

Perbandingan Algoritma NBC, SVM dan Random Forest untuk Analisis Sentimen Implementasi Starlink pada Media Sosial X

Lintang Kencono, Dedi Darwis*

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

Email: ¹kencono@teknokrat.ac.id, ^{2,*}darwisdedi@teknokrat.ac.id

Email Penulis Korespondensi: darwisdedi@teknokrat.ac.id

Submitted: 22/01/2025; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Perkembangan internet di Indonesia terus mengalami kemajuan pesat, namun pemerataan akses masih menjadi tantangan, terutama di wilayah terpencil. *Starlink*, layanan internet satelit dari *SpaceX*, hadir sebagai solusi untuk mengurangi kesenjangan ini dengan menyediakan konektivitas yang cepat dan stabil. Penelitian ini menganalisis sentimen publik terhadap implementasi *Starlink* di platform media sosial X melalui pendekatan komparatif menggunakan tiga algoritma *Machine Learning* *Naive Bayes Classifier*, *Support Vector Machine*, dan *Random Forest*. Data penelitian terdiri dari 6.780 *tweet* berbahasa Indonesia yang dikumpulkan selama periode 1 September hingga 30 November 2024 menggunakan *library tweet harvest* dengan kata kunci "*starlink*," "*internet starlink*," dan "*starlink SpaceX*". Setelah proses *preprocessing*, sebanyak 5.382 *tweet* digunakan, terdiri dari 4.348 *tweet* dengan sentimen negatif dan 884 *tweet* dengan sentimen positif. Untuk mengatasi ketidakseimbangan data, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Sebelum penerapan SMOTE, model *Random Forest* menunjukkan akurasi tertinggi sebesar 92%, diikuti oleh *Support Vector Machine* dengan 91%, dan *Naive Bayes Classifier* dengan 85%. Setelah SMOTE diterapkan, akurasi ketiga model meningkat signifikan, dengan *Random Forest* mencapai 99%, *Support Vector Machine* 98%, dan *Naive Bayes Classifier* 91%. *Random Forest* juga menunjukkan kinerja terbaik dalam mendeteksi sentimen positif, dengan nilai *Precision* dan *Recall* mencapai 100%. Penelitian ini memberikan wawasan mendalam mengenai efektivitas algoritma *Machine Learning* dalam analisis sentimen publik terhadap layanan *Starlink* di media sosial serta menunjukkan bahwa penerapan SMOTE dapat meningkatkan performa model dalam mengklasifikasikan sentimen dengan lebih seimbang.

Kata Kunci: Machine Learning; Media Sosial; Sentimen Publik; SMOTE; Starlink

Abstract—Internet development in Indonesia continues to progress rapidly, but equitable access remains a challenge, especially in remote areas. *Starlink*, a satellite internet service from *SpaceX*, comes as a solution to reduce this gap by providing fast and stable connectivity. This research analyzes public sentiment towards the implementation of *Starlink* on social media platform X through a comparative approach using three *Machine Learning* algorithms: *Naive Bayes Classifier*, *Support Vector Machine*, and *Random Forest*. The research data consisted of 6,780 Indonesian tweets collected during the period September 1 to November 30, 2024 using the *harvest tweet library* with the keywords "*starlink*," "*internet starlink*," and "*SpaceX starlink*". After *preprocessing*, a total of 5,382 tweets were used, consisting of 4,348 tweets with negative sentiment and 884 tweets with positive sentiment. To overcome data imbalance, *Synthetic Minority Over-sampling Technique* (SMOTE) was applied. Before the application of SMOTE, the *Random Forest* model showed the highest accuracy of 92%, followed by *Support Vector Machine* with 91%, and *Naive Bayes Classifier* with 85%. After SMOTE was applied, the accuracy of the three models increased significantly, with *Random Forest* reaching 99%, *Support Vector Machine* 98%, and *Naive Bayes Classifier* 91%. *Random Forest* also showed the best performance in detecting positive sentiment, with *Precision* and *Recall* values reaching 100%. This research provides an in-depth insight into the effectiveness of *Machine Learning* algorithms in analyzing public sentiment towards *Starlink* services on social media and shows that the application of SMOTE can improve the model's performance in classifying sentiment more evenly.

Keywords: Machine Learning; Social Media; Public Sentiment; SMOTE; Starlink

1. PENDAHULUAN

Perkembangan internet di Indonesia telah menunjukkan kemajuan pesat dalam beberapa tahun terakhir. Meskipun demikian, pemerataan akses internet masih menjadi masalah yang belum sepenuhnya terselesaikan. Dengan populasi yang besar dan wilayah yang luas serta beragam, Indonesia menghadapi tantangan dalam memastikan seluruh warganya mendapatkan akses yang memadai terhadap infrastruktur digital. Sebagai negara yang memiliki ambisi untuk menjadi negara maju pada tahun 2045, akses internet menjadi salah satu komponen kunci dalam mewujudkan tujuan tersebut. Transformasi digital yang menjadi prioritas pemerintah membutuhkan infrastruktur internet yang handal dan tersebar merata di seluruh wilayah Indonesia[1].

Di tengah perkembangan ini, kebutuhan akan akses internet cepat dan andal semakin mendesak, baik di Indonesia maupun di seluruh dunia. Indonesia, sebagai negara kepulauan terbesar dengan lebih dari 17.000 pulau, menghadapi tantangan besar dalam menyediakan infrastruktur internet yang merata. Kesenjangan digital antara kota besar dan daerah pedesaan masih menjadi masalah utama. Banyak daerah terpencil yang belum tersentuh jaringan broadband berkualitas tinggi, sehingga memperburuk ketimpangan akses internet[2]. Penyediaan layanan internet di Indonesia menghadapi tantangan khusus, karena penerapan infrastruktur kabel dan serat optik seringkali sulit dan mahal di banyak wilayah. Dalam hal ini, *Starlink* muncul sebagai solusi efisien yang dapat menjangkau daerah-daerah terpencil dengan keterbatasan infrastruktur[3].

Starlink, layanan internet satelit yang dikembangkan oleh *SpaceX*, menawarkan teknologi yang bisa mengatasi kendala tersebut. Dengan menggunakan teknologi satelit Orbit Rendah Bumi (LEO), *Starlink* mampu menyediakan cakupan internet yang luas dan stabil dengan kecepatan tinggi. Hingga saat ini, lebih dari 5.000 satelit telah diluncurkan, yang memberikan potensi besar untuk mengubah lanskap konektivitas internet di Indonesia[4]. Melalui

solusi ini, *Starlink* berpotensi mengatasi kesenjangan akses internet, khususnya di daerah pedesaan dan terpencil, yang sebelumnya kesulitan mendapatkan akses internet memadai.

Selain itu, *Starlink* menawarkan berbagai keunggulan dibandingkan internet satelit tradisional, seperti kecepatan internet yang lebih tinggi, latensi yang lebih rendah, dan cakupan yang lebih luas. Teknologi ini memiliki fleksibilitas yang dapat membantu mengatasi kesenjangan akses internet, terutama di daerah terpencil[5]. Dengan potensi besar ini, *Starlink* dapat meningkatkan kualitas hidup masyarakat, membuka akses pendidikan daring, layanan kesehatan jarak jauh, dan peluang ekonomi digital, termasuk mendukung UMKM untuk memperluas pasar secara online. Namun, keberhasilan *Starlink* tidak hanya bergantung pada teknologinya, tetapi juga pada penerimaan masyarakat. Faktor seperti harga layanan, kesesuaian kebutuhan lokal, dan kesadaran manfaat teknologi ini akan menentukan keberhasilannya. Peran pemerintah juga penting untuk memastikan akses yang merata dan mengurangi kesenjangan digital di Indonesia.

Penerimaan publik terhadap teknologi seperti Starlink menjadi faktor penting untuk memastikan keberhasilan implementasinya di Indonesia. Oleh karena itu, diperlukan evaluasi tentang bagaimana masyarakat merespons dan memandang layanan yang ditawarkan. Salah satu metode yang efektif untuk memahami respons ini adalah analisis sentimen, yang dapat mengidentifikasi apakah opini publik terhadap Starlink bersifat positif atau negatif[6]. Media sosial X, yang memiliki banyak pengguna di Indonesia, merupakan platform yang tepat untuk mengumpulkan data mengenai opini publik. Melalui analisis cuitan dan penggunaan hashtag, kita dapat dengan mudah memahami bagaimana masyarakat membahas dan mendiskusikan Starlink. Hal ini menjadikan media sosial X sebagai sarana yang efektif untuk mendapatkan gambaran penerimaan publik terhadap layanan ini, khususnya sebagai penyedia internet di daerah terpencil dengan kualitas koneksi yang minim[7].

Seiring perkembangan teknologi, layanan internet satelit seperti Starlink menawarkan potensi besar untuk memperluas akses internet, terutama di wilayah-wilayah yang sulit dijangkau oleh layanan tradisional. Teknologi ini menghadirkan peluang untuk mengurangi kesenjangan digital dan mendukung pembangunan sosial dan ekonomi di daerah terpencil. Namun, untuk memastikan keberhasilan implementasi Starlink, penting untuk memahami persepsi dan pandangan masyarakat terhadap teknologi ini. Analisis sentimen menjadi salah satu pendekatan utama yang dapat menggambarkan opini publik secara objektif. Beberapa penelitian sebelumnya telah membuktikan bahwa algoritma Machine Learning sangat efektif untuk menganalisis sentimen publik terhadap berbagai layanan internet, termasuk layanan baru seperti Starlink.

Salah satunya adalah penelitian oleh Bebin Paula dkk (2024) menggunakan Naive Bayes untuk menganalisis sentimen terhadap Starlink di Indonesia mencapai akurasi sebesar 64%[7]. Sardin dkk (2024) menggunakan algoritma SVM pada data sentimen Starlink di Twitter dan mencatat akurasi 76,22%, presisi 77,48%, dan recall 81,38%[4]. M. Khalil Gibran dkk (2024) juga menggunakan Naive Bayes untuk menganalisis sentimen terhadap Starlink, dengan akurasi 80%, recall 0,80, dan presisi 0,90[8]. Selain fokus pada Starlink, ada penelitian lain yang mengkaji sentimen terhadap layanan internet lokal, seperti yang dilakukan Muhammad Nur Akbar dkk (2022) menemukan bahwa algoritma SVM memiliki akurasi tinggi, yakni 91,3%, dalam analisis sentimen pengguna Indihome[9]. Lady Agustin Fitriana dkk (2024) menggunakan Naive Bayes untuk menganalisis sentimen pengguna Twitter terhadap Indosat, menghasilkan akurasi 83,91%[10]. Farin Junita Fauzan dkk (2024) membandingkan SVM dan Random Forest pada ulasan MyTelkomsel di Google Playstore, dengan SVM mencapai akurasi 83% sebagai yang tertinggi[11]. Adrian Nur Ihsan & Shandy Tresnawati (2024) menunjukkan bahwa SVM lebih unggul dibandingkan Naive Bayes dalam analisis sentimen terhadap provider Tri, dengan akurasi tertinggi 76,10%[12].

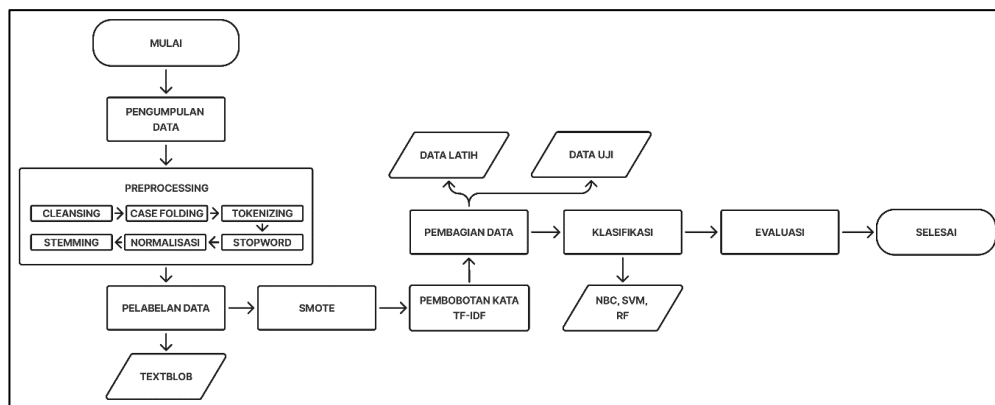
Meskipun banyak penelitian telah menganalisis sentimen menggunakan algoritma *Machine Learning* seperti *Naive Bayes*, *Support Vector Machine*, dan *Random Forest*, masih sedikit yang secara khusus membahas implementasi *Starlink* di media sosial X. Sebagian besar studi hanya menggunakan satu atau dua algoritma tanpa perbandingan menyeluruh, padahal opini publik mengenai layanan internet satelit ini sangat beragam. Penelitian ini mengisi kesenjangan tersebut dengan membandingkan langsung kinerja ketiga algoritma dalam analisis sentimen publik terhadap *Starlink*. Pendekatan komparatif ini memberikan wawasan yang lebih mendalam dibandingkan penelitian sebelumnya karena tidak hanya menilai akurasi, tetapi juga membandingkan kecepatan pemrosesan, ketahanan terhadap ketidakseimbangan data, serta kemampuan menangkap nuansa sentimen yang kompleks. Dengan demikian, penelitian ini tidak hanya menunjukkan algoritma yang paling efektif, tetapi juga menjelaskan faktor-faktor yang mempengaruhi performa dalam konteks analisis sentimen terkait teknologi baru. Hasil penelitian ini dapat memberikan pemahaman lebih luas tentang persepsi publik terhadap *Starlink*, terutama di daerah terpencil, serta membantu penyedia layanan dalam menilai penerimaan masyarakat terhadap teknologi mereka. Selain itu, pendekatan ini dapat menjadi model analisis sentimen untuk inovasi teknologi lainnya di masa depan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini bertujuan untuk menganalisis sentimen terkait isu *Starlink* di platform Media Sosial X, dimulai dengan mengumpulkan data opini yang diberi label positif dan negatif. Data yang diperoleh kemudian diproses melalui tahap *preprocessing*, termasuk pembersihan teks dengan menghapus *stopwords*, melakukan *stemming*, dan menghilangkan karakter-karakter yang tidak relevan. Selanjutnya, teknik SMOTE (*Synthetic Minority Over-sampling Technique*)

digunakan untuk mengatasi ketidakseimbangan data dengan menghasilkan data sintesis untuk kelas minoritas (positif). Kata-kata dalam data dibobot menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*), dan data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model klasifikasi dibangun dengan menggunakan tiga algoritma, yaitu *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Random Forest*. Evaluasi kinerja model dilakukan dengan menggunakan metrik seperti akurasi dan *F1-score*, dan hasil klasifikasi divisualisasikan untuk mempermudah interpretasi. Langkah-langkah penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

2.2 Pengumpulan Data

Dalam penelitian ini, data dikumpulkan melalui metode *crawling* dari media sosial X menggunakan Google Colab dan diproses dengan bahasa pemrograman *Python*. *Python* dipilih karena fleksibilitasnya, sifatnya yang *interpretatif*, dan struktur yang terorganisir, didukung oleh beragam pustaka untuk analisis data skala besar[13]. Proses *crawling* dilakukan menggunakan pustaka *Python* seperti *Harvest* dan *Node.js*, dengan data yang diperoleh disimpan dalam format CSV melalui kode yang dijalankan di *Google Colab*[14]. Data yang digunakan berupa tweet berbahasa Indonesia dengan kata kunci "Starlink," "internet starlink," dan "starlink spaceX." Sebanyak 6.780 tweet berhasil dikumpulkan dalam periode 1 September 2024 hingga 30 November 2024.

2.3 Preprocessing

Preprocessing adalah tahap dalam analisis data, khususnya pada data teks, yang bertujuan untuk mempersiapkan dan membersihkan data sehingga lebih siap untuk dianalisis lebih lanjut[15]. Tahapan *Preprocessing* yaitu *Cleansing*, *Case Folding*, *Tokenizing*, *Stopword Removal*, *Normalisasi* dan *Stemming*. Sebagai berikut :

- Cleansing* adalah proses membersihkan data dari *noise* seperti tanda baca dan karakter yang tidak relevan, Proses ini mencakup penghapusan elemen seperti indikator retweet (RT), *hashtag* (#), *mention* (@username), tautan (URL), simbol dan tanda baca seperti tanda seru (!) dan tanda tanya (?), angka, karakter *non-alfabet*, serta spasi berlebih[16]. Langkah ini dilakukan untuk memastikan data lebih bersih, terstruktur, dan siap untuk analisis lebih lanjut.
- Case folding* adalah proses mengubah semua karakter dalam teks menjadi huruf kecil serta menghapus tanda baca dan angka. Proses ini hanya memproses huruf alfabet dari "a" hingga "z," sehingga karakter selain huruf akan dihapus untuk memastikan data lebih konsisten dan siap untuk analisis[17].
- Tokenizing* adalah tahap di mana kalimat-kalimat dalam teks dipecah menjadi kata-kata tunggal atau token[18]. Tujuan dari proses ini adalah untuk memberikan bobot nilai pada setiap dokumen dengan memecah karakter-karakter dalam teks menjadi satuan kata yang dapat dianalisis lebih lanjut[19].
- Stopword removal* adalah langkah untuk menghapus kata-kata yang tidak relevan, seperti "dan", "atau", dan "yang", karena tidak memberikan informasi penting[20].
- Normalisasi* adalah proses mengubah teks menjadi bentuk yang lebih terstruktur dan standar. Langkah ini dilakukan dalam pemrosesan bahasa alami dan analisis teks untuk memastikan konsistensi dan keseragaman data[21].
- Stemming* adalah proses untuk memperoleh kata dasar dengan cara menghilangkan awalan, akhiran, kata sisipan, dan *confixes* (kombinasi awalan dan akhiran) menggunakan pustaka *Sastrawi* untuk kata-kata dalam bahasa Indonesia[22].

2.4 Pelabelan Data

Proses pelabelan menggunakan *library TextBlob* dilakukan dengan menghitung nilai *polarity* dan *subjectivity* untuk mengklasifikasikan teks secara otomatis [23]. *Polarity* menunjukkan kecenderungan sentimen dalam sebuah teks, apakah bernada positif atau negatif. Sementara itu, *subjectivity* mengukur sejauh mana sebuah teks mengandung opini atau fakta. Semakin tinggi nilai *subjectivity*, semakin besar kemungkinan teks tersebut merupakan opini. Di sisi lain, semakin tinggi nilai *polarity*, semakin positif emosi yang terkandung dalam teks tersebut[24]. Berdasarkan nilai

polarity, teks dapat diklasifikasikan ke dalam dua kelas utama, yaitu positif dan negatif. Dapat dilihat pada persamaan 1 berikut :

$$S_{sentiment} = \begin{cases} \text{positive if } S_{positive} > S_{negative} \\ \text{negative if } S_{positive} < S_{negative} \end{cases} \quad (1)$$

Dimana $S_{sentiment}$ kalimat sentiment pada dataset, $S_{positive}$ Kalimat sentimen positif pada dataset, $S_{negative}$ Kalimat sentimen negatif pada dataset

2.5 SMOTE

Synthetic Minority Oversampling Technique (SMOTE) adalah metode *oversampling* yang dikembangkan oleh Nithes V. Chawla untuk mengatasi ketidakseimbangan kelas dalam dataset[25]. Teknik ini membuat data sintetis dari sampel kelas *minoritas* untuk menyamakan jumlahnya dengan kelas *mayoritas*, sehingga meningkatkan kinerja model klasifikasi. Ketidakseimbangan kelas terjadi ketika jumlah sampel pada satu kelas jauh lebih sedikit dibandingkan kelas lainnya, yang dapat menyebabkan model terlalu terfokus pada kelas mayoritas. SMOTE menghasilkan data baru secara sintetis, bukan hanya menyalin data yang ada, sehingga mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model[26]. Meskipun SMOTE dapat meningkatkan keseimbangan dataset dan kinerja model, teknik ini berpotensi menyebabkan *overfitting* jika data sintetis terlalu mirip dengan sampel asli. *Overfitting* terjadi ketika model terlalu menyesuaikan diri dengan data pelatihan, sehingga kinerjanya menurun pada data uji. Untuk mengatasinya, dapat diterapkan kombinasi SMOTE dengan *undersampling*, validasi silang, dan regularisasi. Selain itu, membandingkan performa model dengan dan tanpa SMOTE membantu menilai efektivitasnya dalam meningkatkan akurasi dan generalisasi.

2.6 Pembobotan Kata TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengukur pentingnya sebuah term dalam dokumen terkait koleksi dokumen yang lebih besar. TF-IDF menghitung bobot term berdasarkan dua komponen utama, *Term Frequency* (TF), yang mengukur frekuensi kemunculan term dalam dokumen, dan *Inverse Document Frequency* (IDF), yang menghitung seberapa jarang term tersebut muncul di seluruh dokumen[27]. Bobot TF-IDF mencerminkan relevansi *term* terhadap dokumen, sehingga sering digunakan dalam pemrosesan bahasa alami dan sistem temu kembali informasi untuk meningkatkan akurasi dan relevansi hasil pencarian[28]. Berikut adalah rumus untuk menghitung *Term Frequency-Inverse Document Frequency* (TF-IDF) dapat dilihat pada persamaan 2.

$$TF * IDF(d, t) = TF(d, t) * \log \frac{N}{df(t)} \quad (2)$$

Dimana $TF(d, t)$ - *Term Frequency* (Frekuensi Kemunculan *Term*), TF mengukur seberapa sering *term t* muncul dalam dokumen *d.*, $IDF(t)$ - *Inverse Document Frequency* (Frekuensi Dokumen Terbalik), IDF mengukur seberapa jarang *term t* muncul dalam semua dokumen di korpus, N Jumlah total dokumen dalam korpus. $df(t)$ Jumlah dokumen yang mengandung *term t*.

2.7 Pembagian Data

Pembagian data menjadi dataset *training* dan dataset *test*, dilakukan untuk mengukur kinerja model pada data yang belum pernah dilihat sebelumnya. Dalam proses ini, data dibagi menjadi 80% untuk *training* dan 20% untuk *test*. Pembagian ini memungkinkan model untuk dilatih pada sebagian besar data, dan diuji secara objektif menggunakan data yang tidak terlibat dalam proses pelatihan, sehingga dapat menilai kemampuan generalisasi model terhadap data baru[29]. Setelah pembagian data, klasifikasi dilakukan menggunakan beberapa algoritma, seperti *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Random Forest*.

2.8 Klasifikasi NBC, SVM, Random Forest

2.8.1 Naive Bayes Classifier

Naive Bayes adalah metode klasifikasi yang sederhana, yang menghitung probabilitas menggunakan *teorema Bayes* dan menggabungkannya dengan frekuensi nilai dari database. Dalam konteks klasifikasi, tujuannya adalah untuk memprediksi label kelas dari sampel yang diberikan berdasarkan sekumpulan fitur atau karakteristik[30]. Secara khusus, metode *Naive Bayes Classifier* sangat efisien untuk klasifikasi teks, karena hanya memerlukan sedikit data masukan selama proses klasifikasi. Tujuan utama dari metode ini adalah untuk meminimalkan varians dalam dokumen, sehingga menghasilkan prediksi yang lebih akurat dan konsisten[31]. Rumus perhitungannya adalah sebagai berikut dapat dilihat pada persamaan 3.

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

Dimana $P(H|X)$ adalah *probabilitas hipotesis H* (kelas) diberikan data X (fitur), $P(X|H)$ adalah *probabilitas likelihood* dari fitur X diberikan *hipotesis H*. $P(H)$ adalah probabilitas apriori dari *hipotesis H*. $P(X)$ adalah *probabilitas total*

dari data X, yang biasanya dianggap konstan dan dapat diabaikan dalam perhitungan jika kita membandingkan beberapa kelas.

2.8.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang diawasi (*supervised learning*) dan sering digunakan untuk tugas klasifikasi serta analisis data karena kecepatan dan hasil yang baik. SVM bekerja dengan mencari *hyperplane* atau bidang pemisah optimal yang memisahkan kelas-kelas dalam dataset, seperti sentimen positif dan negatif[32]. SVM bertujuan memaksimalkan jarak antara *hyperplane* dan data terdekat dari setiap kelas untuk meningkatkan akurasi. Jika data tidak dapat dipisahkan secara linear, SVM menggunakan kernel untuk memetakan data ke ruang fitur berdimensi lebih tinggi, memudahkan pemisahan kelas. SVM dilatih dengan optimasi untuk menemukan *hyperplane* terbaik yang memaksimalkan margin pemisah, menghasilkan klasifikasi yang lebih akurat.[33]. Berikut adalah rumus-rumus utama dalam algoritma *Support Vector Machine* (SVM) disajikan pada persamaan 4,5 dan 6.

Pasangan data dan kelas :

$$\{(x_i, y_i)\}_{i=1}^N \tag{4}$$

Menghitung nilai w dan b :

$$w = \sum_{i=1}^N a_i y_i x_i \quad b = -\frac{1}{2} (w \cdot x^+ + w \cdot x^-) \tag{5}$$

Dimana w Bobot yang menentukan arah *hyperplane*, b Bias yang menggeser *hyperplane* agar sesuai dengan data Fungsi Keputusan klasifikasi *sign* (f(x)) :

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

f(x) Hasil prediksi kelas dari data masukan x., K(x, x_i) Fungsi *kernel* untuk menghitung kesamaan antara x dan x_i, m Jumlah data pelatihan yang relevan.

2.8.3 Random Forest

Random Forest adalah metode *machine learning* yang menggunakan kumpulan pohon keputusan (*Decision Tree*) untuk klasifikasi, dengan keputusan akhir diperoleh melalui *voting mayoritas*. Algoritma ini menerapkan sampling terpandu untuk membangun pohon prediksi dan menggunakan prediktor acak[34]. Hasil prediksi dikombinasikan dengan *majority vote* untuk klasifikasi dan rata-rata untuk regresi. *Random Forest* memiliki akurasi tinggi, tahan terhadap *outliers* dan *noise*, serta lebih cepat dibandingkan *bagging* dan *boosting*[35]. Berikut adalah rumus perhitungan *random forest* menggunakan persamaan 7.

$$f(x) = \text{Average}(f_1(x), f_2(x), \dots, f_n(x)) \tag{7}$$

Dimana f(x) Hasil prediksi, f_{1-n}(x) Hasil prediksi dari setiap pohon keputusan ke-n, (x) Inputan data

2.9 Evaluasi Model

Confusion matrix adalah tabel yang membandingkan hasil prediksi dengan label sesungguhnya untuk mengevaluasi akurasi, presisi, dan recall. Terdapat empat kemungkinan: *true positive* (positif diprediksi positif), *false negative* (positif diprediksi negatif), *true negative* (negatif diprediksi negatif), dan *false positive* (negatif diprediksi positif)[36]. Contoh *confusion matrix* untuk klasifikasi biner dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Actual	Prediction	
	Positive	Negatif
Positive	True Positive (TP)	True Negative (FN)
Negatif	False Positive (FP)	False Negative (TN)

Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Rumus-rumus berikut digunakan untuk menghitung akurasi, presisi, dan recall berdasarkan nilai-nilai dalam *confusion matrix* disajikan pada persamaan 8,9,10 dan 11.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{8}$$

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{10}$$

$$f1 - \text{score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \tag{11}$$

Akurasi mengukur proporsi prediksi yang benar, dihitung dengan membagi jumlah prediksi benar (TP + TN) dengan total prediksi (TP + TN + FP + FN), lalu dikalikan 100%. Presisi mengukur akurasi model dalam memprediksi kelas positif, dihitung dengan membagi TP dengan total prediksi positif (TP + FP). Recall mengukur kemampuan model dalam menemukan semua data positif, dihitung dengan membagi TP dengan total data positif sebenarnya (TP + FN). *F1-Score* adalah rata-rata harmonik antara presisi dan recall, dihitung dengan rumus $2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$, memberikan keseimbangan antara keduanya, terutama pada dataset yang tidak seimbang.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan melalui proses *crawling* media sosial X menggunakan metode *tweet harvest*. Dalam rentang waktu 1 September hingga 30 November 2024, sebanyak 6.780 *tweet* berhasil dikumpulkan dengan kata kunci utama "Starlink". Selanjutnya data dibersihkan dari data duplikat yang sehingga data bersih sebanyak 5.382 *tweet*, kemudian data disimpan dalam format CSV untuk mempermudah proses analisis sentimen. Proses *crawling* ini memanfaatkan Twitter API yang diintegrasikan dengan Google Colab menggunakan bahasa pemrograman Python. Hasil dari pengumpulan data tersebut disajikan dalam Tabel 2 sebagai dasar untuk analisis lebih lanjut.

Tabel 2. Hasil Pengumpulan Data Tweet

No	Username	Tweet
1	_rzkymln	Pertama kali nyoba <i>starlink</i> punya om @elonmusk. Ternyata emg <i>amazing</i> ya <i>speed</i> sama pingnya untuk sekelas internet satelit. Gokill https://t.co/T8kf7DGdbm
5382	Juniorama_	Tapi para tetua protes karena anak muda mengakses <i>pornografi</i> . Duhhh... â€• i• https://t.co/2N5CHsC8Ke

Berdasarkan sampel data pada Tabel 2 yang berhasil dikumpulkan, langkah berikutnya adalah melakukan proses *preprocessing* data untuk mempersiapkan data agar siap dianalisis.

3.2 Preprocessing

Setelah melakukan pengumpulan data, langkah berikutnya adalah tahap *preprocessing*, yang merupakan tahap awal dalam analisis data teks untuk membersihkan dan mempersiapkan data. Langkah-langkah *preprocessing* meliputi *cleansing* untuk menghapus elemen yang tidak relevan, *case folding* untuk menyamakan teks menjadi huruf kecil, dan *tokenization* untuk memisahkan teks menjadi kata-kata. Selanjutnya, *stopword removal* menghapus kata-kata umum yang tidak bermakna, *normalisasi* menyetarakan variasi kata, dan *stemming* mengubah kata menjadi bentuk dasarnya. Semua langkah ini memastikan data teks lebih bersih dan siap untuk analisis lebih lanjut. Hasil *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Preprocessing*

Tahapan	Tweet
<i>Tweet</i>	Pertama kali nyoba <i>starlink</i> punya om @elonmusk. Ternyata emg <i>amazing</i> ya <i>speed</i> sama pingnya untuk sekelas internet satelit. Gokill https://t.co/T8kf7DGdbm .
<i>Cleansing</i>	Pertama kali nyoba <i>starlink</i> punya om elonmusk Ternyata emg <i>amazing</i> ya <i>speed</i> sama pingnya untuk sekelas internet satelit Gokill
<i>Case Folding</i>	pertama kali nyoba <i>starlink</i> punya om elonmusk ternyata emg <i>amazing</i> ya <i>speed</i> sama pingnya untuk sekelas internet satelit gokill
<i>Tokenizing</i>	['pertama', 'kali', 'nyoba', 'starlink', 'punya', 'om', 'elonmusk', 'ternyata', 'emg', 'amazing', 'ya', 'speed', 'sama', 'pingnya', 'untuk', 'sekelas', 'internet', 'satelit', 'gokill']
<i>Stopword</i>	['kali', 'nyoba', 'starlink', 'om', 'elonmusk', 'emg', 'amazing', 'ya', 'speed', 'pingnya', 'sekelas', 'internet', 'satelit', 'gokill']
<i>Normalisasi</i>	['kali', 'mencoba', 'starlink', 'om', 'elonmusk', 'emang', 'menakjubkan', 'ya', 'kecepatan', 'pingnya', 'sekelas', 'internet', 'satelit', 'gokill']
<i>Stemming</i>	kali coba <i>starlink</i> om elon musk emang takjub ya cepat ping kelas internet satelit gokil

Berdasarkan Tabel 2 yang menunjukkan hasil *preprocessing*, proses *preprocessing* dimulai dengan *cleansing*, yang menghapus elemen tak relevan seperti URL. Misalnya, teks "Ternyata emg *amazing* ya *speed* sama pingnya untuk sekelas internet satelit. Gokill <https://t.co/T8kf7DGdbm>." diubah menjadi "Ternyata emg *amazing* ya *speed* sama pingnya untuk sekelas internet satelit Gokill". *Case folding* mengubah semua huruf menjadi kecil untuk konsistensi, seperti "Gokill" menjadi "gokill". Pada *tokenizing*, teks dipecah menjadi kata-kata individual, misalnya "Ternyata emg *amazing*" menjadi ['ternyata', 'emg', 'amazing']. *Stopword removal* menghapus kata-kata umum yang tidak bermakna, seperti "kali" dan "ya", meninggalkan kata-kata relevan. *Normalisasi* mengubah variasi kata tidak baku, seperti "nyoba" menjadi "mencoba", ke bentuk baku. Terakhir, *stemming* menyederhanakan kata-kata menjadi

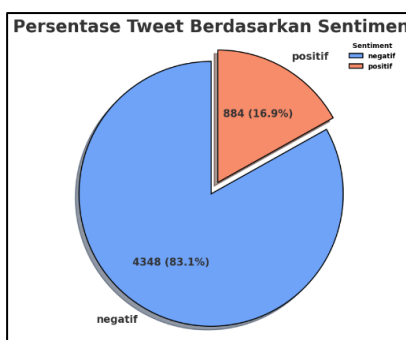
bentuk dasarnya, seperti "mencoba" menjadi "coba". Semua tahapan ini membersihkan, menyelaraskan, dan menyederhanakan teks agar siap untuk analisis lebih lanjut.

3.3 Pelabelan Data

Tahap selanjutnya adalah pelabelan data untuk menentukan apakah teks memiliki sentimen positif atau negatif. Proses ini menggunakan pustaka *Python* seperti *TextBlob*, yang menilai sentimen berdasarkan polaritas dalam rentang -1 hingga 1. Ambang batas 0 dipilih karena secara alami memisahkan sentimen positif dan negatif; nilai di atas 0 menunjukkan opini positif, sedangkan nilai di bawah 0 menunjukkan opini negatif. Namun, menggunakan polaritas sebagai satu-satunya indikator memiliki keterbatasan, seperti kesulitan mengenali konteks, ironi, atau sarkasme. Selain itu, teks dengan polaritas mendekati 0 cenderung bersifat netral, tetapi tetap diklasifikasikan sebagai positif atau negatif, yang dapat memengaruhi akurasi analisis. Hasil pelabelan data dapat dilihat pada Tabel 8.

Tabel 8. Hasil Proses Pelabelan

No	Username	Tweet	Sentiment Score	Label
1	_rzkymln	Pertama kali nyoba starlink punya om @elonmusk. Ternyata emg amazing ya speed sama pingnya untuk sekelas internet satelit. Gokill https://t.co/T8kf7DGdbm	0.6	Positif
5.382	Juniorama_	Suku pedalaman di Amazon akhirnya terhubung dengan internet berkat Starlink Elon Musk. Tapi para tetua protes karena anak muda mengakses pornografi. Duhhh... https://t.co/2N5CHsC8Ke	-0.5	Negatif

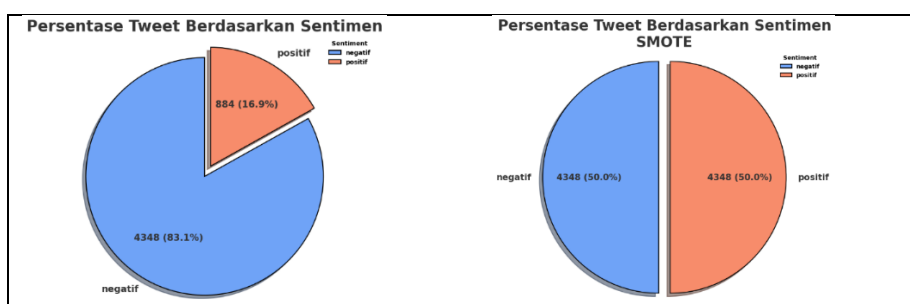


Gambar 2. Hasil Pelabelan

Berdasarkan Gambar 2 yang menunjukkan Analisis distribusi sentimen berdasarkan diagram *pie* menunjukkan adanya ketidakseimbangan yang signifikan dalam dataset. Sentimen positif hanya mencakup 884 *tweet* atau 16,9% dari total data, sedangkan sentimen negatif mendominasi dengan jumlah 4.348 *tweet* atau 83,1%. Ketimpangan proporsi ini mengindikasikan bahwa dataset tidak seimbang, dengan jumlah data sentimen negatif yang jauh lebih besar dibandingkan sentimen positif.

3.4 Teknik SMOTE

Berdasarkan hasil pelabelan awal, distribusi sentimen pada dataset menunjukkan ketidakseimbangan yang mencolok. Sentimen positif hanya terdiri atas 884 *tweet* (16,9%), sementara sentimen negatif mendominasi dengan 4.348 *tweet* (83,1%). Ketidakseimbangan ini mengindikasikan distribusi data yang tidak proporsional, di mana kelas negatif jauh lebih besar dibandingkan kelas positif. Untuk mengatasi masalah ini, diterapkan teknik *oversampling*, yaitu *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini menambahkan data sintesis pada kelas *minoritas* (positif) untuk menyamakan proporsi dengan kelas *majoritas*. Setelah penerapan SMOTE, distribusi data menjadi seimbang, dengan jumlah *tweet* positif dan negatif masing-masing mencapai 4.348 (50%).



Gambar 3. Hasil Optimasi SMOTE

3.5 Klasifikasi NBC, SVM, Random Forest

Pada penelitian ini, dilakukan perbandingan kinerja model klasifikasi sebelum dan sesudah penerapan teknik SMOTE (*Synthetic Minority Oversampling Technique*) untuk mengatasi ketidakseimbangan kelas. Model yang diuji meliputi *Naive Bayes Classifier*, *Support Vector Machine*, dan *Random Forest*, yang diukur menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Data awal memiliki distribusi kelas yang tidak seimbang, di mana jumlah data untuk kelas positif jauh lebih sedikit dibandingkan kelas negatif. Hasil klasifikasi disajikan pada Tabel 8 dan Tabel 9.

Tabel 9. Hasil Klasifikasi Before SMOTE

Model	Class	Accuracy	Precision	Recall	F1-Score
NBC	Negatif	85%	85%	100%	92%
	Positif		91%	16%	28%
SVM	Negatif	91%	91%	99%	95%
	Positif		91%	56%	69%
Random Forest	Negatif	92%	92%	100%	95%
	Positif		96%	57%	71%

Berdasarkan Tabel 8 menunjukkan hasil klasifikasi Sebelum penerapan SMOTE, akurasi masing-masing model adalah sebagai berikut NBC sebesar 85%, SVM sebesar 91%, dan *Random Forest* sebesar 92%. Namun, hasil klasifikasi menunjukkan ketidakseimbangan yang signifikan antara kelas negatif dan positif. Semua model cenderung memiliki performa yang jauh lebih baik dalam mengklasifikasikan kelas negatif dibandingkan kelas positif. *Naive Bayes Classifier* (NBC) mencatatkan performa terburuk dalam mendeteksi kelas positif, dengan *Precision* 16% dan *Recall* 28%, menunjukkan kesulitan dalam mengidentifikasi kelas *minoritas*. Sementara itu, *Random Forest* mencatatkan performa terbaik di antara ketiga model dengan akurasi tertinggi, meskipun deteksi pada kelas positif masih rendah, sebagaimana terlihat dari *F1-Score* 71%.

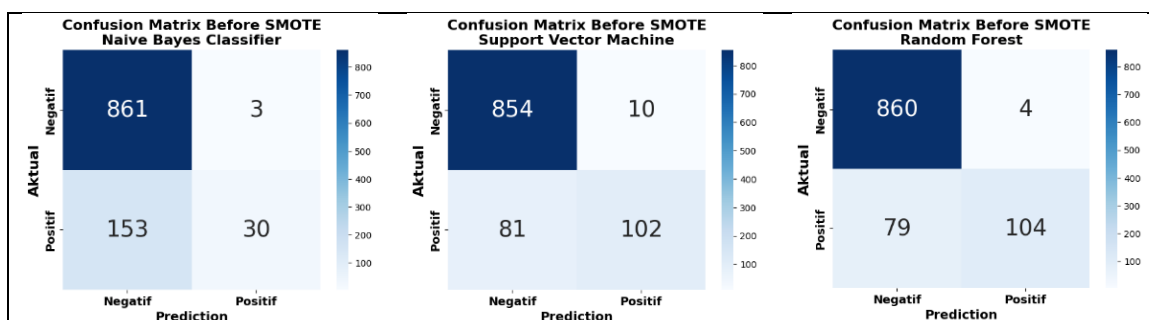
Tabel 9. Hasil Klasifikasi After SMOTE

Model	Class	Accuracy	Precision	Recall	F1-Score
NBC	Negatif	91%	93%	89%	91%
	Positif		89%	94%	91%
SVM	Negatif	98%	98%	97%	98%
	Positif		97%	98%	98%
Random Forest	Negatif	99%	100%	98%	99%
	Positif		98%	100%	99%

Berdasarkan Tabel 9 yang menunjukkan hasil Setelah penerapan SMOTE, terjadi peningkatan signifikan pada kinerja model dalam mendeteksi kelas positif. Pada *Naive Bayes Classifier* (NBC), akurasi meningkat menjadi 91%, dengan peningkatan pada kelas positif mencapai *Precision* 94% dan *Recall* 91%. Model *Support Vector Machine* (SVM) menunjukkan kinerja yang sangat baik dengan akurasi 98%, serta *Precision* dan *Recall* untuk kelas positif masing-masing 98%. *Random Forest* menjadi model terbaik dengan akurasi tertinggi 99%, *Precision* 100%, dan *Recall* 100% pada kelas positif, menunjukkan performa yang sangat baik dalam menangani ketidakseimbangan data. Secara keseluruhan, penerapan SMOTE berhasil meningkatkan kemampuan semua model dalam mendeteksi kelas positif, mengurangi ketidakseimbangan data, dan memperbaiki performa model secara signifikan.

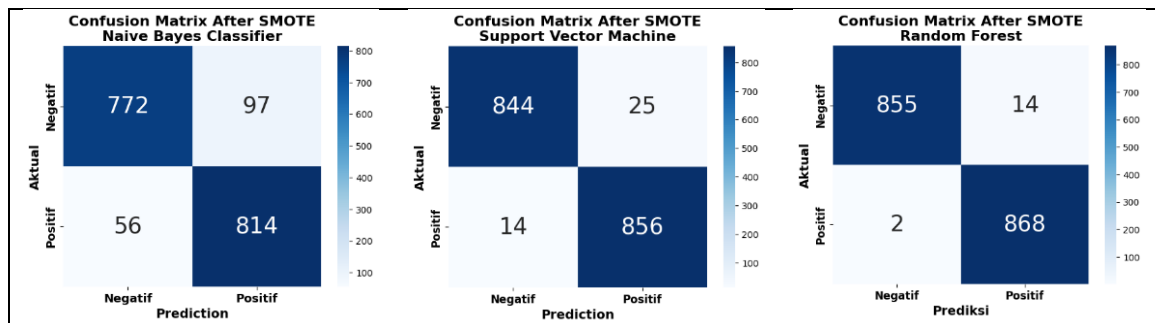
3.6 Evaluasi

Setelah melakukan klasifikasi dengan *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Random Forest*, evaluasi kinerja model dilakukan menggunakan *Confusion Matrix*. Evaluasi kinerja model dilakukan menggunakan *Confusion Matrix* untuk mengukur *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang digunakan untuk menghitung metrik seperti *Precision*, *Recall*, dan *F1-Score*.



Gambar 4. Confusion Matrix Before SMOTE

Berdasarkan Gambar 4 yang menunjukkan hasil evaluasi menggunakan *confusion matrix* Sebelum penerapan SMOTE, hasil evaluasi menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki *True Positive* (TP) sebanyak 102, *True Negative* (TN) sebanyak 854, *False Positive* (FP) sebanyak 10, dan *False Negative* (FN) sebanyak 81, menunjukkan kelemahan dalam mendeteksi kelas positif. *Random Forest* menunjukkan kinerja terbaik dengan TP sebanyak 104, TN sebanyak 860, FP sebanyak 4, dan FN sebanyak 79, dengan kesalahan prediksi yang paling rendah. Sementara itu, *Naive Bayes Classifier* (NBC) memiliki TP sebanyak 30, TN sebanyak 861, FP sebanyak 3, dan FN sebanyak 153, yang menunjukkan kesulitan signifikan dalam mendeteksi kelas positif.



Gambar 5. *Confusion Matrix After SMOTE*

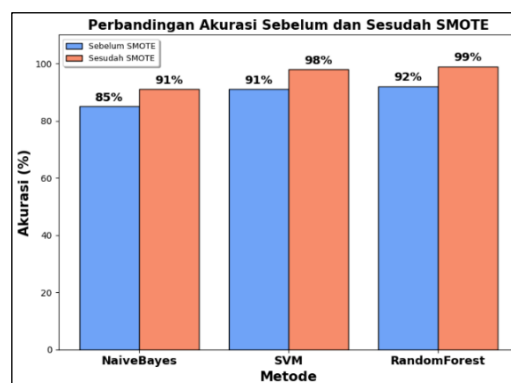
Berdasarkan Gambar 5 setelah penerapan SMOTE, algoritma *Random Forest* menunjukkan kinerja terbaik dengan TP sebanyak 868, TN sebanyak 855, FP sebanyak 14, dan FN sebanyak 2, yang menghasilkan jumlah kesalahan prediksi paling rendah. *Support Vector Machine* (SVM) memiliki TP sebanyak 856, TN sebanyak 844, FP sebanyak 25, dan FN sebanyak 14, menunjukkan kinerja yang baik meskipun ada lebih banyak kesalahan prediksi dibandingkan dengan *Random Forest*. *Naive Bayes Classifier* (NBC) menunjukkan kinerja terburuk dengan TP sebanyak 814, TN sebanyak 772, FP sebanyak 97, dan FN sebanyak 56, meskipun terdapat peningkatan dibandingkan sebelum penerapan SMOTE, sedangkan *Naive Bayes* memiliki performa yang paling lemah. Secara keseluruhan, *Random Forest* menunjukkan kinerja terbaik dalam menangani ketidakseimbangan kelas, baik sebelum maupun setelah penerapan SMOTE.

3.7 Visualisasi

Setelah melalui tahap klasifikasi dan evaluasi model, selanjutnya adalah tahap visualisasi. Pada tahap ini, beberapa metode digunakan untuk menggambarkan hasil analisis data teks. Antaranya adalah sebagai berikut :

3.7.1 Perbandingan Akurasi

Visualisasi perbandingan adalah alat yang berguna untuk menggambarkan perbedaan atau perubahan antara beberapa kelompok data atau kondisi tertentu, sehingga mempermudah pemahaman dan pengambilan keputusan. Dalam konteks analisis data, visualisasi sering digunakan untuk menyoroti dampak dari suatu perubahan atau intervensi, seperti penerapan teknik tertentu pada data. Hasil perbandingan akurasi sebelum dan sesudah smote dapat dilihat pada Gambar 6.



Gambar 6. Perbandingan Klasifikasi

Gambar 6 menunjukkan perbandingan klasifikasi, hasil penerapan SMOTE secara konsisten meningkatkan akurasi untuk semua metode pembelajaran mesin. Pada metode *Naive Bayes*, akurasi meningkat dari 85% menjadi 91%, sedangkan pada metode SVM, akurasi naik dari 91% menjadi 98%. Sementara itu, metode *Random Forest* menunjukkan peningkatan akurasi dari 92% menjadi 99%, sekaligus mencatat akurasi tertinggi baik sebelum maupun setelah SMOTE diterapkan.

Gambar 9 menunjukkan bahwa kata "starlink" memiliki frekuensi tertinggi, muncul sebanyak 5965 kali, menandakan fokus utama ulasan adalah pada layanan ini. Kata "internet" berada di posisi kedua dengan 3167 kali, yang mengindikasikan banyak ulasan terkait konektivitas atau akses internet. Selanjutnya, kata "indonesia" muncul sebanyak 2916 kali, menunjukkan bahwa banyak ulasan berasal dari pengguna di Indonesia. Kata-kata lainnya, seperti "elon", "satelit", "musk", "masuk", "layan", "cepat", dan "pakai", memiliki frekuensi yang lebih rendah tetapi tetap signifikan. Kesimpulannya, grafik ini menggambarkan bahwa ulasan terkait layanan Starlink terutama berfokus pada akses internet dan pengguna di Indonesia. Selain itu, kata-kata seperti "elon", "musk", dan "satelit" menunjukkan adanya kaitan ulasan dengan figur pendiri (Elon Musk) dan teknologi yang digunakan (satelit).

4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa penerapan *Synthetic Minority Oversampling Technique* (SMOTE) berhasil mengatasi ketidakseimbangan distribusi sentimen dalam dataset X yang cenderung didominasi oleh sentimen negatif. Penelitian ini membandingkan kinerja tiga model klasifikasi, yaitu *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Random Forest*, sebelum dan setelah penerapan SMOTE. Evaluasi dilakukan dengan membagi data menjadi 80% untuk *training* dan 20% untuk *testing*. Sebelum penerapan SMOTE, ketiga model menunjukkan ketidakseimbangan kinerja, dengan akurasi *Naive Bayes* sebesar 85%, SVM 91%, dan *Random Forest* 92%. Model lebih unggul dalam mengklasifikasikan kelas negatif, sebagaimana ditunjukkan oleh nilai *Precision*, *Recall*, dan *F1-Score* yang lebih tinggi pada kelas tersebut. Setelah penerapan SMOTE, akurasi meningkat menjadi 91% untuk *Naive Bayes*, 98% untuk SVM, dan 99% untuk *Random Forest*, dengan peningkatan signifikan pada metrik *Precision*, *Recall*, dan *F1-Score* untuk kelas positif. Model *Random Forest* menunjukkan hasil terbaik dengan *Precision* dan *Recall* mencapai 100%, menegaskan bahwa SMOTE efektif dalam meningkatkan performa model, terutama dalam mendeteksi kelas positif yang sebelumnya kurang seimbang. Dibandingkan dengan penelitian sebelumnya oleh Bebin Paula dkk. (2024), Sardin dkk. (2024), dan M. Khalil Gibran dkk. (2024), yang menganalisis sentimen terhadap *Starlink* menggunakan *Naive Bayes* dan SVM tanpa SMOTE, penelitian ini menunjukkan hasil yang lebih baik karena membandingkan tiga algoritma dengan jumlah data yang lebih besar serta menerapkan teknik SMOTE. Namun, penelitian ini memiliki beberapa keterbatasan, terutama dalam cakupan dataset yang terbatas pada platform X (*Twitter*). Untuk meningkatkan generalisasi temuan, penelitian selanjutnya disarankan untuk memperluas cakupan data dengan mencakup platform lain seperti *Facebook*, *Instagram*, atau bahkan platform dengan karakteristik bahasa dan gaya komunikasi yang berbeda. Selain itu, eksplorasi teknik penyeimbangan data lainnya perlu dilakukan guna mengurangi potensi *overfitting*. Penelitian mendatang juga dapat bereksperimen dengan model *deep learning*, seperti LSTM atau BERT, yang lebih canggih dalam menangani data teks dan berpotensi meningkatkan akurasi dalam analisis sentimen.

REFERENCES

- [1] M. R. Julianto, Y. Akbar, and T. Wahyudi, "Analisis Sentimen Respon Publik Terhadap Program Internet Gratis di Platform X Melalui Pendekatan Algoritma Naive Bayes," *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, vol. 5, no. 3, pp. 2940–2950, 2024. doi.org/10.35870/jimik.v5i3.981.
- [2] Lisnawati, "Kehadiran Starlink di Indonesia : Manfaat dan Dampak," *Info Singkat : Kajian Singkat Terhadap isu Aktual dan Strategies*, vol. 16, no. 11, pp. 16–20, 2024.
- [3] M. Ridwan *et al.*, "Oligopoli Telekomunikasi dan Inovasi : Analisis Dampak Masuknya STARLINK bagi Industri Telekomunikasi di Indonesia," *Jurnal Ilmiah Ekonomi Dan Manajemen*, vol. 2, no. 12, pp. 306–312, 2024. doi.org/10.61722/jiem.v2i12.3206.
- [4] A. Nugroho and N. T. Kurniadi, "Sentiment Analysis of Starlink on Twitter Using Support Vector Machine Algorithm," *Journal of Computer Networks , Architecture and High Performance Computing*, vol. 6, no. 3, pp. 1321–1332, 2024. doi.org/10.47709/cnahpc.v6i3.4348
- [5] M. Y. Khaliq, M. Rakib, and V. Aris, "Sentiment Analysis to Starlink Services in Indonesia," *International Journal of Latest Technology in Engineering & Management*, vol. 9, no. 6, pp. 13–21, 2024. doi: 10.56581/ijltem.9.6.02-12.
- [6] C. Candra, K. W. Chandra, and H. Irsyad, "Efektifitas SMOTE dalam Mengatasi Imbalanced Class Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 31–42, 2024, doi: 10.54082/jiki.132.
- [7] B. Paula, M. Fawzan, and H. Irsyad, "Analisis Sentiment Masyarakat Terhadap penyebaran Starlink di Indonesia Menggunakan Algoritma Naive Bayes," *Journal Information & Computer*, vol. 02, no. 2, pp. 141–148, 2024. doi.org/10.32493/jicomisc.v2i2.43077.
- [8] M. Khalil Gibran, Rifki, M. I., Hasugian, A. H., Siahaan, A. T. A. A., Afandi Sahputra, & Ong, R., "Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes," *Instal : Jurnal Komputer*, vol. 16, no. 03, pp. 210–220, 2024, doi.org/10.54209/jurnalinstall.v16i03.240
- [9] M. Nur Akbar, N. A. S. Yusuf, N. Nasrullah, and M. Mubarak, "Analisis Sentimen Pengguna Indihome dengan Metode Klasifikasi Support Vector Machine (SVM)," *Journal Software, Hardware and Information Technology*, vol. 2, no. 1, pp. 13–21, 2022, doi: 10.24252/shift.v2i1.18.
- [10] Lady Agustine Fitriana, S. Linawati, N. Herlinawati, R. Sa'adah, and S. Seimahuria, "Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naive Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 4291–4297, 2024, doi: 10.36040/jati.v8i3.9866.



- [11] F. J. Fauzan, M Afdal, Rice Novita, and Mustakim, “Penerapan Machine Learning Pada Analisis Sentimen Aplikasi Mytelkomsel Menggunakan Data Ulasan Google Playstore,” *Indonesian Journal of Computer Science*, vol. 13, no. 3, pp. 4747–4761, 2024, doi: 10.33022/ijcs.v13i3.4024.
- [12] H. Handayani, “Analisis Sentimen Pengguna X Terhadap Layanan Provider Iconnet Menggunakan Naïve Bayes Dan Support Vector Machine,” *Artificial Intelligence and Mechatronics Systems (AIMS)*, vol. 12, no. 112, pp. 33–38, 2024, doi:10.1109/aims61812.2024.10512562
- [13] P. D. Aprilia and S. Lestari, “Analisa Sentimen Drama Korea Melalui Media Sosial X dengan Menggunakan Algoritma Naïve Bayes,” *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, vol. 5, no. 3, pp. 3248–3261, 2024, doi.org/10.35870/jimik.v5i3.997.
- [14] P. Cahyani and L. Abdillah, “Perbandingan Performa Algoritma Naïve Bayes , SVM dan Random Forest : Studi Kasus Analisis Sentimen Pengguna Sosial Media X,” *Kalbiscientia Jurnal Sains Dan Teknologi*, vol. 11, no. 02, pp. 12–21, 2024, doi.org/10.53008/kalbiscientia.v11i02.3624.
- [15] A. Rhamadanti, A. Rifa’i, F. Dikananda, and K. Anam, “Analisis Sentimen Pada Ulasan Access By Kereta Api Indonesia Dengan K-Nearest Neighbor,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3961.
- [16] P. Wahyuningtias, H. Warih Utami, U. Ahda Raihan, H. Nur Hanifah, and Y. Nicholas Adanson, “Comparison of Random Forest and Support Vector Machine Methods on Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Venny Escape From Quarantine),” *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 1, pp. 141–145, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.1.168>
- [17] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [18] L. Rofiqi and M. Akbar, “Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine,” *JEKIN - Jurnal Teknik Informatika*, vol. 4, no. 3, pp. 529–538, 2024, doi: 10.58794/jekin.v4i3.824.
- [19] D. N. Herisnan and M. Elwinda, “Analisis sentimen terhadap resesi ekonomi global di indonesia menggunakan hybrid linear regression,” *INTECOMS: Journal of Information Technology and Computer Science*, vol. 7, pp. 1495–1501, 2024., Doi:10.31539/Intecom.v7i5.11889.
- [20] D. Toresa, S. Rico Francisco Sitorus, I. Muzdalifah, F. Wiza, and R. Syelly, “Analisis Sentimen Terhadap Ulasan Penggunaan Dompot Digital Dana Menggunakan Metode Klasifikasi Support Vector Machine,” *Jurnal Technologica*, vol. 3, no. 2, pp. 64–74, 2024, doi: 10.55043/technologica.v3i2.163.
- [21] R. Nurlaely, S. D. Sartika, Kamdan, and I. L. Kharisma, “Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM),” *Jurnal Computer Science and Information Technology(CoSciTech)*, vol. 4, no. 2, pp. 376–384, 2023, doi.org/10.37859/coscitech.v4i2.5161
- [22] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings of Engineering*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [23] A. Saepudin, A. Faqih, and G. Dwilestari, “Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee,” *Jurnal Tekno Kompak*, vol. 18, no. 1, p. 178, 2024, doi: 10.33365/jtk.v18i1.3764.
- [24] A. Baita, Y. Pristyanto, and N. Cahyono, “Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (Svm) Dan K-Nearest Neighbor (Knn),” *Information System Journal*, vol. 4, no. 2, pp. 42–42, 2021. doi.org/10.24076/infosjournal.2021v4i2.687.
- [25] H. Azizah, B. S. Rintyarna, and T. A. Cahyanto, “Sentimen Analisis Untuk Mengukur Kepercayaan Masyarakat Terhadap Pengadaan Vaksin Covid-19 Berbasis Bernoulli Naive Bayes,” *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 3, no. 1, pp. 23–29, 2022, doi: 10.37148/bios.v3i1.36.
- [26] 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, 2023, doi: 10.57152/malcom.v3i2.897.
- [27] L. Annisa, A. D. Kalifia, F. Bisnis, D. Humaniora, and U. T. Yogyakarta, “Analisis Teknik TF-IDF Dalam Identifikasi Faktor-Faktor Penyebab Depresi Pada Individu,” *Gudang Jurnal Multidisiplin Ilmu*, vol. 2, pp. 302–307, 2024. <https://doi.org/10.59435/gjmi.v2i1.249>.
- [28] I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, “Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen,” *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, pp. 145–149, 2022. [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/349>
- [29] R. A. Raharjo, I. M. G. Sunarya, and D. G. H. Divayana, “Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter,” *Elkom : Jurnal Elektronika dan Komputer*, vol. 15, no. 2, pp. 456–464, 2022, doi: 10.51903/elkom.v15i2.918.
- [30] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *Jurnal KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [31] R. Afandi, M. Afdal, and R. Novita, “Analisis Sentimen Masyarakat Terhadap Pinjaman Online di Twitter Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 596–605, 2024, doi: 10.47065/bits.v6i2.5300.
- [32] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [33] A. Ramadhani, I. Permana, M. Afdal, and M. Fronita, “Analisis Sentimen Tanggapan Publik di Twitter Terkait Program Kerja Makan Siang Gratis Prabowo – Gibran Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector



- Machine,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1509–1516, 2024, doi: 10.47065/bits.v6i3.6188.
- [34] I. Afdhal, R. Kurniawan, I. Iskandar, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- [35] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- [36] E. Suryati, Styawati, and A. A. Aldino, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Dan Sistem Informasi*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>