ANALISIS SENTIMEN PUBLIK PADA PEMERINTAH DALAM SERANGAN RANSOMWARE DENGAN PENDEKATAN SMOTE

Mohammad Indra Prayugah[1)], Uce Indahyanti[2)], Novia Ariyanti[3)]
[1-3]Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Muhammadiyah Sidoarjo, Jl. Raya Gelam No.250, Pagerwaja, gelam, kec. Candi, kabupaten Sidoarjo, jawa Timur
email: [1]211080200115@umsida.ac.id, [2]uceindahyanti@umsida.ac.id, [3]noviaariyanti@umsida.ac.id

Abstract

The ransomware attack on Indonesia's national data center has become a widely discussed topic in society. YouTube has become the main platform for disseminating information and giving people opinions. This research aims to identify public sentiment regarding the government's handling of ransomware attacks through analysis of comments on the CNN Indonesia and MetroTV YouTube channels. Data was collected using web scraping techniques and entered into a classification model with three labels positive, neutral and negative sentiment. The three machine learning models that will be used are SVM, Random Forest, and Naïve Bayes, with two test scenarios, using Synthetic Minority Over-sampling Technique (SMOTE) and without SMOTE. Applying SMOTE increases model accuracy, especially in SVM which reaches 96%. The research results show that the majority of comments express negative sentiments towards the government's performance. This research will provide an understanding of public perceptions of cyber security issues in Indonesia and the effectiveness of SMOTE in sentiment analysis.

Keywords: *ransomware, support vector machine, random forest, naïve bayes, SMOTE*

Abstrak

Serangan ransomware pada pusat data nasional Indonesia menjadi topik yang banyak dibicarakan di masyarakat. YouTube menjadi platform utama untuk menyebarkan informasi dan masyarakat beropini. Penelitian ini bertujuan untuk mengidentifikasi sentimen publik mengenai penanganan pemerintah terhadap serangan ransomware melalui analisis komentar di kanal YouTube CNN Indonesia dan MetroTV. Data dikumpulkan menggunakan teknik web scraping dan dimasukkan ke dalam model klasifikasi dengan tiga label yaitu sentimen positif, netral, dan negatif. Tiga model machine learning yang akan digunakan adalah SVM, Random Forest, dan Naïve Bayes, dengan dua skenario pengujian yaitu menggunakan Synthetic Minority Over-sampling Technique (SMOTE) dan tanpa SMOTE. Penerapan SMOTE meningkatkan akurasi model, terutama pada SVM yang mencapai 96%. Hasil penelitian menunjukkan bahwa mayoritas komentar mengungkapkan sentimen negatif terhadap kinerja pemerintah. Penelitian ini akan memberikan pemahaman mengenai persepsi publik terhadap isu keamanan siber di Indonesia dan efektivitas SMOTE dalam analisis sentimen.

Kata Kunci: *ransomware, support vector machine, random forest, naïve bayes, SMOTE*

1. PENDAHULUAN

Teknologi informasi yang berkembang dengan cepat telah mengubah cara masyarakat dalam menyampaikan opini atau pendapat mereka. Salah satu platform yang sering digunakan masyarakat untuk beropini adalah *Youtube*. *YouTube* adalah situs berbagi media semacam hiburan *virtual* untuk berbagi media video dan suara [1]. Youtube sebagai platform berbagi video dengan pengguna aktif lebih dari jutaan setiap harinya menjadi tempat dimana berbagai informasi beredar. *Youtube* menyediakan fitur komentar untuk pengguna bisa memberi tanggapan ataupun opini [2]. Opini yang diberikan oleh para pengguna *YouTube* seringkali digunakan untuk membuat penilaian yang merujuk pada suatu topik atau tujuan tertentu dikenal dengan istilah analisis sentimen [3].

Baru-baru ini beredar informasi mengenai serangan *ransomware* oleh *hacker* pada pusat data nasional yang berdampak ke masyarakat dan pemerintah. *YouTube* berperan sebagai salah satu media utama tempat informasi beredar dan masyarakat menyuarakan opini melalui komentar pada video

berita terkait insiden tersebut. Sebagai lembaga yang bertugas melayani masyarakat, instansi pemerintah sangat mengandalkan opini publik untuk melakukan evaluasi dan meningkatkan kinerja. Salah satu contohnya adalah Kementerian Komunikasi dan Informatika (Kominfo), yang memiliki tanggung jawab menjaga keamanan data. Serangan ini memunculkan berbagai persepsi dari masyarakat terkait langkah-langkah yang diambil oleh Kominfo dalam menangani situasi tersebut. Oleh karena itu, analisis sentimen menjadi penting untuk memahami reaksi dan persepsi publik terhadap kinerja pemerintah, yang pada gilirannya dapat menjadi bahan evaluasi dan peningkatan layanan publik di masa depan.

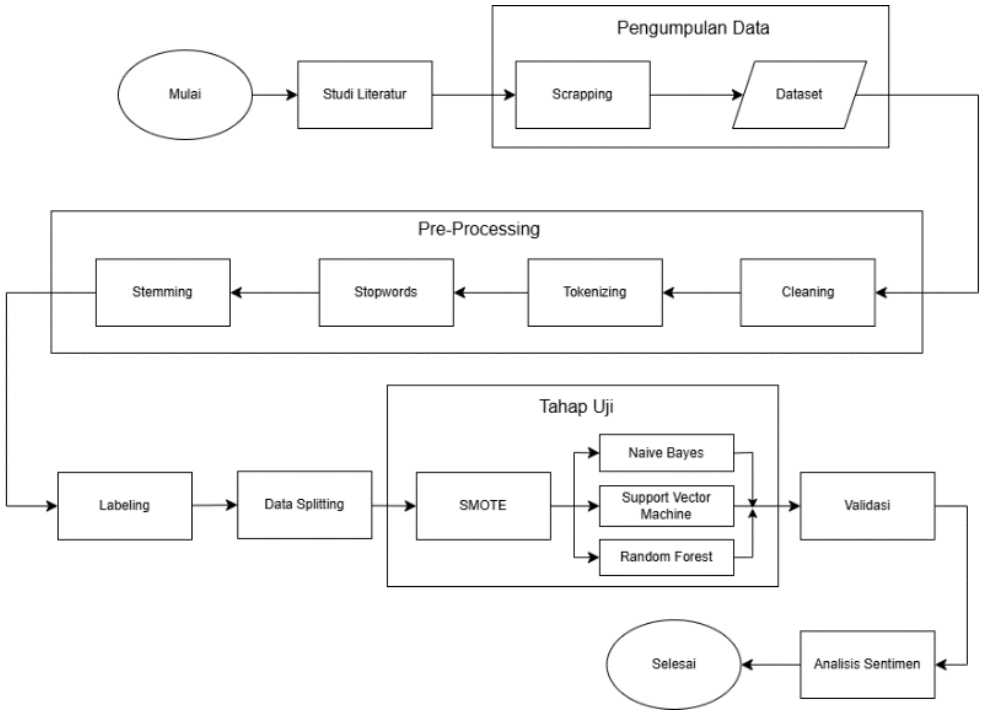
Analisis sentimen publik melalui komentar *YouTube* menjadi informasi langsung tentang persepsi masyarakat terhadap kinerja instansi pemerintah. Analisis sentimen adalah kegiatan untuk menganalisis opini tentang suatu topik [4]. Analisis sentimen secara otomatis mengidentifikasi, mengekstrak, dan mengolah informasi dalam bentuk teks untuk menghasilkan data mengenai sentiment [5]. Analisa sentimen dapat dilakukan dengan beberapa tahap dimulai dari pengumpulan data (*scrapping data*), dilanjutkan tahap *pre-processing* data dan kemudian masuk ke tahap analisa data menggunakan beberapa algoritma *machine learning*.

Penelitian sebelumnya yang dilakukan Farid Naufal pernah melakukan perbandingan dari tiga metode algoritma tentang analisis *cyberbullying* pada media sosial. Peneliti membandingkan antara Support Vector Machine, Random Forest, dan Naive Bayers setelah dilakukan pre-processing tanpa dilakukan oversampling SMOTE. Oleh karena itu, penelitian ini ditujukan untuk mengembangkan penelitian yang sebelumnya dilakukan Farid Naufal dengan membandingkan 3 metode algoritma Support Vector Machine, Random Forest, dan Naive Bayers tapi dengan penambahan oversampling SMOTE [6]. Penelitian serupa juga pernah dilakukan oleh putri yang dimana penelitian tersebut membahas tentang analisis data dengan tahap pre-processing pada kometar tentang pelayanan masyarakat DKI Jakarta dan dilakukan pembandingan antara dua metode algoritma Naïve Bayes dan SVM. Didapati hasil dari SVM mengungguli sedikit lebih akurat dari Naïve Bayes. Tetapi dataset yang dipakai dalam penelitian hanya bersumber dari 3 video membuat penelitian terkesan tidak *representative* [7]. Penelitian oleh Oktaviami Manullang membandingkan metode algoritma dari random forest dengan lexicon based untuk prediksi hasil calon pemilu presiden berdasarkan data twitter. Dalam penelitian tersebut didapat hasil bahwa *Random Forest* memberikan hasil yang akurat dibanding dengan Lexicon Based yang kurang efektif dalam menangani analisi sentimen kalimat ambigu [8]. Penelitian Muhammad Yasir membandingkan model *Naïve bayes, Decision Tree, Random Forest* tentang analisis sentimen kenaikan biaya haji 2023 dan didapati model *Naïve Bayes* memberikan hasil terbaik [9]. Dalam Penelitian oleh Syah dibahas tentang analisis sentimen twitter terhadap *indihomecare* dengan melakukan uji memakai metode algoritma SVM. Data yang diuji dilakukan beberapa skenario dengan membandingkan hasil menggunakan SMOTE, AdaBoost, PSO. Didapati hasil penggunaan SMOTE memberikan hasil yang lebih unggul [10].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kinerja pemerintah dalam menangani kasus serangan ransomware pada pusat data nasional. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi pengaruh penggunaan teknik SMOTE (Synthetic Minority Oversampling Technique) dalam meningkatkan akurasi analisis sentimen. Untuk mencapai tujuan tersebut, beberapa model machine learning seperti Support Vector Machine, Random Forest, dan Naïve Bayes digunakan, baik dengan maupun tanpa penerapan SMOTE. Hasil penelitian diharapkan menjadi bahan evaluasi kinerja pemerintah dalam penanganan kasus serupa di masa depan.

2. METODE PENELITIAN

Penelitian dilakukan melalui beberapa tahap sesuai dengan yang ditunjukkan pada Gambar 1. Metode penelitian dimulai dari tahap Pengumpulan Data, Pre-Processing, Tahap Uji, hingga Validasi. Dalam penelitian ini akan dilakukan menggunakan bahasa python. Python adalah bahasa pemrograman interpretatif yang sederhana untuk dipelajari, dapat digunakan di berbagai platform, dan dirancang dengan penekanan pada keterbacaan kode [11]. Program akan dijalankan melalui *Google Colab*. *Google Colab* merupakan platform cloud untuk menulis, menjalankan, dan membagikan kode python langsung melalui web browser [4].



Gambar 1 Flowchart Alur Penelitian

2.1 Studi Literatur

Studi literatur dilakukan untuk memenuhi referensi penelitian dari beberapa sumber mulai dari buku, jurnal, hingga video dengan tujuan menguatkan pemahaman mendasar penelitian. Proses ini memungkinkan peneliti untuk menggali informasi yang relevan untuk meninjau literatur sebelumnya [12].

2.2 Pengumpulan Data

Data dikumpulkan dengan mengambil komentar video berita dari *YouTube* tentang serangan *ransomware* pada pusat data nasional Indonesia dari channel berita tertentu seperti CNN dan MetroTV dengan batas waktu antara 20 Juni 2024 hingga 30 Agustus 2024. Data dikumpulkan melalui metode scraping menggunakan YouTube Data APIv3 yang diakses melalui Google Developers Console. Di dalam Google Developers Console, terdapat fitur APIs & Services yang perlu diaktifkan terlebih dahulu. Setelah itu, pengguna akan diarahkan ke API Library untuk mencari YouTube Data APIv3 dan membuat kredensial baru. Proses ini menghasilkan API Key yang digunakan untuk mengakses data dari YouTube [1]. API Key tersebut yang akan digunakan untuk mendapatkan data dengan akses dari id setiap video. Pertama dilakukan pengumpulan sumber video yang akan dipakai sesuai dengan kebutuhan penelitian. Kemudian diambil id dari setiap video untuk dilakukan proses scrapping data menggunakan Google Colab. Data yang diambil meliputi komentar video saja yang kemudian akan diubah ke dalam format CSV dengan berisi 6484 data.

Tabel 1 Dataset

no	text
1	Masak gak ada back up
2	pas make w7 pernah kena ransome
3	Seharusnya semua pejabat dinegeri ini dipilih…
4	Gk mau berinovasi dan hny mau terima bersih..
5	Kelalaian TIDAK Adanya Sistem Backup Data Nasi...

2.3 Pre-Processing

Data yang diperoleh kemudian diproses untuk menghilangkan noise dan memastikan struktur yang terstruktur. Pada *pre-processing* dataset yang sudah dikumpulkan akan diolah untuk dijadikan ke format yang lebih efisien dan terstruktur [13]. Tahap ini berisi pengubahan dataset menjadi mudah dipahami untuk hasil penelitian yang lebih optimal. *Pre-processing* mencakup beberapa langkah mulai dari *cleaning*, *tokenizing*, *stopwords*, hingga *stemming*. Proses ini dibutuhkan untuk mengatasi permasalahan yang akan datang saat tahap uji [14].

2.3.1 Cleaning

Data cleaning adalah penghapusan simbol atau karakter yang tidak relevan untuk membantu akurasi klasifikasi data [7]. *Data cleaning* akan mencakup simbol, *mention*, *URL*, emoji, dan data kosong. Pendekatan ini dirancang untuk mengurangi kemungkinan kemiripan dengan sumber lain sekaligus meningkatkan akurasi dan kualitas data yang diolah [15].

Tabel 2 Pre-Processing Data Cleaning

text	text_clean
Ga mau bayar tebusan yg hanya beberapa M , sed...	ga mau bayar tebusan yg hanya beberapa m , sed...
Pecat Menkominfo dan PENJARAKAN seumur hidup...	pecat menkominfo dan penjarakan seumur hidup...
Kebanyakan buka web gak jelas tuh.. main downl...	kebanyakan buka web gak jelas tuh.. main downl...

2.3.2 Tokenizing

Tokenizing adalah proses pemisahan teks menjadi kata maupun karakter agar dapat dianalisa dengan mengubah dari semula berupa dataset utuh berupa kalimat menjadi beberapa kata [16]. *Tokenize* dilakukan untuk mempermudah proses stopword [17].

Tabel 3 Pre-Processing Tokenizing

text_clean	text_token
ga mau bayar tebusan yg hanya beberapa m , sed...	[ga, mau, bayar, tebusan, yg, hanya, beberapa,...
pecat menkominfo dan penjarakan seumur hidup...	[pecat, menkominfo, dan, penjarakan, seumur, h...
kebanyakan buka web gak jelas tuh.. main downl...	[kebanyakan, buka, web, gak, jelas, tuh, main,...

2.3.3 Stopwords

Dataset yang telah dilakukan *tokenizing* masuk tahap *stopword*. *Stopword* adalah proses mengeliminasi kata-kata yang dianggap kurang penting dalam menentukan sentiment [18]. Kata tersebut dianggap sebagai noise dan harus dihilangkan untuk membantu proses uji [19].

Tabel 4 Pre-Processing Stopwords

text_token	text_stop
[ga, mau, bayar, tebusan, yg, hanya, beberapa,...	[ga, bayar, tebusan, backup, data, ga, ...
[pecat, menkominfo, dan, penjarakan, seumur, h...	[pecat, menkominfo, penjarakan, seumur, hidup,...
[kebanyakan, buka, web, gak, jelas, tuh, main,...	[kebanyakan, buka, web, gak, tuh, main, downlo..
[ini, akibat, politik, balas, budi, memilih, m...	[akibat, politik, balas, budi, memilih, mentri...

2.3.4 Stemming

Kata yang telah diproses *stopwords* langsung diproses lagi dalam tahap *stemming. Stemming* adalah tahapan mengubah kalimat dari sebuah kata menjadi bentuk dasarnya [20]. Penelitian kali ini akan menggunakan *library* sastrawi dalam proses *stemming.* Sastrawi adalah library yang dirancang menggunakan algoritma Nazief dan Adriani (NA). Algoritma ini mengikuti kaidah Bahasa Indonesia [21].

Tabel 5 Pre-Processing Stemming

text_stop	stemmed
[ga, bayar, tebusan, backup, data, ga, ...	[ga, bayar, tebus, backup, data, ga, ta...
[pecat, menkominfo, penjarakan, seumur, hidup,...	[pecat, menkominfo, penjara, umur, hidup, pdn,...
[kebanyakan, buka, web, gak, tuh, main, downlo..	[banyak, buka, web, gak, tuh, main, download, ...
[akibat, politik, balas, budi, memilih, mentri...	[akibat, politik, balas, budi, pilih, mentri...

2.4 Labelling

Masuk tahap *Labelling* yang dimana hasil dari tahapan *stemming* akan dilakukan perhitungan dari ulasan yang terambil, sehingga dapat menghasilkan sebuah label [22]. *Labelling* dilakukan dengan daftar kata positif dan negative dari kamus sentiment, di mana skor ditambahkan berdasarkan kecocokan kata [23]. Dalam penelitian kali ini labelling akan dilakukan berdasarkan sentiment_score dengan tiga sentimen yaitu positif, negatif, dan netral.

Tabel 6 Labelling Klasifikasi

filtered_text	sentiment_score	sentiment
[ga, bayar, tebus, backup, data, ga, ta...	-3	negatif
[pecat, menkominfo, penjara, umur, hidup, pdn,..	-3	negatif
[banyak, buka, web, gak, tuh, main, download, ...	-4	negatif
[akibat, politik, balas, budi, pilih, mentri...	-3	negatif

2.5 Data Splitting

Data Splitting digunakan untuk membagi data menjadi data latih dan data uji dengan tujuan mencari akurasi terbaik [24]. Data akan dibagi menjadi 3 pembagian seperti yang terlampir pada tabel berikut.

Tabel 7 Splitting Data Train & Data Test

Data Train	*Data Test*
90%	10%
80%	20%
70%	30%

2.6 Tahap Uji

Dataset yang telah melalui proses *pre-processing* dan juga *labelling* akan diuji dengan perbandingan tiga model yakni *Support Vector Machine*, *Naïve Bayes*, dan *Random Forest*. Tahap uji penelitian akan dilakukan dengan dua tahap skenario yang berbeda. Pada tahap pertama dataset akan diuji menggunakan *oversampling SMOTE.* Pada tahap kedua dataset akan diuji tanpa menggunakan *oversampling SMOTE.* Synthetic Minority Oversampling Technique (SMOTE) merupakan algoritma yang bekerja untuk mengontrol pemerataan distribusi data pada suatu kelas mayoritas dengan membuat sampel buatan dari kelas minoritas untuk menyeimbangkan perbandingan data [25]. Penggunaan SMOTE pada penelitian kali ini bertujuan untuk mengetahui pengaruh SMOTE dalam analisis sentimen.

2.7 Validasi

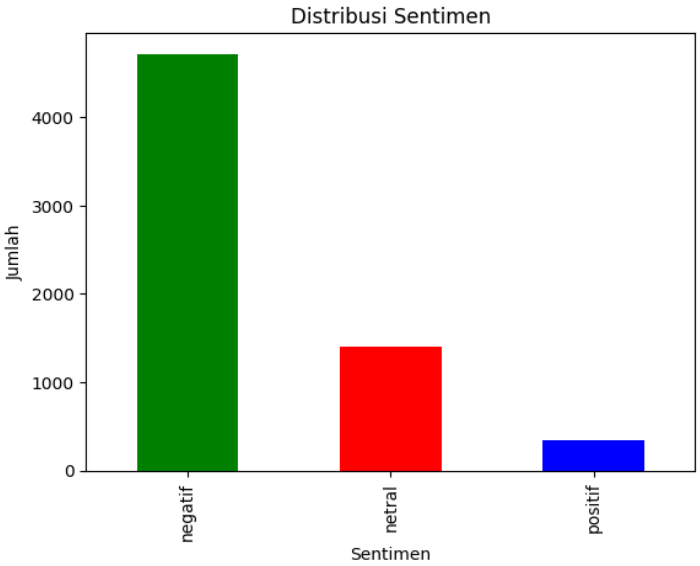
Validasi dilakukan untuk mengukur keakuratan hasil dari model *machine learning* sebelumnya. Tahap evaluasi akan dilakukan menggunakan visualisasi *confussion matrix* dengan bentuk table 3x3 yang dibagi indikator sentimen positif, negatif, dan netral. Melalui metode confusion matrix bisa didapat nilai akurasi dalam mengategorikan data dengan tepat dengan keseluruhan data [2].

3. HASIL DAN PEMBAHASAN

Pada bagian hasil akan berisi hasil dari penelitian meliputi tahap uji dan validasi dari setiap model *machine learning.* Pada tahap *labelling* didapat bahwa untuk klasifikasi sentimen negatif memiliki jumlah total lebih banyak daripada klasifikasi sentimen lainnya dengan negatif 4728 data, netral 1406 data, positif 345 data. Hasil klasifikasi tertera pada Tabel 8 dan Gambar 2.

Tabel 8 Dataset setelah Pre-Processing

filtered_text	**sentiment_score**	**sentiment**
[ga, bayar, tebus, backup, data, ga, ta...	-3	negatif
[pecat, menkominfo, penjara, umur, hidup, pdn,..	-3	negatif
[banyak, buka, web, gak, tuh, main, download, ...	-4	negatif
[akibat, politik, balas, budi, pilih, mentri...	-3	negatif



Gambar 2 Grafik Perbandingan Data Sentimen

Dataset yang telah diklasifikasikan akan diuji dengan model *machine learning* yang ditentukan yakni *Support Vector Machine*, *Random Forest*, dan *Naive bayes*. Tahap uji dibagi menjadi dua skenario dengan skenario 1 tahap uji menggunakan *SMOTE* dan skenario 2 tahap uji tanpa *SMOTE*.

3.1 Skenario 1 (SMOTE)

Dari hasil penelitian setelah melalui tahap uji dengan *SMOTE*, diperoleh akurasi sebesar 96% untuk *Support Vector Machine*, 95% untuk, *Random Forest* dan 89% untuk *Naïve Bayes*. Setelah dilakukan uji terlihat bahwa keakuratan prediksi paling rendah terdapat pada 89% untuk *Naïve bayes* dan yang tertinggi pada 96% untuk *Support Vector Machine*. Hasil perbandingan terlihat pada Tabel Hasil uji dan Gambar Visualisasi hasil uji setiap model.

Tabel 9 Hasil Uji SVM (Skenario 1)

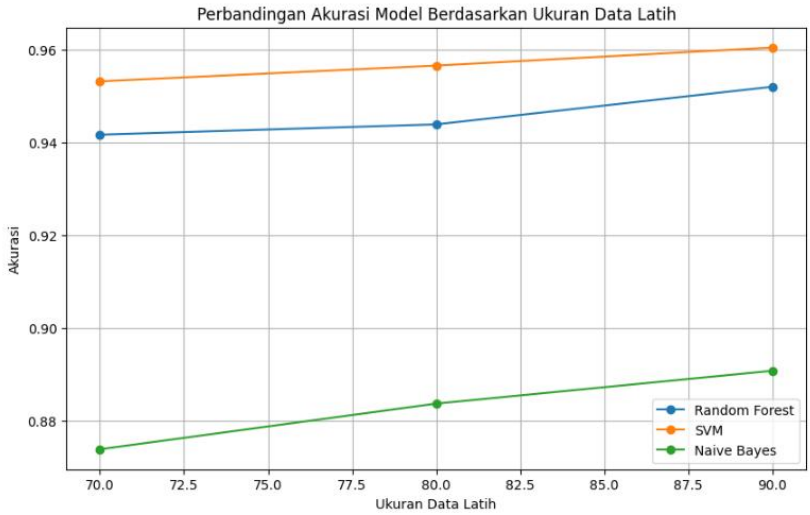
Data Train	***Data Test***	**(*Support Vector Machine*)**			
		Akurasi	**Precision**	**Recall**	**F1-Score**
90%	10%	0.960536	0.961894	0.960536	0.960501
80%	20%	0.956644	0.958596	0.956644	0.956561
70%	30%	0.953242	0.955187	0.953242	0.953141

Tabel 10 Hasil Uji RF (Skenario 1)

Data Train	***Data Test***	**(*Random Forest*)**			
		Akurasi	**Precision**	**Recall**	**F1-Score**
90%	10%	0.952079	0.954322	0.952079	0.951810
80%	20%	0.943955	0.946839	0.943955	0.943667
70%	30%	0.941729	0.944020	0.941729	0.941555

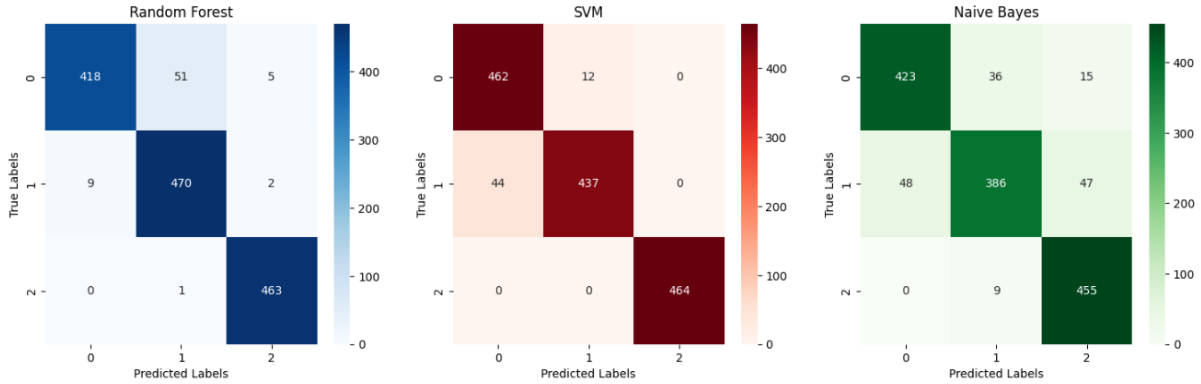
Tabel 11 Hasil Uji NB (Skenario 1)

Data Train	***Data Test***	**(*Naïve Bayes*)**			
		Akurasi	**Precision**	**Recall**	**F1-Score**
90%	10%	0.890768	0.891353	0.890768	0.889304
80%	20%	0.883680	0.884614	0.883680	0.882145
70%	30%	0.873825	0.874402	0.873825	0.871949



Gambar 3 Grafik Hasil Uji (Skenario 1)

Dari hasil uji didapat *data splitting* dengan perbandingan 90:10 menghasilkan akurasi tertinggi dari masing-masing model *machine learning*. Untuk mengetahui keakuratan hasil dari tahap uji diperlukan tahap validasi. Validasi akan dilakukan menggunakan *confussion matrix*. Pada tahap validasi didapat hasil yang lumayan akurat untuk setiap model dan setiap klasifikasi seperti yang tertera pada Gambar hasil confussion matrix berikut.



Gambar 4 Confussion Matrix Hasil Uji (Skenario 1)

3.2 Skenario 2 (Tanpa SMOTE)

Dari hasil penelitian setelah melalui tahap uji tanpa *SMOTE*, diperoleh akurasi sebesar 80% untuk *Random Forest*, 79% untuk *Support Vector Machine*, dan 76% untuk *Naïve Bayes.* Setelah dilakukan uji terlihat bahwa keakuratan prediksi paling rendah terdapat pada 76% untuk *Naïve bayes* dan yang tertinggi pada 80% untuk *Random Forest*. Hasil perbandingan terlihat pada Tabel Hasil uji dan Gambar Visualisasi hasil uji setiap model

Tabel 12 Hasil Uji SVM (Skenario 2)

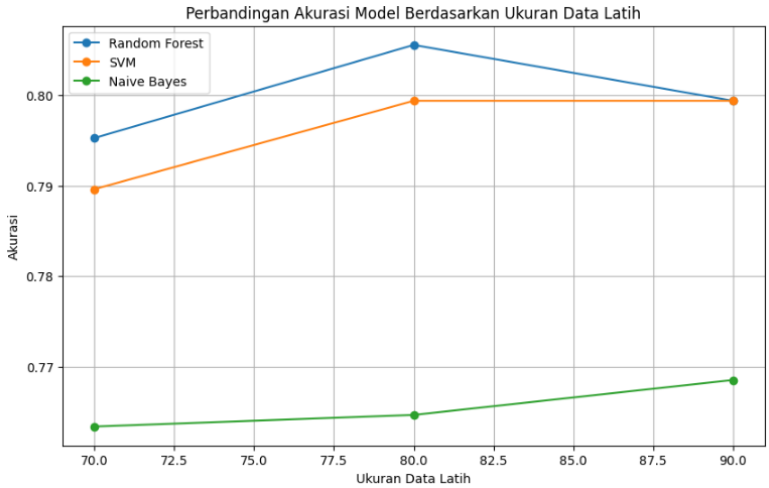
Data Train	***Data Test***	**(*Support Vector Machine*)** **Akurasi**	**Precision**	**Recall**	**F1-Score**
90%	10%	0.799383	0.793786	0.799383	0.758865
80%	20%	0.799383	0.793892	0.799383	0.756308
70%	30%	0.789609	0.785061	0.789609	0.742886

Tabel 13 Hasil Uji RF (Skenario 2)

Data Train	*Data Test*	(*Random Forest*) Akurasi	Precision	Recall	F1-Score
90%	10%	0.799383	0.776912	0.799383	0.766577
80%	20%	0.805556	0.794664	0.805556	0.772613
70%	30%	0.795267	0.771970	0.795267	0.761556

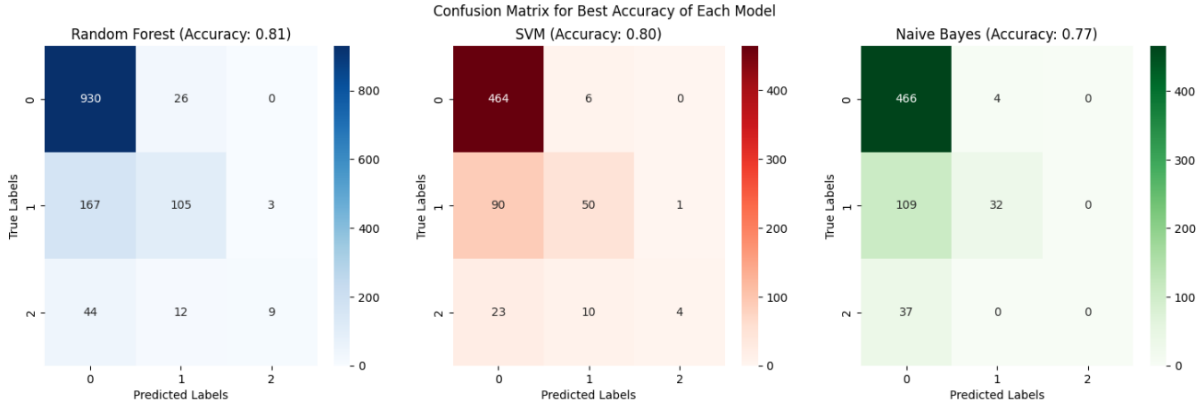
Tabel 14 Hasil Uji NB (Skenario 2)

Data Train	*Data Test*	(*Naïve Bayes*) Akurasi	Precision	Recall	F1-Score
90%	10%	0.768519	0.745693	0.768519	0.703435
80%	20%	0.764660	0.719146	0.764660	0.699062
70%	30%	0.763374	0.714351	0.763374	0.697906



Gambar 5 Grafik Hasil Uji (Skenario 2)

Dari hasil uji didapat *data splitting* dengan perbandingan 90:10 menghasilkan akurasi tertinggi untuk model *Support Vector Machine* dan *Naïve Bayes*. Sedangkan pada perbandingan 80:20 menghasilkan akurasi tertinggi untuk model *Random Forest*. Untuk mengetahui keakuratan hasil dari tahap uji diperlukan tahap validasi. Validasi akan dilakukan menggunakan *confussion matrix*. Pada tahap validasi didapat hasil yang terbilang kurang akurat untuk setiap model dan setiap klasifikasi seperti yang tertera pada Gambar hasil confussion matrix berikut.



Gambar 6 Confussion Matrix Hasil Uji Skenario 2)

4. SIMPULAN

Dari tahap validasi dikonfirmasi bahwa model yang dilatih menggunakan SMOTE memberikan akurasi prediksi yang lebih tinggi pada semua kategori sentimen dibandingkan model tanpa SMOTE. Dengan SMOTE, akurasi model berkisar antara 89% - 96%, sedangkan tanpa SMOTE hanya mencapai 76% - 80%. Metode Support Vector Machine (SVM) memberikan performa terbaik dengan akurasi tertinggi mencapai 96% saat SMOTE diterapkan. Dengan demikian, penelitian ini memberikan kontribusi signifikan dalam mendemonstrasikan keunggulan SMOTE dalam meningkatkan akurasi model machine learning pada dataset yang tidak seimbang. Analisis sentimen menunjukkan bahwa mayoritas komentar memiliki sentimen negatif (4728 data), diikuti oleh sentimen netral (1406 data) dan sentimen positif (345 data). Hasil penelitian ini diharapkan dapat menjadi landasan bagi evaluasi dan pengambilan kebijakan pemerintah, khususnya dalam memperbaiki kinerja dan respons terhadap isu-isu kritis seperti keamanan data masyarakat.

5. UCAPAN TERIMAKASIH

Peneliti mengucapkan syukur kepada Tuhan Yang Maha Esa atas rahmat dan anugerah-Nya yang telah memudahkan proses penelitian dan penyusunan manuskrip ini hingga selesai dengan baik. Rasa terima kasih yang tulus juga ditujukan kepada semua pihak yang telah memberikan dukungan sepanjang proses penelitian. Peneliti juga berterima kasih pada lab informatika UMSIDA yang sudah menyediakan fasilitas pendukung penelitian ini dan juga pada JOISIE (*Journal Of Information Systems And Informatics Engineering*) atas kesempatan untuk mempublikasikan hasil penelitian ini.

6. DAFTAR PUSTAKA

[1] A. A. Ningtyas, A. Solichin, and R. Pradana, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME NAÏVE BAYES," 2023.

[2] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 207–215, Apr. 2023, doi: 10.37859/coscitech.v4i1.4790.

[3] S. Faira Huwaida, R. Kusumawati, B. Isnaini, P. Korespondensi, and R. Artikel, "Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes," *Jambura J. Informatics*, vol. 6, no. 1, pp. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.

[4] E. Darwisah Harahap and R. Kurniawan, "Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes," 2024.

[5] E. Indrayuni and A. Nurhadi, "OPTIMASI NAIVE BAYES BERBASIS PSO UNTUK ANALISA SENTIMEN PERKEMBANGAN ARTIFICIAL INTELLIGENCE DI TWITTER," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 65–70, Aug. 2023, doi: 10.33480/inti.v18i1.4282.

[6] M. F. Naufal, T. Arifin, and H. Wirjawan, "Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial," *2023*, vol. 8, p. 82, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

[7] R. R. Putri and N. Cahyono, "ANALISIS SENTIMEN KOMENTAR MASYARAKAT TERHADAP PELAYANAN PUBLIK PEMERINTAH DKI JAKARTA DENGAN ALGORITMA SUPER VECTOR MACHINE DAN NAIVE BAYES," 2024.

[8] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest," 2023.

[9] M. Yasir and R. Suraji, "Perbandingan Metode Klasifikasi Naive Bayes, Decision, Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 pada Media Sosial

Youtube," *J. Cahaya Mandalika*, vol. 3, no. 2, pp. 180–192, 2023.
[10] F. Syah, H. Fajrin, A. N. Afif, R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *J. Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, 2023, doi: 10.35870/jti.
[11] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
[12] H. Hidayat, F. Santoso, and L. F. Lidimillah, "Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine)," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1729–1738, 2024, doi: 10.33379/gtech.v8i3.4497.
[13] E. Andrian and A. Rahman Isnain, "JURNAL MEDIA INFORMATIKA BUDIDARMA Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier," 2024, doi: 10.30865/mib.v8i2.7530.
[14] D. D. S. Dika Wardhani, Rika Astuti, "Optimasi Feature Selection Text Mining: Stemming dan Stopword," *Innov. J. Soc. Sci. Res.*, vol. 4, pp. 7537–7548, 2024.
[15] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.
[16] Ardiansyah and Nur'aini, "JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0) IMPLEMENTASI ANALISIS SENTIMEN MASYARAKAT MENGENAI KENAIKAN HARGA BBM DENGAN METODE NAIVE BAYES," *J. Inf. Syst. Informatics Eng.*, vol. 8, no. 1, pp. 1–9, 2024, doi: 10.35145/joisie.v8i1.3838.
[17] A. Syakir and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, p. 1796, Oct. 2023, doi: 10.30865/mib.v7i4.6648.
[18] T. Aura Azzahra *et al.*, "JURNAL MEDIA INFORMATIKA BUDIDARMA Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," 2024, doi: 10.30865/mib.v8i1.7186.
[19] I. Aida Sapitri and M. Fikry, "Pengklasifikasian Sentimen Ulasan Aplikasi Whatsapp Pada Google Play Store Menggunakan Support Vector Machine," *J. TEKINKOM*, vol. 6, no. 1, pp. 1–7, 2023, doi: 10.37600/tekinkom.v6i1.773.
[20] P. Kumala Sari and R. Randy Suryono, "KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE," 2024.
[21] S. S. Murni Pardede, Tulus Pramita Sihaloho, Jenheri Rejeki Tarigan, "Implementasi Algoritma Rabin Karp Dan Optimasi Dengan Algoritma Stemmer Sastrawi Dalam Deteksi Plagiat Pada Jurnal Skripsi Mahasiswa," *J. Armada Inform.*, vol. 6, no. 2, pp. 1–6, 2022, [Online]. Available: https://journal.stekom.ac.id/index.php/elkom/article/download/372/308/
[22] Mualfah Desti, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz AbdulSomad Dideportasi Dari Singapura MenggunakanAlgoritma SVM," 2023.
[23] R. Y. Pratama and P. F. Aryani, "KOMINFO TENTANG PEMBLOKIRAN GAME KEKERASAN SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON KOMINFO ' S," vol. 3, no. September, pp. 743–752, 2024.
[24] M. Muadin and H. Asnal, "IMPLEMENTASI METODE SUPPORT VECTOR MACHINE PADA OPINION MINING MASYARAKAT TERKAIT CHATGPT," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 7, no. 1, pp. 78–84, 2023.
[25] N. Sharfina and N. Ghaniaviyanto Ramadhan, "Terakreditasi SINTA Peringkat 3 Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes," 2026.