

ANALISIS SENTIMEN PUBLIK PADA PEMERINTAH DALAM SERANGAN RANSOMWARE DENGAN PENDEKATAN SMOTE

Oleh:

Mohammad Indra Prayugah (211080200115)

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

2024



Latar Belakang

- Dalam era digitalisasi ini Youtube menjadi media berita utama bagi masyarakat untuk menuaikan opini melalui komentar di setiap video. Reaksi publik terhadap peristiwa tersebut dapat diukur melalui analisis sentiment komentar di Youtube. Analisis sentimen publik melalui komentar youtube menjadi informasi langsung tentang persepsi masyarakat terhadap kinerja instansi pemerintah.
- Penelitian ini bertujuan untuk menganalisis sentiment publik terhadap kinerja instansi pemerintah pada kasus ransomware PDN dengan menggunakan perbandingan metode seperti Support Vector Machine, Random Forest, dan Naïve Bayers. **“Hasil dari penelitian ini diharapkan dapat mengevaluasi kinerja instansi pemerintah dalam menangansi sebuah kasus”**

Rumusan Masalah

Rumusan masalah akan lebih memfokuskan permasalahan utama pada penelitian ini

“Metode mana yang paling efektif untuk analisis sentimen publik berdasarkan komentar Youtube?”

Tujuan Penelitian

Untuk menemukan metode mana yang paling akurat untuk menghitung sentimen publik berdasarkan komentar youtube.

Batasan masalah

Terdapat beberapa batasan dalam penelitian ini diantaranya:

- Penelitian ini menggunakan dataset dari opini kometar video youtube dengan topik serangan ransomware pada pusat data nasional Indonesia
- Dataset terbatas didapat dari channel berita CNN dan MetroTV.
- Pelabelan dari opini dibagi menjadi 3 yaitu positif, negatif, dan netral.
- Klasifikasi data diberikan berdasarkan sebuah Pustaka Bahasa positif dan negatif didapat dari github.

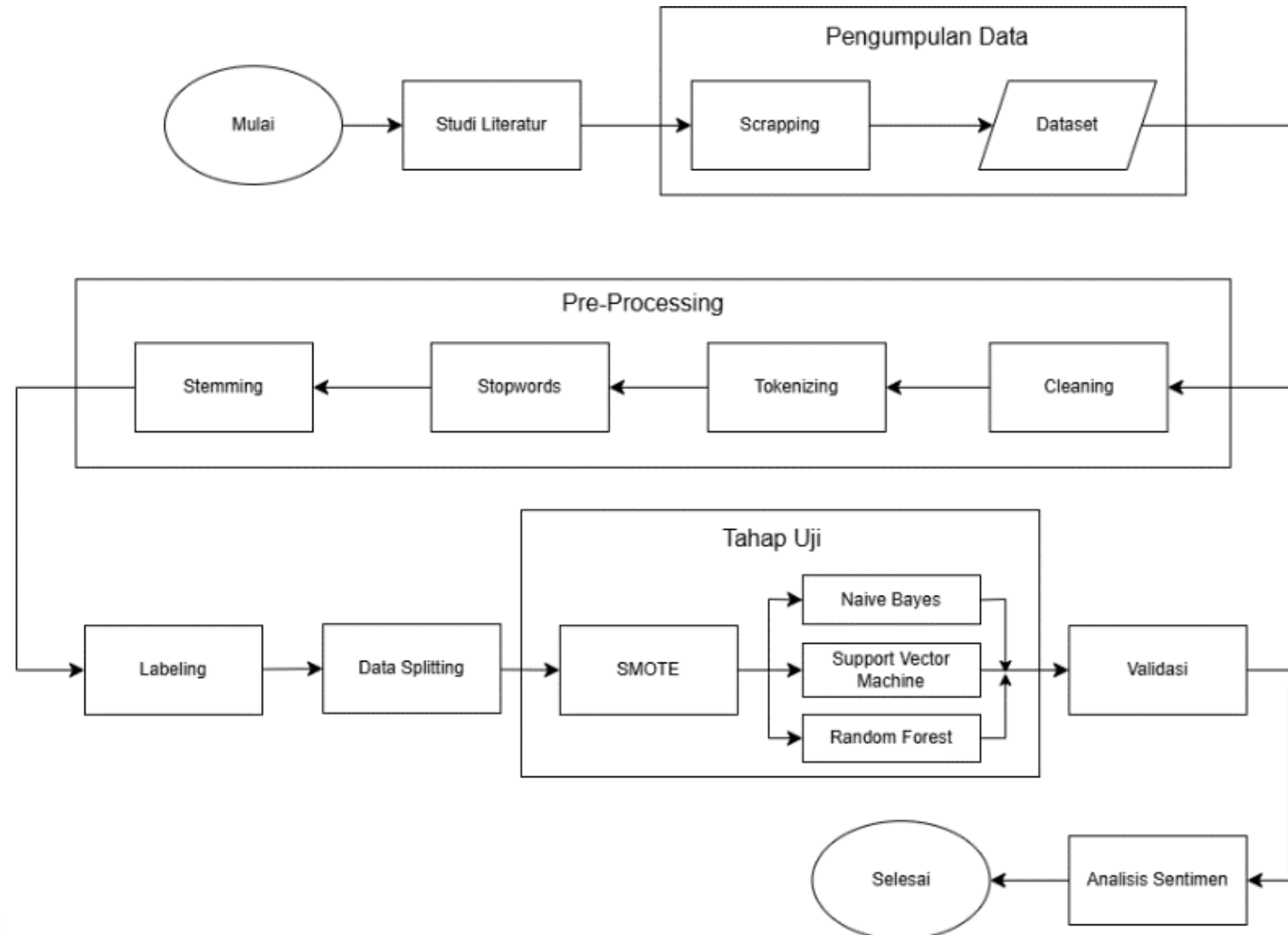
Penelitian Sebelumnya

Judul : Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial

Penulis : (Farid Naufal, 2023)

Penelitian ini membahas tentang pe,bandingan 3 metode algoritma dan bertujuan untuk menemukan algoritma yang memiliki akurasi paling akurat dengan membandingkan algoritma untuk memproses data cyberbullying pada media sosial

Sistematis Penelitian



Studi Literatur

- Tahap Studi Literatur dilakukan untuk memenuhi referensi penelitian dari berbagai sumber media mulai dari buku, jurnal, dan juga video guna dukungan penelitian dan bertujuan memberikan sebuah dasar tentang analisis sentimen.

Pengumpulan Data

- Data dikumpulkan menggunakan metode *crawling* dengan menggunakan YouTube Data API v3 dengan konteks dibatasi seputar serangan ransomware pada Pusat Data Nasional Indonesia yang diambil dari video channel berita terverifikasi.
- Data memiliki hingga **6000** Review data

	text	date
0	Ga mau bayar tebusan yg hanya beberapa M , sed...	2024-07-02T08:36:19Z
1	Pecat Menkominfo dan PENJARAKAN seumur hidup....	2024-07-02T06:27:55Z
2	Kebanyakan buka web gak jelas tuh.. main downl...	2024-07-01T19:02:12Z
3	ini akibat politik balas budi, memilih mentri ...	2024-06-30T16:14:50Z
4	Mana kronologinya? Penulis bohong	2024-06-30T05:36:53Z

Data Cleansing

- Pembersihan data, atau data *cleaning*, adalah proses krusial dalam pengolahan data yang melibatkan penghapusan symbol, *mention*, *URL*, emoji, dan karakter lainnya serta juga untuk mengatasi data kosong yang berpotensi mengganggu jalan penelitian.

“tinggal instal ulang 🤪 🤪”



“tinggal instal ulang”

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. Tokenize dilakukan untuk mempermudah proses stopwords.

“ga mau bayar tebusan yg hanya beberapa”



[ga, mau, bayar, tebusan, yg, hanya, beberapa]

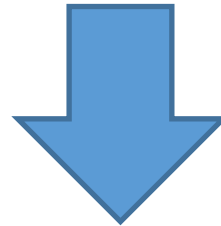
Stopword Removal

- *Stopwords* menghapus beberapa kata yang dianggap tidak berpengaruh saat proses analisis untuk meningkatkan keakurasian. Misalnya “pun”, “akan”, “di”, “nya”, “me”, “oleh”, “dan”, dll.

Stemming

- *Stemming* merupakan tahapan untuk memproses kalimat dari sebuah kata menjadi bentuk dasarnya dengan tujuan untuk menyederhanakan bentuk kata.

“Kekurangan masih banyak tapi sok bisa ngatasi”



“kurang masih banyak tapi sok bisa atasi”

Labelling

- Labelling adalah proses untuk memberikan label pada setiap data dalam dataset berdasarkan nilai komentar dan akan dibagi menjadi 3 sentimen antara lain positif, negatif, dan netral.

Data Splitting

- Dataset akan dibagi menjadi data train dan data test dengan pendekatan mencari akurasi terbaik sehingga terdapat beberapa perbandingan yang akan diterapkan.

<i>Data train</i>	90%	80%	70%
<i>Data test</i>	10%	20%	30%

Uji Coba

- Dataset yang telah dibagi mulai diuji memakai 3 metode yakni Random Forest, Support Vector Machine, Naïve Bayes dengan pengujian dibagi menjadi 2 scenario:
 1. Data diuji setelah diterapkan *oversampling SMOTE*
 2. Data diuji tanpa *oversampling SMOTE*

Validasi

- Tahap Validasi dilakukan untuk mengukur keakuratan dan menverifikasi hasil metode algoritma yang sebelumnya sudah memproses dataset.

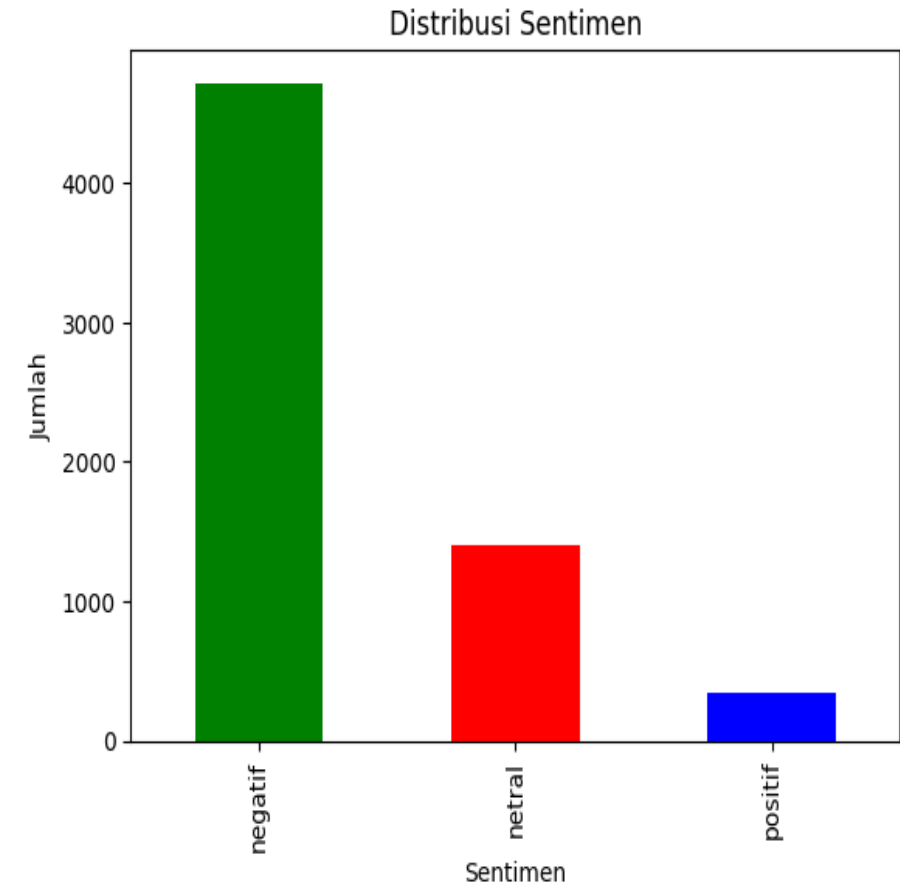
CONFUSION MATRIX

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Hasil Sentimen

Pada bagian hasil 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.

<u>filtered_text</u>	<u>sentiment_score</u>	<u>sentiment</u>
[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



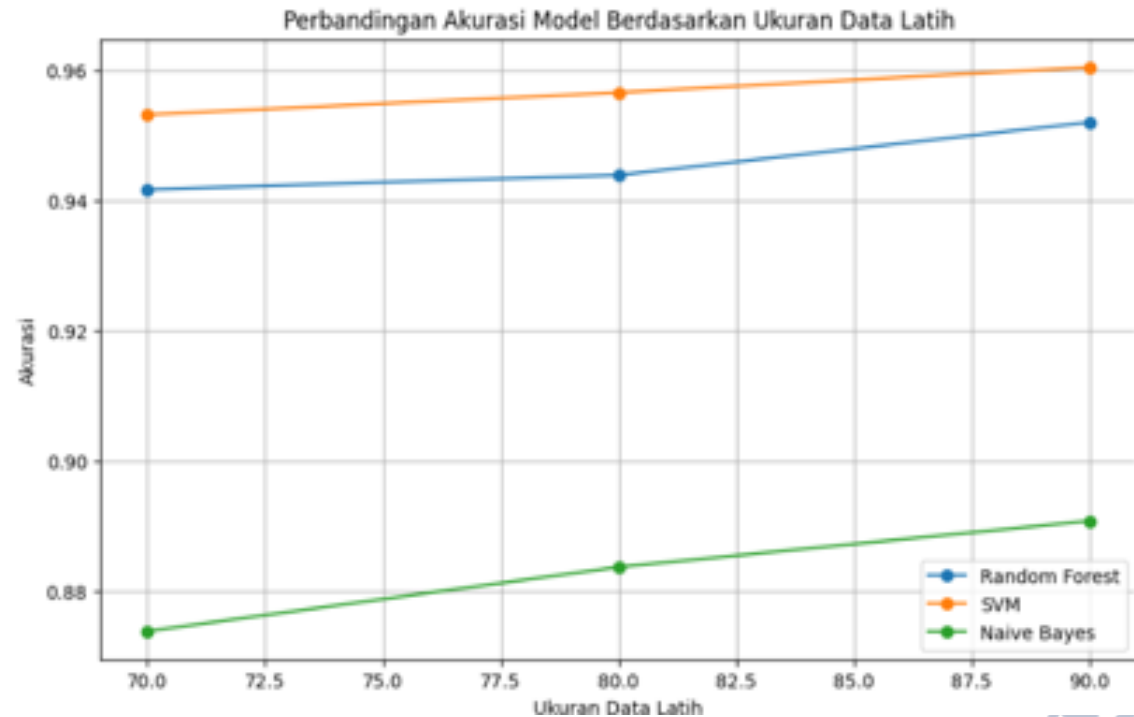
Hasil Uji Skenario 1 (SMOTE)

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 uii setiap model.

<i>Data Train</i>	<i>Data Test</i>	(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

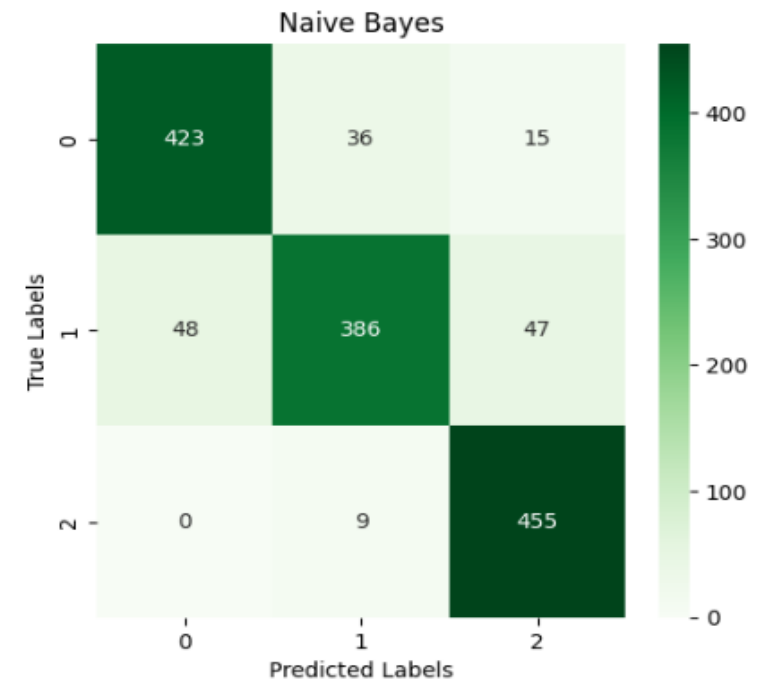
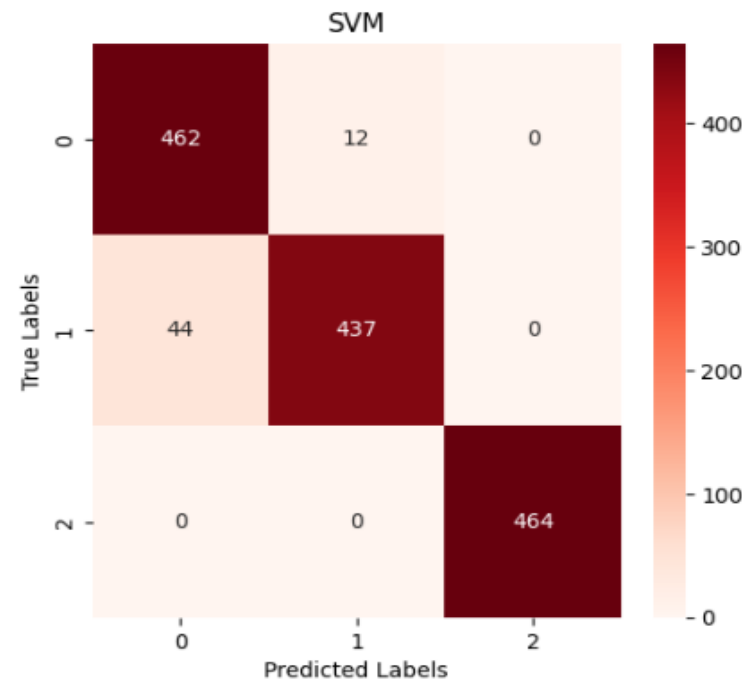
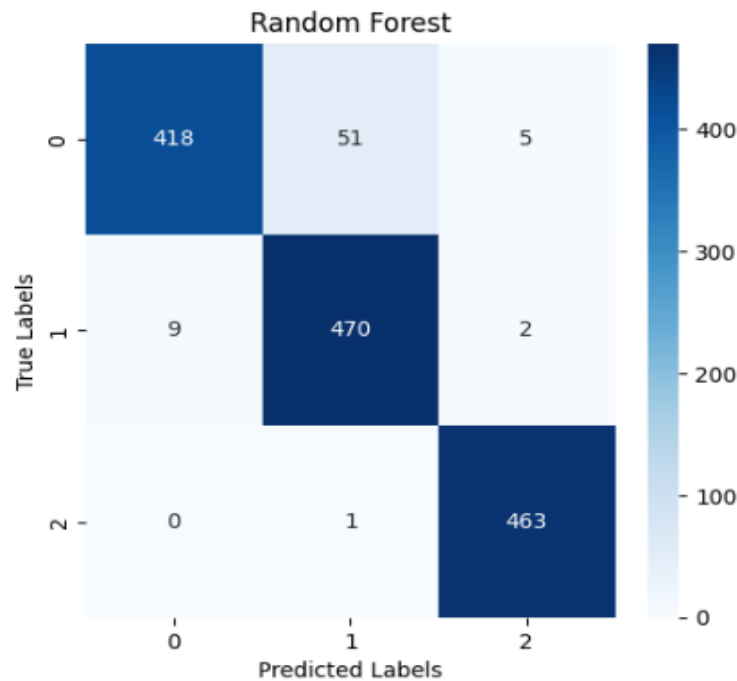
<i>Data Train</i>	<i>Data Test</i>	(Random Forest)			
		Akurasi	Precision	Recall	F1-Score
90%	10%	0.952079	0.954322	0.952079	0.951810
80%	20%	0.943955	0.946839	0.943955	0.943667
70%	30%	0.941729	0.944020	0.941729	0.941555

<i>Data Train</i>	<i>Data Test</i>	(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



Validasi Skenario 1 (SMOTE)

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.



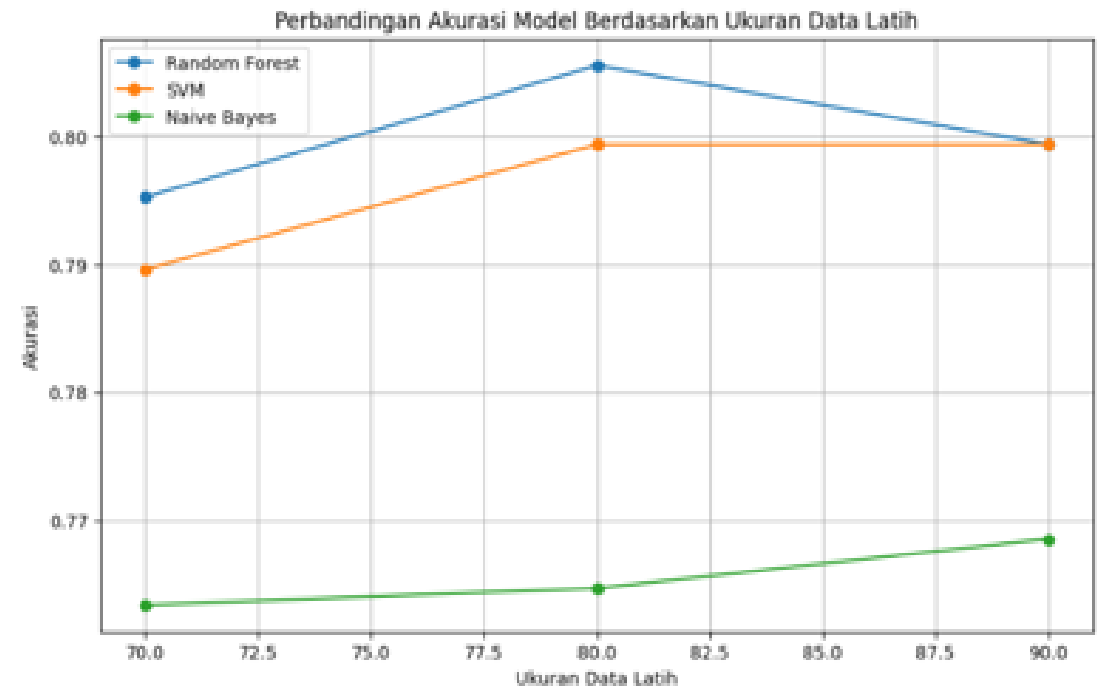
Hasil Uji Skenario 2 (Tanpa SMOTE)

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

<i>Data Train</i>	<i>Data Test</i>	<i>Akurasi</i>	<i>(Support Vector Machine)</i>		
			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

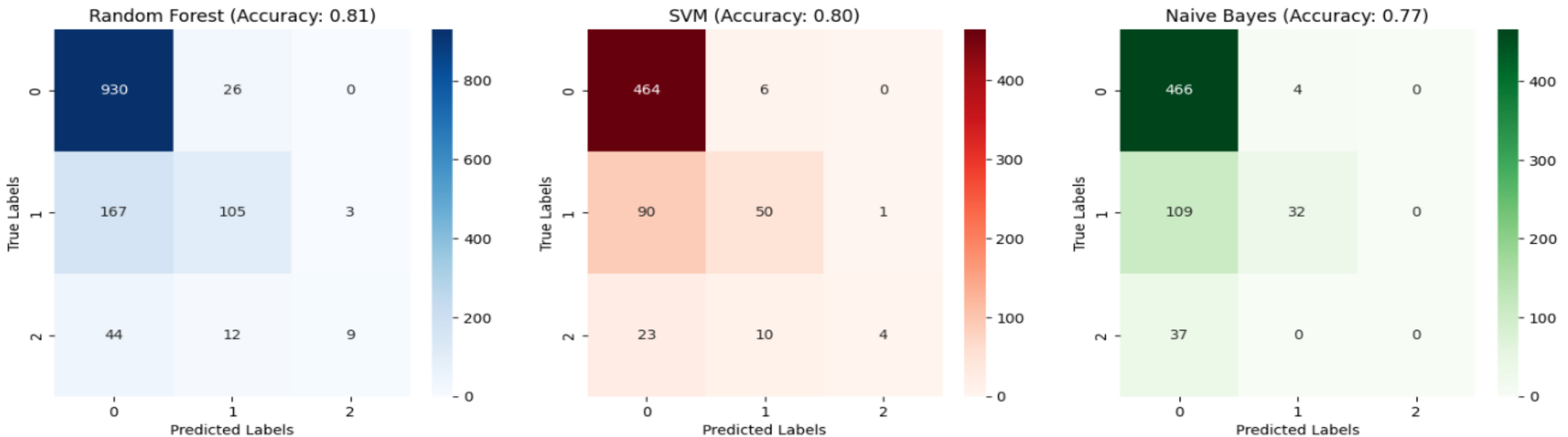
<i>Data Train</i>	<i>Data Test</i>	<i>Akurasi</i>	<i>(Random Forest)</i>		
			Precision	Recall	F1-Score
90%	10%	0.799383	0.776912	0.799383	0.766577
80%	20%	0.805556	0.794664	0.805556	0.772613
70%	30%	0.795267	0.771970	0.795267	0.761556

<i>Data Train</i>	<i>Data Test</i>	<i>Akurasi</i>	<i>(Naïve Bayes)</i>		
			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



Validasi Skenario 2 (Tanpa SMOTE)

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.



Kesimpulan Hasil Uji

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.

Referensi

- Aida Sapitri, I., & Fikry, M. (2023). Pengklasifikasian Sentimen Ulasan Aplikasi Whatsapp Pada Google Play Store Menggunakan Support Vector Machine. *Jurnal TEKINKOM*, 6(1), 1–7. <https://doi.org/10.37600/tekinkom.v6i1.773>
- Alfandi Safira, & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Andrian, E., & Rahman Isnain, A. (2024). *JURNAL MEDIA INFORMATIKA BUDIDARMA Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier*. <https://doi.org/10.30865/mib.v8i2.7530>
- Ardiansyah, & Nur'aini. (2024). 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. *Journal Of Information Systems And Informatics Engineering*, 8(1), 1–9. <https://doi.org/10.35145/joisie.v8i1.3838>
- Aura Azzahra, T., Anisa Sri Winarsih, N., Wilujeng Saraswati, G., Ocky Saputra, F., Syaifur Rohman, M., Oka Ratmana, D., Anggi Pramunendar, R., & Fajar Shidik, G. (2024). *JURNAL MEDIA INFORMATIKA BUDIDARMA Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube*. <https://doi.org/10.30865/mib.v8i1.7186>
- Chely Aulia Misrun, Haerani, E., Fikry, M., & Budianita, E. (2023). Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 207–215. <https://doi.org/10.37859/coscitech.v4i1.4790>
- Darwisah Harahap, E., & Kurniawan, R. (2024). *Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes*.
- Dika Wardhani, Rika Astuti, D. D. S. (2024). Optimasi Feature Selection Text Mining: Stemming dan Stopword. *INNOVATIVE: Journal Of Social Science Research*, 4, 7537–7548.
- Faira Huwaida, S., Kusumawati, R., Isnaini, B., Korespondensi, P., & Artikel, R. (2024). Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes. *Jambura Journal of Informatics*, 6(1), 26–39. <https://doi.org/10.37905/jji.v6i1.24718>
- Hidayat, H., Santoso, F., & Lidimillah, L. F. (2024). Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine). *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1729–1738. <https://doi.org/10.33379/gtech.v8i3.4497>
- Indrayuni, E., & Nurhadi, A. (2023). OPTIMASI NAIVE BAYES BERBASIS PSO UNTUK ANALISA SENTIMEN PERKEMBANGAN ARTIFICIAL INTELLIGENCE DI TWITTER. *INTI Nusa Mandiri*, 18(1), 65–70. <https://doi.org/10.33480/inti.v18i1.4282>

Referensi

- Kevin, K., Enjeli, M., & Wijaya, A. (2024). Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Computer Science*, 2(2), 89–98. <https://doi.org/10.58602/jics.v2i2.24>
- Kumala Sari, P., & Randy Suryono, R. (2024). *KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE* (Vol. 7, Issue 1).
- Manullang, O., Prianto, C., & Harani, N. H. (2023). *Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest*.
- Muadin, M., & Asnal, H. (2023). IMPLEMENTASI METODE SUPPORT VECTOR MACHINE PADA OPINION MINING MASYARAKAT TERKAIT CHATGPT. *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), 78–84.
- Mualfah Desti. (2023). *Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz AbdulSomad Dideportasi Dari Singapura Menggunakan Algoritma SVM*.
- Murni Pardede, Tulus Pramita Sihalohe, Jenheri Rejeki Tarigan, S. S. (2022). Implementasi Algoritma Rabin Karp Dan Optimasi Dengan Algoritma Stemmer Sastrawi Dalam Deteksi Plagiat Pada Jurnal Skripsi Mahasiswa. *Jurnal Armada Informatika*, 6(2), 1–6. <https://journal.stekom.ac.id/index.php/elkom/article/download/372/308/>
- Naufal, M. F., Arifin, T., & Wirjawan, H. (n.d.). Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial. 2023, 8, 82. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- Ningtyas, A. A., Solichin, A., & Pradana, R. (2023). *ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME NAÏVE BAYES* (Vol. 20, Issue 1).
- Pratama, R. Y., & Aryani, P. F. (2024). *KOMINFO TENTANG PEMBLOKIRAN GAME KEKERASAN SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON KOMINFO ' S*. 3(September), 743–752.
- Putri, R. R., & Cahyono, N. (2024). ANALISIS SENTIMEN KOMENTAR MASYARAKAT TERHADAP PELAYANAN PUBLIK PEMERINTAH DKI JAKARTA DENGAN ALGORITMA SUPER VECTOR MACHINE DAN NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).
- Sharfina, N., & Ghaniaviyanto Ramadhan, N. (2026). *Terakreditasi SINTA Peringkat 3 Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes* (Vol. 7, Issue 1).
- Syah, F., Fajrin, H., Afif, A. N., Saeputra, R., Mirranty, D., & Saputra, D. D. (2023). Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization. *Jurnal Teknologi Informasi Dan Komunikasi*, 7(1). <https://doi.org/10.35870/jti>

Referensi

- Syakir, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(4), 1796. <https://doi.org/10.30865/mib.v7i4.6648>
- Yasir, M., & Suraji, R. (2023). Perbandingan Metode Klasifikasi Naive Bayes, Decision, Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 pada Media Sosial Youtube. *Jurnal Cahaya Mandalika (JCM)*, 3(2), 180–192.

