya dalam kehidupan sehari-hari. Keragaman data yang terdapat di Media sosial X menjadikannya sumber informasi yang sangat berharga untuk melakukan analisis sentimen [5].

Analisis sentimen adalah metode yang digunakan untuk memahami emosi, pendapat, atau sikap yang terkandung dalam teks, baik yang bernuansa positif, negatif, maupun netral. Dalam penelitian ini, analisis sentimen diterapkan untuk mengungkap faktor-faktor yang memengaruhi persepsi publik terhadap *diabetes*. Pendekatan ini sangat relevan, karena opini masyarakat tidak hanya terbatas pada polaritas positif dan negatif, tetapi juga dapat mencerminkan kombinasi polaritas yang lebih kompleks [6].

Beberapa penelitian sebelumnya terkait masalah dan solusi serupa yang telah dilakukan. Salah satunya adalah studi yang menganalisis sentiment penyakit *diabetes* di media sosial twitter menggunakan algoritma *Naïve Bayes* hasil penelitian menunjukkan bahwa sentiment positif mendapatkan hasil 22% sedangkan sentiment negative 14% dan sentiment netral 64% serta nilai akurasi yaitu 87% [7]. Penelitian sebelumnya menganalisis sentiment penyakit hepatitis akut menggunakan algoritma *Support Vector Machine*, data yang di gunakan merupakan data tweet yang di dapat dari media twitter yang mendapatkan hasil sentiment positif yaitu 1.590 dan sentiment negatif yaitu 468 serta mendapatkan hasil akurasi yaitu 81,06%, *precession* 81,88% dan *recall* 97,19% penelitian ini juga melakukan perbandingan 80:20 [8]. Penelitian lain melakukan perbandingan algoritma *Naïve Bayes* dan *Random forest* pada analisis prediksi kanker paru mendapatkan hasil bahwa algoritma *Random forest* lebih unggul di dibandingkan dengan *Naïve Bayes* dengan hasil akurasi 98,4% dan mendapatkan nilai *recall* 100% untuk kelas positif dan 80% untuk kelas negatif serta mendapatkan nilai prediksi 100% dilihat dari nilai AUC yaitu 1 [9].

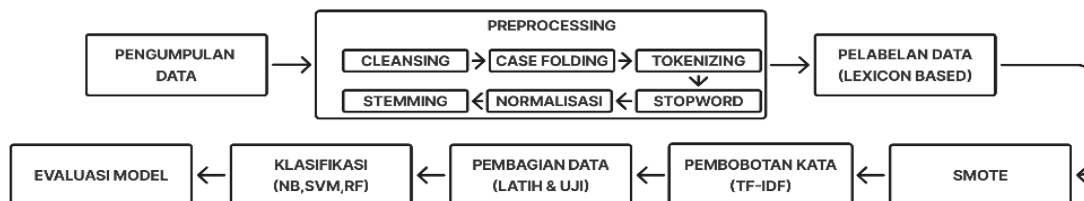
Penelitian serupa juga membahas perbandingan algoritma *Naïve Bayes*, *Random Forest* dan SVM dalam analisis sentiment kesehatan mental. Hasilnya model SVM mendapatkan nilai akurasi paling tinggi yaitu sebesar 86,11%, *Random Forest* mendapatkan nilai akurasi 82,55% dan *Naïve Bayes* mendapatkan nilai akurasi sebesar 78,19% [10]. Selanjutnya penelitian lain juga membahas tentang analisis sentiment terhadap vaksinasi penyakit covid 19 pada twitter menggunakan algoritma *Random Forest*, penelitian ini divalidasi menggunakan *K-fold cross validation* dan *Confusion Matrix* yang mendapatkan hasil sentiment positif oleh masyarakat dengan nilai akurasi mencapai 79% nilai *Precession* sebesar 85% *Recall* sebesar 90% dan F1 Score sebesar 88% [11].

Berdasarkan masalah yang telah diuraikan sebelumnya serta penelitian penelitian terkait. Penelitian ini bertujuan untuk menganalisis sentiment serta membandingkan tiga algoritma utama, yaitu *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine (SVM)*. Ketiga algoritma ini dipilih karena keandalannya dalam mengolah data teks dan menghasilkan analisis yang akurat. Proses analisis mencakup beberapa langkah penting, seperti tokenisasi, stemming, dan penghapusan kata-kata yang tidak relevan, untuk memastikan kualitas data yang digunakan. Dengan pendekatan ini, diharapkan dapat ditemukan metode yang paling efektif dalam menggali opini masyarakat mengenai penyakit *diabetes* berdasarkan data yang diambil dari Media sosial X

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Langkah-langkah penelitian dalam studi ini disusun untuk memberikan pemahaman yang komprehensif terkait analisis sentimen terhadap penyakit *diabetes* berdasarkan data dari media sosial X atau Twitter. Penelitian ini membandingkan kinerja algoritma *Random Forest*, SVM, dan *Naive Bayes*. Gambar 1 memperlihatkan tahapan-tahapan penelitian yang mencakup proses pengumpulan data (*crawling*), *praprosesing* data, pelabelan, penerapan SMOTE, ekstraksi fitur, pembagian data, pemodelan, evaluasi, dan visualisasi. Penjelasan lebih rinci mengenai tahapan metode penelitian ini dapat ditemukan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.3 Pengumpulan Data

Data Pada penelitian ini di dapat dari proses *crawling* menggunakan *library harves* pada media social X atau twitter, Proses pengumpulan data dilakukan melalui langkah code program yang melibatkan penggunaan kata kunci untuk mencari tweet opini publik tentang *diabetes* yang sesuai dengan preferensi yang diinginkan [12].

2.4 Pre-processing

Pada tahap ini dilakukan proses pembersihan data mentah menjadi data yang siap di olah untuk ke tahap tahap berikutnya, data akan melewati beberapa proses yaitu proses *cleaning*, *casefolding*, normalisasi kata, *Tokenizing* *stopword removal* dan *stemming*.

2.4.1 Cleaning

Tahap *cleaning*, yang juga sering disebut sebagai tahap pembersihan data, merupakan langkah penting dalam analisis sentimen. Data yang berasal dari media social X sering kali mengandung noise seperti karakter khusus, tautan, atau teks yang kurang relevan[13]

2.4.2 Case Folding

Case Folding merupakan tahapan mengganti seluruh case dalam sebuah dokumen menjadi bentuk standar (huruf kecil). Sedangkan karakter lainnya dianggap sebagai delimiter atau pembatas[14] proses ini dilakukan untuk mempercepat perbandingan selama proses pengindeksan.

2.4.3 Tokenizing

Merupakan suatu tahapan untuk memecah suatu kumpulan teks menjadi sebuah kata. Analisis kata-kata dalam teks menjadi penting karena dengan melakukan tokenisasi, makna dari teks dapat dengan mudah ditentukan. Dalam pemrosesan bahasa alami, *tokenizing* sering digunakan untuk memecah teks menjadi unit-unit yang lebih kecil dan mudah diolah oleh mesin. Beberapa algoritma tokenizing yang umum digunakan adalah pemotongan karakter, pemotongan kata, dan pemotongan frase[15]

2.4.4 Normalisasi Kata

Normalisasi kata merupakan proses mengubah kata-kata yang berbeda bentuk menjadi satu bentuk yang seragam. Normalisasi dapat membantu meningkatkan akurasi dalam analisis teks dan membuat data lebih mudah diolah[15]

2.4.5 StopWords

Stopword Removal digunakan untuk Menghilangkan kata-kata yang umumnya tidak memberikan informasi penting dalam pemrosesan teks, seperti "dan", "atau", "yang", dan lain-lain. Proses *stopword removal* biasanya dilakukan setelah tokenisasi, di mana teks telah dibagi menjadi kata-kata individual. Setelah itu, setiap kata akan diperiksa dan kata-kata yang terdaftar dalam daftar *stopWords* akan dihapus[16].

2.4.6 Stemming

Proses *stemming* merupakan proses dimana dilakukannya pemetaan dan penguraian berbagai bentuk (variants) menjadi bentuk stem (kata dasar) dari kata hasil pemrosesan *stopword removal* atau *filtering* Proses *stemming* digunakan untuk meningkatkan kualitas informasi dari data yang didapatkan. contoh kata “mengerjakan”, “dikerjakan” dan “pekerja” yang semula mengandung arti yang berbeda dapat dilakukan stem menjadi sebuah kata “kerja” yang memiliki arti yang sama sehingga kata tersebut berhubungan[17].

2.5 Pelabelan Data

Tahap pelabelan menggunakan metode berbasis leksikon dilakukan dengan memanfaatkan kamus yang berisi kata-kata positif dan negatif. Proses ini didasarkan pada skor sentimen: jika skor lebih dari nol, maka sentimen dikategorikan sebagai positif; jika skor kurang nol, sentimen dikategorikan sebagai negatif. Sementara itu, jika skor tidak memenuhi kedua kriteria tersebut, sentimen akan dimasukkan ke dalam kategori netral[18].

2.6 SMOTE

Synthetic Minority Oversampling Technique (SMOTE) merupakan suatu metode yang dirancang untuk mengatasi masalah ketidakseimbangan kelas pada dataset. Ketidakseimbangan ini terjadi ketika jumlah data pada satu kelas jauh lebih sedikit dibandingkan kelas lainnya, yang dapat mengakibatkan model pembelajaran mesin menjadi bias. SMOTE berfungsi dengan menciptakan sampel sintesis untuk kelas minoritas melalui proses interpolasi di antara data minoritas yang sudah ada. Pendekatan ini meningkatkan proporsi kelas minoritas dalam dataset, sehingga model pembelajaran mesin dapat lebih memahami pola dari kelas tersebut dan menghasilkan prediksi yang lebih akurat[19].

2.7 Ekstraksi Fitur

Ekstraksi Fitur merupakan sebuah proses merubah sebuah data yang di proyeksikan ke dalam fitur baru dengan dimensi yang lebih rendah[20] Metode yang digunakan dalam ekstraksi fitur ini adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF merupakan teknik yang sering diterapkan dalam pemrosesan bahasa alami dan pencarian informasi untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen dibandingkan dengan keseluruhan dokumen dalam korpus. Berikut adalah cara perhitungan pembobotan menggunakan metode TF-IDF. dapat dilihat pada persamaan 1.

$$TF * IDF(d, t) = TF(d, t) * \log \frac{N}{df(t)} \quad (1)$$

Dimana:

Nilai Term Frequency (TF) menunjukkan seberapa sering suatu term muncul dalam sebuah dokumen, sementara Inverse Document Frequency $IDF(d, t)$ mengukur seberapa penting term tersebut dalam seluruh kumpulan dokumen. Dalam hal ini, (d, t) merepresentasikan jumlah kemunculan term t dalam dokumen, d , N adalah total jumlah dokumen dalam koleksi, dan $df(t)$ menunjukkan jumlah dokumen yang mengandung term t .

2.8 Pembagian Data

Dalam penelitian ini, pembagian data dilakukan dengan rasio 70:30, di mana 70% data digunakan sebagai data training, sedangkan 30% sisanya digunakan sebagai data testing. Pembagian data ini menggunakan teknik stratified sampling untuk memastikan proporsi kelas sentimen (negatif, netral, positif) tetap konsisten di data pelatihan dan pengujian. Data latih berfungsi untuk membangun model klasifikasi dengan menggunakan data yang label kelasnya sudah diketahui, yaitu positif, negatif, dan netral. Sementara itu, data uji digunakan untuk mengevaluasi performa model yang telah dibuat[21].

2.9 Pembangunan Model

Pada tahap ini, klasifikasi model dilakukan dengan menggunakan tiga algoritma, yaitu *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine*. Proses pemodelan ini memanfaatkan bahasa pemrograman *Python* dengan *library sklearn*. Berikut ini adalah penjelasan lebih rinci mengenai ketiga algoritma tersebut:

2.9.1 Naïve Bayes

Naïve Bayes adalah metode klasifikasi sederhana yang menghitung peluang berdasarkan kombinasi data yang ada[22]. Teorema Bayes itu sendiri adalah rumus yang digunakan untuk menghitung probabilitas suatu kejadian berdasarkan pengetahuan sebelumnya. Dalam konteks klasifikasi, *Naïve Bayes* digunakan untuk memprediksi kelas atau kategori dari data berdasarkan fitur yang ada[23]. Persamaannya dapat ditunjukkan pada persamaan 2 berikut:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (2)$$

Probabilitas posterior dari kelas C berdasarkan fitur X dinyatakan sebagai $P(C|X)$, sementara $P(X|C)$ merepresentasikan probabilitas munculnya fitur X jika diketahui bahwa data berasal dari kelas C . Selain itu, $P(C)$ menggambarkan probabilitas umum dari kelas C , sedangkan $P(X)$ menunjukkan probabilitas keseluruhan dari fitur X .

2.9.2 Random Forest

Random Forest merupakan algoritma *ensemble* yang diterapkan untuk tugas klasifikasi dan regresi dengan menggabungkan sejumlah pohon keputusan. Pendekatan ini bertujuan untuk meningkatkan akurasi prediksi secara keseluruhan. Algoritma ini sangat baik dalam mengatasi masalah *overfitting* yang sering terjadi pada pohon keputusan tunggal dan mampu menangani variabel-variabel yang kurang relevan pada dataset berskala besar[24]. persamaan untuk *Random forest* dapat dilihat pada persamaan 3.

$$f(x) = \text{Average}f_1(x), f_2(x), \dots, f_n(x) \quad (3)$$

Fungsi $f(x)$ merepresentasikan hasil prediksi, sementara $f_{1-n}(x)$ adalah hasil prediksi yang diperoleh dari setiap pohon keputusan ke- n . Adapun (x) merupakan data yang digunakan sebagai input dalam proses prediksi.

2.9.3 Support Machine Vektor (SVM)

SVM merupakan algoritma klasifikasi berbasis diskriminatif yang bertujuan menemukan batas pemisah terbaik, yang dikenal sebagai *hyperplane*, untuk membedakan kelas[10]. Pada kernel linear, *hyperplane* optimal direpresentasikan oleh persamaan 4 dan 5.

$$(w \cdot x_i + b \leq 1, y_i = -1) \quad (4)$$

$$(w \cdot x_i + b \leq 1, y_i = 1) \quad (5)$$

Data ke- i dilambangkan sebagai x_i , dengan $w \cdot x_i$ merepresentasikan nilai bobot untuk kelas data tersebut. Sementara itu, b menunjukkan nilai bias, dan y_i merupakan kelas dari data ke- i .

2.10 Evaluasi Model

Peguian model untuk mengukur kinerja metode *Naïve Bayes*, *Random Forest*, dan *Support Vektor Machine* dilakukan untuk sejauh mana model tersebut mampu memprediksi data dengan akurat[25]. Evaluasi kinerja model memanfaatkan *Confusion Matrix* untuk menampilkan hasil prediksi model yang sedang diuji, serta menghitung metrik evaluasi seperti presisi, recall, skor F1, dan akurasi pada setiap datatest[26], persamaan *confusion matrix* dapat dilihat pada persamaan berikut

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{6}$$

$$Precision = \frac{TP}{TF+FP} \tag{7}$$

$$Recall = \frac{TF}{TF+FN} \tag{8}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan metode *crawling* menggunakan *google collab* dengan bahasa pemrograman *python*, menggunakan *library pandas* sebagai perintah untuk pengambilan data pada media social X atau twitter, data di *crawling* dengan menggunakan kata kunci “*diabetes*”, “*Penyakit diabetes*” dan “*diabetes Melitus*”, data tweet yang berhasil dikumpulkan yaitu 8.401 data tweet dengan rentan waktu dari 2023 sampai 2025. Kemudian data tersebut di simpan dalam format csv. Untuk di olah ke proses analisis sentiment selanjutnya. Hasil pengumpulan data dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

Username	Tweet
Celsiamgrt	@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu
Fshaffik	Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya. Saat ini lebih dari 19 juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak. https://t.co/0FuO37cbAO

3.2 Preprocessing

Preprocessing data merupakan tahap penting dalam membersihkan dan menyiapkan data untuk analisis sentimen. Tujuan utamanya adalah meningkatkan kualitas data, menghapus elemen yang tidak relevan, serta memformat teks agar lebih efisien dalam pemrosesan oleh algoritma. Berikut ini adalah beberapa langkah umum dalam praproses data untuk analisis sentimen:

3.2.1 Cleansing

Cleansing merupakan tahapan awal dari proses *preprocessing*, *cleansing* bertujuan menghilangkan data-data yang tidak relevan seperti data yang mengandung simbol, emoticon nama pengguna atau mention (@username) (#)hashtag, retweet dan tautan URL, sehingga data lebih bersih, konsisten dan layak digunakan untuk proses selanjutnya. Hasil dari proses cleaning dalam di lihat pada Tabel 2.

Tabel 2. Hasil Cleansing

Cleansing	
Sebelum Cleansing	Sesudah Cleansing
@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO	<i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok Coba pahami di sisi itu Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO

3.2.2 Case Folding

Proses *case folding* adalah proses mengubah seluruh teks dalam data yang mengandung huruf besar menjadi huruf kecil hal ini dilakukan agar data teks menjadi lebih konsisten dalam membandingkan sensitivitas huruf besar dan huruf kecil. Hasil case Folding dapat dilihat pada Tabel 3.

Tabel 3. Hasil Case Folding

<i>Case Folding</i>	
<i>Sebelum Case Folding</i>	<i>Sesudah Case Folding</i>
<p>@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu</p> <p>Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO</p>	<p><i>diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok coba pahami di sisi itu</p> <p>kasus penyakit <i>diabetes</i> di indonesia terus meningkat setiap tahunnya saat ini lebih dari juta penduduk dewasa indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcofuocbao</p>

3.2.3 Normalisasi Kata

Tahapan normalisasi kata bertujuan untuk meratakan atau mengubah kata-kata yang tidak resmi, istilah gaul, atau kata-kata yang tidak seragam menjadi bentuk yang baku atau standar. Hasil dari proses normalisasi kata dapat dilihat pada Tabel 4.

Tabel 4. Hasil Normalisasi Kata

Normalisasi	
Sesebelum Normalisasi	Sesudah Normalisasi
<p>@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu</p> <p>Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO</p>	<p><i>diabetes</i> memang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma sih perokok coba pahami di sisi itu</p> <p>kasus penyakit <i>diabetes</i> di indonesia terus meningkat setiap tahunnya saat ini lebih dari juta penduduk dewasa indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcofuocbao</p>

3.2.4 Tokenizing

Tokenizing merupakan proses pembagian kalimat atau teks secara terpisah. kalimat atau teks akan dibagi dengan tujuan untuk lebih mudah menghitung frekuensi kemunculan setiap kata dalam teks tersebut. Hasil dari proses tokenizing dapat dilihat pada Tabel 5.

Tabel 5. Hasil Tokenizing

<i>Tokenizing</i>	
<i>Sebelum Tokenizing</i>	<i>Sesudah Tokenizing</i>
<p>@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu</p> <p>Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap <i>diabetes</i> sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO</p>	<p>['<i>diabetes</i>', 'memang', 'penyakit', 'yang', 'dicari', 'sendiri', 'tapi', 'dampaknya', 'juga', 'dirasakan', 'sendiri', 'kalau', 'merokok', 'penyakit', 'dicari', 'sendiri', 'dampaknya', 'juga', 'merugikan', 'orang', 'lain', 'bukan', 'cuma', 'sih', 'perokok', 'coba', 'pahami', 'di', 'sisi', 'itu']</p> <p>['kasus', 'penyakit', '<i>diabetes</i>', 'di', 'indonesia', 'terus', 'meningkat', 'setiap', 'tahunnya', 'saat', 'ini', 'lebih', 'dari', 'juta', 'penduduk', 'dewasa', 'indonesia', 'mengidap', '<i>diabetes</i>', 'sehingga', 'menempati', 'peringkat', 'kelima', 'di', 'dunia', 'dengan', 'kasus', '<i>diabetes</i>', 'terbanyak', 'httpstcofuocbao']</p>

3.2.5 StopWord

Tujuan dari proses *stopword* adalah untuk mempermudah dan mempercepat pengolahan teks dengan cara mengeliminasi kata-kata yang tidak memiliki makna signifikan dan kurang relevan, serta menyimpan kata-kata yang memiliki arti penting. hasil dari *stopword* dapat dilihat pada Tabel 6.

Tabel 6. Hasil Stopword

<i>StopWord</i>

Sebelum Stop Word

@WidasSatyo *Diabetes* emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu

Kasus penyakit diabetes di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap diabetes sehingga menempati peringkat kelima di dunia dengan kasus *diabetes* terbanyak httpstcoFuOcbAO

Sesudah Stop Word

['*diabetes*', 'penyakit', 'dicari', 'dampaknya', 'dirasakan', 'merokok', 'penyakit', 'dicari', 'dampaknya', 'merugikan', 'orang', 'sih', 'perokok', 'coba', 'pahami', 'sisi']

['penyakit', '*diabetes*', 'indonesia', 'meningkat', 'tahunnya', 'juta', 'penduduk', 'dewasa', 'indonesia', 'mengidap', '*diabetes*', 'menempati', 'peringkat', 'dunia', '*diabetes*', 'httpstcofuocbao']

3.2.6 Stemming

Proses *stemming* merupakan tindakan menghapus afiks dari suatu kata sehingga didapatkan bentuk dasarnya. Tindakan ini dilakukan dengan cara membuang awalan, akhiran, sisipan, atau kombinasi dari imbuhan yang ada di sebuah kata. Proses *stemming* ini di lakukan dengan menggunakan *library* sastrawi. Hasil dari *Stemming* dapat dilihat pada Tabel 7.

Tabel 7. Hasil Stemming

Stemming	
Sebelum Stemming	Sesudah Stemming
@WidasSatyo <i>Diabetes</i> emang penyakit yang dicari sendiri tapi dampaknya juga dirasakan sendiri. Kalau merokok penyakit dicari sendiri dampaknya juga merugikan orang lain bukan cuma si perokok. Coba pahami di sisi itu	<i>diabetes</i> sakit cari dampak rasa rokok sakit cari dampak rugi orang sih okok coba paham sisi
Kasus penyakit <i>diabetes</i> di Indonesia terus meningkat setiap tahunnya Saat ini lebih dari juta penduduk dewasa Indonesia mengidap diabetes sehingga menempati peringkat kelima di dunia dengan kasus <i>diabetes</i> terbanyak httpstcoFuOcbAO	sakit <i>diabetes</i> indonesia tingkat tahun juta duduk dewasa indonesia idap diabetes tempat peringkat dunia <i>diabetes</i> httpstcofuocbao

3.3 Pelabelan Data

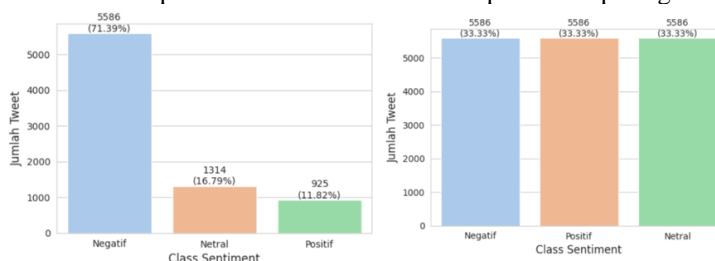
Pelabelan sentiment pada tahap ini dibagi menjadi tiga kategori, yaitu sentimen positif, sentimen negatif, dan sentimen netral. Proses pelabelan data dilakukan dengan menggunakan kamus leksikon serta program *Python* yang dijalankan di *Google Colaboratory*. Data yang telah ditandai menunjukkan 5586 data negatif, 1314 data netral, dan 925 data positif. Hasil dari pelabelan tersebut dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pelabelan Data

Tweet	Label
turun risiko <i>diabetes</i> jaga sehat jantung manfaat makan roti turun risiko <i>diabetes</i> roti kandungan serat bantu lambat serap gula darah bantu jaga kadar gula darah stabil cegah	Positif
<i>diabetes</i> sakit cari dampak rasa rokok sakit cari dampak rugi orang sih okok coba paham sisi	Negatif
hindar komplikasi <i>diabetes melitus</i> cerdas cegahptm dukunggermas <i>diabetes melitus</i> httpstcoubxnnns	Netral

3.4 Perbandingan Akurasi

Dari Penelitian ini data yang dihasilkan setelah tahapan pre-Processing yaitu 7825 datates dengan sentiment negatif yaitu 5586, sentiment netral yaitu 1314 dan sentiment positif yaitu 925 data. Data yang dihasilkan cenderung berlabel negatif sehingga menyebabkan ketidakseimbangan, ketidakseimbangan ini mengakibatkan kinerja model dalam mengklasifikasikan emosi positif dan netral menjadi kurang optimal. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan teknik SMOTE sebagai solusi untuk mengatasi ketidakseimbangan yang terjadi. Perbandingan sebelum dan sesudah implementasi Teknik SMOTE dapat dilihat pada gambar 2.



Gambar 2. Perbandingan Sebelum dan Sesudah SMOTE

Berdasarkan gambar 2 dapat dilihat bahwa perbandingan sebelum dan sesudah implelementasi SMOTE Untuk mengatasi ketidakseimbangan data pada kelas minoritas di tambahkan sebanyak 4.661 untuk data yang berlabel positif dan 4.272 untuk data yang berlabel netral. Penambahan data ini bertujuan agar model dapat memepelajari sentiment positif dan negatif secara seimbang. Setelah penerapan teknik SMOTE ini data untuk sentiment positif negatif dan netral menjadi setara yaitu 5586 data. Setelah data seimbang, langkah berikutnya adalah melatih dan menguji model algoritma. Dalam penelitian ini, setiap algoritma *Naïve Bayes*, *Random Forest*, dan *SVM* menggunakan pembagian data dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian. Data yang digunakan mencakup sebelum dan sesudah penerapan SMOTE, sehingga hasil dari masing-masing algoritma dapat dibandingkan untuk menentukan algoritma dengan performa terbaik. Hasil dari model algoritma sesudah dan sebelum penerapan SMOTE dapat di lihat pada tabel 9 dan 10.

Tabel 9. Hasil Akurasi Before SMOTE

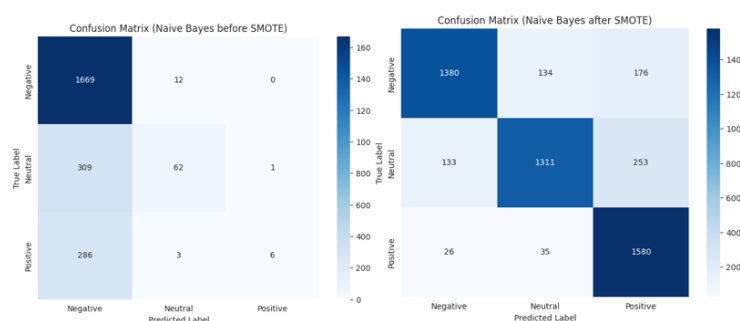
Model	Class	Accuracy	Precesion	Recalll	F1-Score
NBC	Negatif	74%	74%	99%	85%
	Netral		81%	17%	28%
	Positif		86%	2%	4%
RF	Negatif	82%	82%	98%	89%
	Netral		75%	45%	57%
	Positif		87%	33%	48%
SVM	Negatif	85%	88%	97%	92%
	Netral		68%	51%	59%
	positif		82%	57%	67%

Setelah dilakukan pembagian data dengan rasio 70:30 untuk pengujian model sebelum penerapan SMOTE, diperoleh hasil sebagai berikut: algoritma *Naïve Bayes* mencapai tingkat akurasi sebesar 74%, algoritma *Random Forest* memperoleh akurasi sebesar 82%, sedangkan model *SVM* menunjukkan akurasi sebesar 85%. Dari ketiga algoritma tersebut, *SVM* memiliki tingkat akurasi tertinggi.

Tabel 10. Hasil Akurasi After SMOTE

Model	Class	Accuracy	Precesion	Recalll	F1-Score
NBC	Negatif	85%	90%	82%	85%
	Netral		89%	77%	83%
	Positif		90%	86%	87%
RF	Negatif	95%	94%	94%	94%
	Netral		95%	94%	95%
	Positif		97%	98%	97%
SVM	Negatif	96%	97%	93%	95%
	Netral		94%	97%	95%
	Positif		97%	99%	96%

Setelah diterapkan *Synthetic Minority Oversampling Technique* (SMOTE), terjadi perubahan pada nilai evaluasi ketiga algoritma. Algoritma *Naïve Bayes* mengalami peningkatan akurasi dari 74% menjadi 85%, disertai dengan peningkatan nilai *precision*, *recall*, dan *F1-score*. Algoritma *Random Forest* juga mengalami peningkatan akurasi dari 82% menjadi 95%, serta peningkatan pada nilai *precision* dan *F1-score*. Namun, nilai *recall* untuk label negatif mengalami penurunan dari 98% menjadi 94%. Terakhir, algoritma *SVM* menunjukkan peningkatan akurasi dari 85% menjadi 96%, dengan kenaikan pada nilai *precision*, *recall*, dan *F1-score*. Selanjutnya, untuk menentukan algoritma terbaik, dilakukan perbandingan menggunakan *confusion matrix* di antara ketiga algoritma. *Confusion matrix* berfungsi untuk menggambarkan kinerja model dari ketiga algoritma yang digunakan, yaitu *Naïve Bayes*, *Random Forest*, dan *SVM*, baik sebelum maupun setelah penerapan SMOTE. Hasil perbandingan tersebut dapat dilihat pada Gambar 3, Gambar 4 dan Gambar 5.

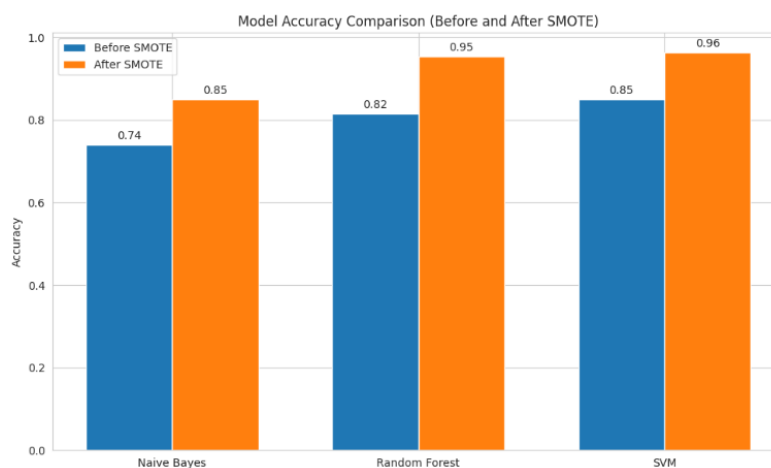


Gambar 3. Perbandingan Confusion Matrix Algoritma Naïve Bayes sesudah dan Sebelum SMOTE

Gambar 7 menampilkan wordcloud gabungan dari sentimen negatif, positif, dan netral terkait *diabetes*. Kata-kata seperti "*diabetes*", "komplikasi", "gestasional", "sakit", "darah", "kencing", "manis", "dampak", dan "gula" mendominasi. Istilah-istilah tersebut mencerminkan berbagai asumsi masyarakat, baik yang bersifat positif, netral, maupun negatif, mengenai penyakit *diabetes*.

3.5 Visualisasi Perbandingan Model

Visualisasi model dalam penelitian ini digunakan untuk menunjukkan perbandingan akurasi antara model *Naïve Bayes*, SVM, dan *Random Forest*. Penerapan *Synthetic Minority Oversampling Technique* (SMOTE) menghasilkan variasi kinerja pada setiap metode klasifikasi. Metode *Naïve Bayes* mengalami peningkatan akurasi dari 74% sebelum penerapan SMOTE menjadi 85% setelahnya. *Random Forest* juga menunjukkan peningkatan, dari 82% sebelum SMOTE menjadi 95% setelah penerapan. Demikian pula, akurasi SVM meningkat dari 85% menjadi 96% setelah SMOTE diterapkan. Secara keseluruhan, penerapan SMOTE terbukti efektif dalam meningkatkan akurasi pada ketiga metode tersebut. Perbandingan akurasi model dapat dilihat pada gambar 8.



Gambar 8. Hasil Perbandingan

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, disimpulkan bahwa pengumpulan data tweet terkait penyakit diabetes pada periode 2024–2025 menggunakan kata kunci tertentu dan metode crawling menghasilkan 8.401 data. Setelah tahap pra-pemrosesan, pelabelan, dan pembagian data dengan proporsi 70:30 (70% untuk training dan 30% untuk testing), diperoleh 5.586 data dengan sentimen negatif, 1.314 data netral, dan 925 data positif. Ketidakseimbangan distribusi sentimen ini menjadi tantangan utama dalam penelitian, sehingga diterapkan teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk menyeimbangkan jumlah data pada kelas minoritas dan mayoritas. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki kinerja lebih unggul dibandingkan *Naïve Bayes* dan *Random Forest*, baik sebelum maupun sesudah penerapan SMOTE. Sebelum SMOTE, *Naïve Bayes* mencapai akurasi 74%, *Random Forest* 82%, dan SVM 85%. Setelah SMOTE, terjadi peningkatan akurasi pada semua algoritma, dengan *Naïve Bayes* 85%, *Random Forest* 95%, dan SVM mencapai akurasi tertinggi 96%. Hasil ini menegaskan efektivitas SMOTE dalam mengatasi ketidakseimbangan data, terutama dalam meningkatkan kinerja model untuk klasifikasi sentimen positif. Untuk penelitian selanjutnya, studi ini dapat diperluas dengan membandingkan lebih banyak algoritma machine learning atau deep learning seperti LSTM, BERT, atau model berbasis Transformer yang umum digunakan dalam pemrosesan bahasa alami (NLP). Selain itu, dapat dieksplorasi teknik pra-pemrosesan teks yang lebih canggih seperti lemmatization dan word embeddings (Word2Vec, FastText, atau BERT embeddings) guna meningkatkan pemahaman konteks dalam analisis sentimen.

REFERENCES

- [1] Y. F. Wijaya and A. Triayudi, "Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 165–174, Nov. 2023, doi: 10.47065/josyc.v5i1.4614.
- [2] Y. Siagian, J. Hutahaean, A. Zikra Syah, J. Efendi Hutagalung, and A. Karim, "Implementasi Metode K-Nearest Neighbours (KNN) Untuk Klasifikasi Penyakit Diabetes," *Jurnal Informatika dan Teknologi Informasi*, vol. 2, no. 3, pp. 253–262, Jan. 2024, doi: 10.56854/jt.v2i3.331.
- [3] H. Nugroho, S. Fadwa Syamlan Fakultas Kesehatan, U. Ichlan Satya, J. Jombang Raya No, and T. Selatan, "Penyuluhan Kesehatan tentang Pengetahuan Pencegahan Diabetes Militus di Desa Rawat Rengas," *Jurnal Peduli Masyarakat*, vol. 5, no. 4, pp. 1063–1070, Oct. 2023, doi: 10.37287/JPM.V5I4.2358.



- [4] P. Keperawatan and M. Bedah, “Skripsi Gambaran Kepatuhan Diet Pasien Diabetes Melitus Tipe 2 Di Puskesmas Poto Tano Kabupaten Sumbawa Barat Tahun 2020,” 2020.
- [5] R. M. Harun and F. Fahmi, “Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree Pada Analisis Sentimen Terhadap Ulasan Aplikasi KitaLulus,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 2033–2042, 2024, doi: 10.47065/bits.v6i3.6367.
- [6] S. A. Sutresno, “Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, Mar. 2023, doi: 10.47065/bits.v4i4.3149.
- [7] A. A. Asyer and M. A. I. P. I. Pakereng, “Analisis Sentimen Tweet Pengguna Twitter Terkait Diabetes Menggunakan Metode Naive Bayes,” *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 12, no. 2, pp. 627–636, 2023, doi: 10.35889/jutisi.v12i2.1234.
- [8] J. Adiputra and D. Mahdiana, “Analisis Sentimen Dengan Algoritma Support Vector Machine Terhadap Penyakit Hepatitis Akut Misterius,” *IDEALIS: InDonEsiA journal Information System*, vol. 6, no. 1, pp. 1–8, Jan. 2023, doi: 10.36080/IDEALIS.V6I1.2985.
- [9] L. Sari, A. Romadloni, and R. Listyaningrum, “Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest,” *Infotekmesin*, vol. 14, no. 1, pp. 155–162, Jan. 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [10] P. Elisa and A. Rahman Isnain, “Comparison Of Random Forest, Support Vector Machine And Naive Bayes Algorithms To Analyze Sentiment Towards Mental Health Stigma,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 321–329, 2024, doi: 10.52436/1.jutif.2024.5.1.1817.
- [11] M. Y. Aldean, P. Paradise, and N. A. S. Nugraha, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac),” *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 4, no. 2, pp. 64–72, Jun. 2022, doi: 10.20895/INISTA.V4I2.575.
- [12] R. Nurhidayat and N. Hendrastuty, “Analisis Sentimen Komentar Media Sosial Twitter Terhadap Tes CPNS dengan Algoritma Naive Bayes,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1477–1489, 2024, doi: 10.47065/bits.v6i3.6148.
- [13] J. Khab Sulaiman Dalam, A. Oktavia Praneswara, N. Cahyono, and U. Amikom Yogyakarta, “Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes,” *The Indonesian Journal of Computer Science*, vol. 12, no. 6, p. 3925, Dec. 2023, doi: 10.33022/IJCS.V12I6.3473.
- [14] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, Oct. 2021, doi: 10.30591/JPIT.V6I3.2870.
- [15] R. R. Salam, M. F. Jamil, Y. Ibrahim, R. Rahmaddeni, S. Soni, and H. Herianto, “Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 27–35, May 2023, doi: 10.57152/MALCOM.V3I1.590.
- [16] D. Toresa, S. Rico, F. Sitorus, I. Muzdalifah, F. Wiza, and R. Syelly, “Analisis Sentimen Terhadap Ulasan Penggunaan Dompot Digital Dana Menggunakan Metode Klasifikasi Support Vector Machine,” *Technologica*, vol. 3, no. 2, pp. 64–74, Jul. 2024, doi: 10.55043/TECHNOLOGICA.V3I2.163.
- [17] S. Sari *et al.*, “Analisis Sentimen Terhadap Komentar Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength,” *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 1, no. 1, pp. 71–78, Apr. 2021, doi: 10.57152/IJIRSE.V1I1.55.
- [18] A. Syah, F. Nurdiansyah, and A. Y. Rahman, “Analisis Sentimen Aplikasi Shopee, Tokopedia, Lazada Dan Bibli Menggunakan Leksikon Dan Random Forest,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5155.
- [19] I. K. Dharmendra, I. M. A. W. Putra, and Y. P. Atmojo, “Evaluation of the Effectiveness of SMOTE and Random Under Sampling in Emotion Classification of Tweets,” *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, vol. 9, no. 2, pp. 182–193, Dec. 2024, doi: 10.51211/ITBI.V9I2.3183.
- [20] D. Alita and A. R. Isnain, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *Jurnal Komputasi*, vol. 8, no. 2, pp. 50–58, 2020.
- [21] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [22] E. Dwi, K. Wardani, F. Febiola Yo, W. N. Meylugita, F. Sains, and D. Teknologi, “Implementasi Algoritma Naïve Bayes Untuk Analisis Ulasan Pengguna Untuk Aplikasi Seabank Di Google Play Store Implementation Of The Naive Bayes Algorithm For User Review Analysis Of The Seabank Application On Google Play Store,” *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 1, pp. 2963–6191, 2025, doi: 10.69916/jkbt.v4i1.193.
- [23] S. A. Riyandona, N. Rahaningsih, R. D. Dana, and - Mulyawan, “Implementasi Model Analisis Sentimen Terhadap Grup Musik Bts Menggunakan Metode Naïve Bayes,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5816.
- [24] R. Z. Firdaus, S. H. Wijoyo, and W. Purnomo, “Analisis Sentimen Berbasis Aspek Ulasan Pengguna Aplikasi Alfagift Menggunakan Metode Random Forest dan Pemodelan Topik Latent Dirichlet Allocation,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, 2025.
- [25] I. Amelia, Sugiyono, F. M. Sarimole, and Tundo, “Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah dengan Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 2994–3003, Sep. 2024, doi: 10.35870/JIMIK.V5I3.990.
- [26] M. G. Andriawan and T. Ernawati, “PENGUNAAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN KONFLIK PALESTINA DAN ISRAEL PADA PLATFORM X,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4943.