

Analisis Sentimen Publik terhadap Nadiem Makarim sebagai Mendikbudristek menggunakan *Support Vector Machine* (SVM)

Public Sentiment Analysis of Nadiem Makarim as Minister of Education, Culture, Research, and Technology using Support Vector Machine (SVM)

¹Shasha Ramadhani Putri*, ²Muhammad Arifin, ³Supriyono

^{1,2,3}Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus

^{1,2,3}Jl. Lkr. Utara, Kayuapu Kulon, Gondangmanis, Bae, Kudus, Jawa Tengah, Indonesia 59327

*e-mail: 202053048@std.umk.ac.id, arifin.m@umk.ac.id, supriyono.si@umk.ac.id

(received: 7 February 2025, revised: 12 February 2025, accepted: 13 February 2025)

Abstrak

Media sosial menjadi platform utama untuk menyampaikan opini tentang kinerja pejabat publik, termasuk Nadiem Makarim, Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi. Opini di Twitter mencerminkan persepsi publik yang beragam, sehingga analisis sentimen diperlukan untuk memahami tren tersebut. Penelitian ini bertujuan menganalisis sentimen publik terhadap kinerja Nadiem Makarim dan mengoptimalkan model klasifikasi sentimen dalam menghadapi ketidakseimbangan data. Metode yang digunakan adalah *Support Vector Machine* (SVM) dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) melalui tiga skenario: tuning parameter TF-IDF, pemilihan kernel terbaik pada SVM, dan penerapan *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan data. Hasil eksperimen menunjukkan kombinasi $\text{max_features} = 2000$ dan $\text{min_df} = 2$ menghasilkan F1-score terbaik sebesar 68% dengan kernel linear paling stabil. Meskipun SMOTE menyeimbangkan distribusi kelas, akurasi sedikit menurun dari 68% menjadi 66%.

Kata kunci: analisis sentimen, *support vector machine*, twitter, nadiem makarim

Abstract

Social media has become a primary platform for expressing opinions on the performance of public officials, including Nadiem Makarim, the Minister of Education, Culture, Research, and Technology. Opinions on Twitter reflect diverse public perceptions, making sentiment analysis essential to understanding these trends. This study aims to analyze public sentiment toward Nadiem Makarim's performance and optimize sentiment classification models in handling data imbalance. The methodology employs a *Support Vector Machine* (SVM) with *Term Frequency-Inverse Document Frequency* (TF-IDF) through three scenarios: tuning TF-IDF parameters, selecting the best SVM kernel, and applying the *Synthetic Minority Oversampling Technique* (SMOTE) to address data imbalance. Experimental results indicate that the combination of $\text{max_features} = 2000$ and $\text{min_df} = 2$ yields the best F1-score of 68%, with the linear kernel being the most stable. Although SMOTE successfully balances class distribution, accuracy slightly decreases from 68% to 66%.

Keywords: sentiment analysis, *support vector machine*, twitter, nadiem makarim

1 Pendahuluan

Perkembangan teknologi digital telah mengubah cara masyarakat mengekspresikan pendapat dan pandangan mereka terhadap berbagai isu, termasuk kebijakan publik. Media sosial, terutama platform seperti Twitter (sekarang dikenal sebagai X), telah menjadi saluran utama bagi publik untuk menyampaikan opini secara cepat dan terbuka. Di Indonesia, penggunaan media sosial berperan penting dalam diskusi tentang isu-isu sosial, politik, dan pendidikan. Berdasarkan laporan *Digital 2024 Indonesia* yang diterbitkan oleh We Are Social dan Meltwater, terdapat lebih dari 24,85 juta

<http://sistemasi.ftik.unisi.ac.id>

pengguna Twitter di Indonesia [1]. Hal ini menjadikan Twitter sebagai platform penting untuk menganalisis opini publik.

Nadiem Makarim, yang menjabat sebagai Menteri Pendidikan dan Kebudayaan, Riset, dan Teknologi dari tahun 2019 hingga 2024, telah meluncurkan sejumlah kebijakan inovatif yang bertujuan untuk mereformasi sistem pendidikan Indonesia. Salah satu kebijakan yang paling dikenal adalah program Merdeka Belajar yang dirancang untuk memberikan lebih banyak fleksibilitas kepada sekolah dan siswa dalam proses pembelajaran. Kebijakan ini menimbulkan beragam reaksi dari masyarakat, baik dalam bentuk dukungan maupun kritik, yang kemudian banyak diungkapkan melalui media sosial.

Analisis sentimen terhadap opini publik yang berkembang di media sosial menjadi penting untuk memahami penerimaan kebijakan yang telah diterapkan. Dengan analisis sentimen, dapat diperoleh gambaran umum mengenai pola opini publik terhadap kebijakan tertentu. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi sentimen publik terhadap kinerja Nadiem Makarim sebagai Mendikbudristek berdasarkan opini di media sosial Twitter serta mengoptimalkan model klasifikasi sentimen untuk meningkatkan akurasi pada data yang tidak seimbang. Penelitian ini juga berkontribusi dalam pengembangan model klasifikasi sentimen dengan menguji beberapa strategi optimasi untuk meningkatkan performa model analisis sentimen terhadap opini publik.

2 Tinjauan Literatur

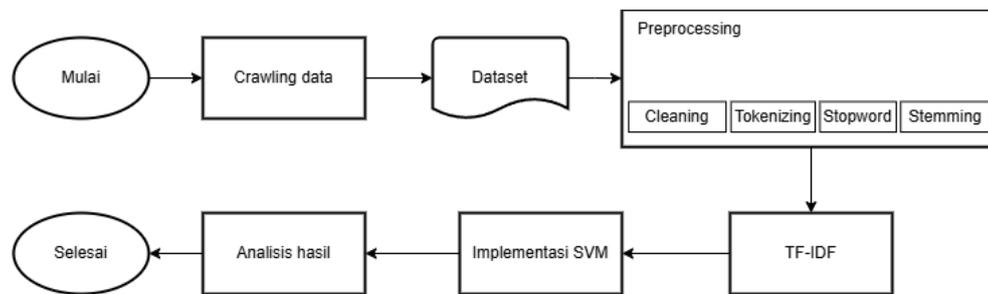
Support Vector Machine (SVM) adalah salah satu metode dalam machine learning yang digunakan secara luas untuk klasifikasi teks, termasuk analisis sentimen. SVM bekerja dengan menentukan hyperplane atau fungsi pemisah yang optimal untuk membagi kelas-kelas data dalam ruang multidimensi [2]. Beberapa penelitian sebelumnya menunjukkan keunggulan SVM dalam analisis sentimen dibandingkan metode lain. Penelitian yang dilakukan terhadap ulasan aplikasi Shopee menunjukkan bahwa SVM mampu mencapai akurasi sebesar 84,71%, lebih tinggi dibandingkan dengan Random Forest yang hanya mencapai 82,21% [3]. Selain itu, studi yang menganalisis sentimen di Twitter terkait Komisi Pemberantasan Korupsi menemukan bahwa SVM mampu memberikan akurasi sebesar 82%, dengan distribusi sentimen negatif yang dominan [4]. Penelitian lainnya oleh [5] juga mengungkapkan bahwa SVM mampu mencapai akurasi 84,37% dalam analisis sentimen maskapai penerbangan di Twitter.

Namun, penelitian terdahulu dalam analisis sentimen masih menghadapi beberapa keterbatasan. Salah satunya adalah ketidakseimbangan kelas dalam dataset, yang menyebabkan model cenderung lebih akurat dalam mengenali kelas mayoritas dibandingkan kelas minoritas. Selain itu, banyak penelitian sebelumnya yang belum mengeksplorasi optimasi fitur, seperti seleksi parameter dalam TF-IDF sehingga model tetap mempertimbangkan fitur yang kurang relevan. Pemilihan kernel SVM juga sering diabaikan, padahal pemilihan kernel yang sesuai dapat meningkatkan performa model secara signifikan.

Penelitian ini berusaha mengatasi keterbatasan tersebut dengan melakukan tiga skenario pengujian utama: tuning parameter TF-IDF, pemilihan kernel terbaik untuk klasifikasi sentimen, serta penerapan SMOTE untuk menangani ketidakseimbangan kelas dalam dataset. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan wawasan yang lebih komprehensif mengenai persepsi publik terhadap kebijakan pendidikan serta mengembangkan model analisis sentimen yang lebih akurat dan tahan terhadap ketidakseimbangan kelas.

3 Metode Penelitian

Proses penelitian ini dimulai dengan proses *crawling data* untuk mengumpulkan teks dari platform Twitter/X, yang kemudian disimpan dalam dataset. Selanjutnya, data mentah diproses melalui tahap *preprocessing*, yang mencakup *cleaning data*, *tokenizing*, *stopwords removal*, dan *stemming*. Data yang telah bersih diubah menjadi representasi numerik menggunakan metode TF-IDF. Representasi ini kemudian digunakan untuk melatih model Support Vector Machine (SVM) guna mengklasifikasikan sentimen menjadi positif, negatif, atau netral. Tahap akhir adalah analisis hasil, di mana performa model dievaluasi, dan *insight* dari hasil klasifikasi digunakan sebagai dasar untuk kesimpulan penelitian.



Gambar 1. Tahapan alur penelitian

3.1 Crawling Data

Crawling adalah sebuah proses mengumpulkan data dengan memanfaatkan Search API pada platform Twitter [6]. Data yang di-*crawl* yaitu kumpulan *tweet* dengan kata kunci ‘Nadiem Makarim’ dan ‘kinerja/kerja/prestasi’. Pengumpulan data dilakukan menggunakan *tools* Jupyter Notebook dan bahasa pemrograman Python, yang memungkinkan ekstraksi data secara otomatis dari Twitter. Data yang berhasil diambil adalah sebanyak 1524 *tweet* dalam waktu mulai tahun 2021. Contoh hasil *crawling data* Twitter dapat dilihat pada Tabel 1.

Tabel 1. Data tweet nadiem makarim

username	created_at	full_text
xxxxxxxxxxx	Tue Oct 08 05:28:06 +0000 2024	@Mdy_Asmara1701 NADIEM MAKARIM MENTERI PALING BERPRESTASI & KINERJA TERBAIK !!

3.2 Text Preprocessing

Setelah data terkumpul melalui proses *crawling*, tahap selanjutnya adalah *text preprocessing*. Tujuan dari tahap ini adalah untuk menghilangkan *noise* atau gangguan dalam data, meningkatkan kualitas data, dan memastikan bahwa data lebih mudah diolah oleh model [7]. Adapun alur proses *text preprocessing* antara lain:

- Cleaning Data* merupakan suatu tahap untuk menghilangkan komponen seperti tautan URL, mention seperti “@username”, *hashtag*, emoji, serta simbol khusus seperti tanda baca yang tidak diperlukan pada proses analisis sentimen [8]. Tahap ini juga termasuk proses *case folding*, yaitu mengubah semua huruf teks menjadi *lowercase*.
- Stopwords removal* merupakan tahap untuk menghilangkan kata-kata yang dianggap tidak penting atau tidak memiliki makna signifikan dalam analisis teks, seperti "dan", "di", "ke", "yang", "dengan", dll. Tujuan proses ini adalah memastikan agar kata-kata tersebut tidak memengaruhi hasil pada tahap selanjutnya [9].
- Tokenizing* adalah suatu proses untuk memecah teks menjadi unit-unit kecil yang disebut token, seperti kata, frasa, atau kalimat. Tahap ini juga bisa disebut proses pemenggalan kata, menghasilkan kata tunggal yang dapat memudahkan proses selanjutnya [8].
- Stemming* adalah mengubah kata-kata menjadi bentuk dasarnya atau bentuk akar. Tujuannya adalah untuk mengurangi variasi kata dengan makna yang sama, sehingga kata seperti “mendukung”, “dukung”, dan “didukung” semuanya diubah menjadi “dukung”.

Hasil *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *text preprocessing*

Sebelum	Sesudah
@Mdy_Asmara1701 NADIEM MAKARIM MENTERI PALING BERPRESTASI & KINERJA TERBAIK !!	nadiem makarim menteri prestasi baik

3.3 Pembobotan Kata TF-IDF

Penelitian ini menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk memberikan bobot pada kata-kata dalam dokumen atau tweet. Teknik ini menggabungkan dua model perhitungan bobot yaitu frekuensi kemunculan kata dalam sebuah dokumen dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut [10].

Model *Term Frequency* (TF) mengukur seberapa sering suatu kata muncul dalam satu dokumen. Semakin sering kata itu muncul, semakin tinggi bobotnya [11]. Sedangkan *Inverse Document*

Frequency (IDF) mengukur seberapa jarang sebuah kata muncul di seluruh dokumen [12]. Jika kata muncul di banyak dokumen, maka bobot IDF-nya akan rendah, karena dianggap sebagai kata umum yang tidak begitu informatif. Sebaliknya, jika kata hanya muncul di sedikit dokumen, IDF-nya akan tinggi, menandakan kata itu penting dan spesifik. Skor dapat diperoleh dengan menggunakan persamaan (1):

$$W_{t,d} = TF_{t,d} \times IDF_t \quad (1)$$

$W_{t,d}$ = bobot dari t (*term*) dalam suatu dokumen

$TF_{t,d}$ = frekuensi kemunculan t (*term*) dalam dokumen d

IDF_t = *Inverse Document Frequency*. Di mana persamaan IDF seperti pada persamaan (2).

$$IDF_t = \log \left(\frac{N}{nt} \right) \quad (2)$$

N = Jumlah dokumen, dibagi dengan nt = jumlah dokumen yang mengandung *term* t

3.4 Synthetic Minority Oversampling Technique (SMOTE)

Ketidakeimbangan kelas dalam dataset dapat menyebabkan model klasifikasi cenderung lebih akurat dalam memprediksi kelas mayoritas, sementara performanya menurun dalam memprediksi kelas minoritas. Dalam penelitian ini, teknik *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan untuk menangani ketidakseimbangan data sentimen. SMOTE bekerja dengan menciptakan data sintetis baru untuk kelas minoritas berdasarkan kedekatan antara titik data dalam ruang vektor, sehingga menghasilkan distribusi data yang lebih seimbang [13]. Teknik ini berbeda dari metode oversampling konvensional yang hanya menggandakan data minoritas, karena SMOTE secara aktif menghasilkan sampel baru dengan melakukan interpolasi antara sampel yang ada.

3.5 Klasifikasi Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode *supervised learning* yang digunakan untuk mengklasifikasikan data teks dengan tingkat akurasi yang cukup tinggi [14]. Algoritma klasifikasi SVM membangun model berdasarkan data training untuk menemukan hyperplane optimal yang memisahkan kelas-kelas dalam data. Model yang telah dilatih ini kemudian digunakan untuk memprediksi kelas dari data baru (data testing) yang belum pernah dilihat sebelumnya [15].

3.6 Evaluasi

Tahap evaluasi bertujuan untuk mengukur tingkat akurasi model yang telah dibangun menggunakan data latih. Selanjutnya, dilakukan perbandingan hasil dari dua dataset yang berbeda dengan menggunakan *confusion matrix* untuk menghitung nilai *precision*, *recall*, *f1-score*, dan *accuracy* [16].

Terdapat empat istilah yang digunakan digunakan dalam *confusion matrix* untuk mengevaluasi akurasi dan kesalahan prediksi suatu model klasifikasi yaitu *True Positive* (TP) yang berarti hasil dari prediksi klasifikasi di mana kelas positif diprediksi dengan benar, *True Negative* (TN) adalah hasil ketika kelas negatif diprediksi dengan benar, *False Positive* (FP) terjadi ketika kelas yang sebenarnya negatif diprediksi sebagai positif, dan *False Negative* (FN) adalah kasus di mana kelas yang sebenarnya positif justru diprediksi sebagai negatif [17]. Distribusi prediksi model terhadap kelas sebenarnya dirangkum dalam *Confusion Matrix* pada Tabel 3.

Tabel 3. Confusion matrix

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Dari hasil proses klasifikasi pada *confusion matrix*, dihitung tiap nilai dengan *accuracy*, *precision*, *recall*, dan *F1-score*. Akurasi (*accuracy*) merupakan rasio antara jumlah data yang diklasifikasikan dengan benar dan total keseluruhan data [18], rumus akurasi dapat dilihat pada persamaan (3):

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \quad (3)$$

Presisi (*precision*) adalah rasio antara jumlah data positif yang diklasifikasikan dengan benar dan jumlah data yang diprediksi positif [18], dapat dihitung dengan persamaan (4):

$$Precision = \left(\frac{TP}{TP + FP} \right) \times 100\% \quad (4)$$

Recall adalah rasio antara jumlah data positif yang diklasifikasikan dengan benar dan total data yang sebenarnya berlabel positif [18], rumus dapat dilihat pada persamaan (5):

$$Recall = \left(\frac{TP}{TP + FN} \right) \times 100\% \quad (5)$$

F1-score merupakan rata-rata harmonis dari *recall* dan *precision* [18]. Rumus *F1-score* dapat dihitung dengan persamaan (6):

$$F1 = \left(\frac{2 \times precision \times recall}{precision + recall} \right) \times 100\% \quad (6)$$

4 Hasil dan Pembahasan

Pada penelitian ini, dilakukan tiga skenario pengujian untuk mengevaluasi performa model pembelajaran mesin menggunakan SVM yaitu tuning parameter pada TF-IDF, pengujian pengaruh pemilihan kernel, dan pengujian pengaruh SMOTE. Setiap tahap pengujian menggunakan hasil terbaik dari tahap sebelumnya sebagai dasar untuk pengujian selanjutnya.

4.1 Parameter Tuning TF-IDF

Pengujian pertama bertujuan untuk mengevaluasi pengaruh parameter pada TF-IDF terhadap performa model. Parameter yang diuji meliputi *min_df* dan *max_features*, dengan *ngram_range* (1,2). Parameter *min_df* digunakan untuk menghapus kata yang jarang muncul di sejumlah dokumen, dan *max_features* menentukan jumlah maksimal kata teratas yang dipertahankan berdasarkan frekuensinya. Sebagai contoh, jika *min_df* = 3, maka kata-kata yang muncul di kurang dari 3 dokumen akan dihapus, dan *max_features* = 3000 hanya mempertahankan 3000 kata dengan frekuensi tertinggi.

Tabel 4. Parameter tuning TF-IDF

<i>min_df</i>	<i>max_features</i>	<i>f1-score</i>
0.01	2000	64%
0.05	3000	57%
2	2000	68%
3	3000	66%

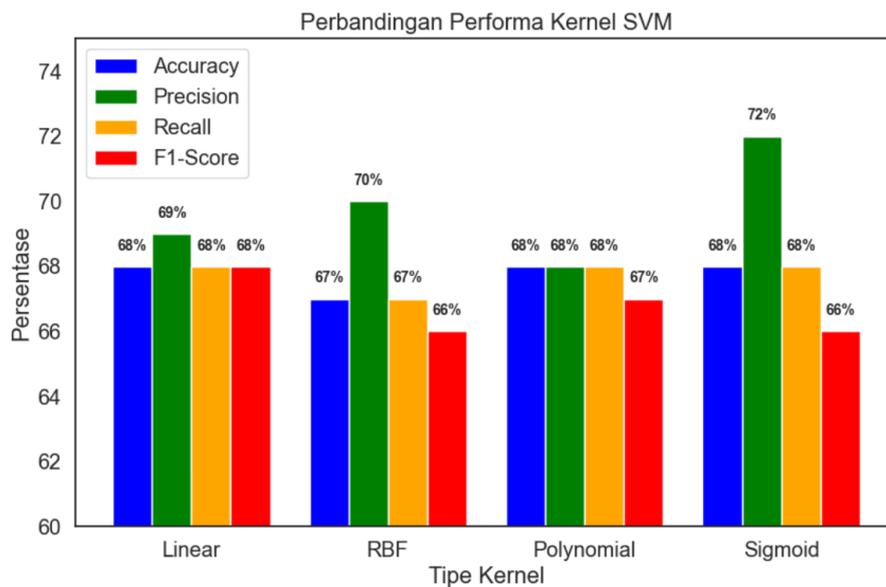
Dari **Tabel 4** dapat dilihat hasil pengujian menunjukkan bahwa parameter *min_df* = 2 dan *max_features* = 2000 memberikan nilai *f1-score* terbaik yaitu 68%, sehingga pada pengujian selanjutnya akan menggunakan parameter tersebut.

4.2 Pengujian Pengaruh Pemilihan Kernel

Pemilihan kernel dalam SVM berperan penting dalam menentukan pola pemisahan data. Untuk itu, penelitian ini menguji empat kernel berbeda, yaitu Linear, RBF, Polynomial, dan Sigmoid, guna menilai pengaruhnya terhadap performa klasifikasi. Kernel dengan hasil terbaik akan digunakan pada tahap pengujian berikutnya.

Tabel 5. Pengujian pengaruh pemilihan kernel

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	68%	69%	68%	68%
RBF	67%	70%	67%	66%
Polynomial	68%	68%	68%	67%
Sigmoid	68%	72%	68%	66%



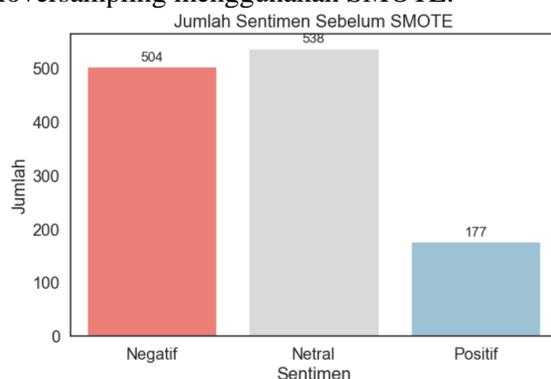
Gambar 2. Hasil pengujian pemilihan kernel

Tabel 5 dan **Gambar 2** menunjukkan hasil pengujian berbagai kernel dalam klasifikasi sentimen menggunakan metode SVM. Evaluasi dilakukan berdasarkan empat metrik utama yaitu accuracy, precision, recall, dan f1-score. Pemilihan kernel yang tepat menjadi faktor krusial, terutama ketika menghadapi ketidakseimbangan kelas (*class imbalance*) dalam dataset. Dalam situasi seperti ini, F1-score menjadi metrik yang sangat penting karena mempertimbangkan keseimbangan antara precision dan recall. F1-score memastikan bahwa model tidak hanya fokus pada kelas mayoritas tetapi juga tetap responsif terhadap kelas minoritas, sehingga hasil klasifikasi lebih adil dan representatif. Metrik ini sangat membantu dalam menghindari bias model terhadap kelas dengan jumlah data yang lebih dominan.

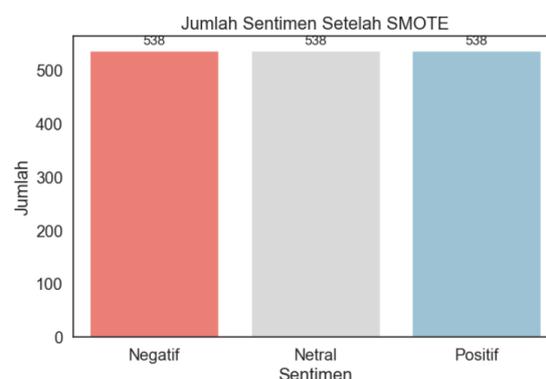
Berdasarkan hasil pengujian, kernel linear menunjukkan performa yang paling konsisten dibandingkan dengan kernel lainnya. Dengan F1-score sebesar 68%, kernel ini memberikan keseimbangan optimal antara akurasi dan kemampuan mengenali seluruh kelas dengan baik. Oleh karena itu, kernel linear menjadi pilihan terbaik dalam penelitian ini, mengingat kestabilannya dalam menangani dataset dengan distribusi sentimen yang tidak merata.

4.3 Pengujian Pengaruh SMOTE

Pengujian bertujuan untuk mengevaluasi pengaruh penerapan SMOTE pada data yang tidak seimbang terhadap performa model. SMOTE digunakan untuk meningkatkan distribusi kelas minoritas dengan menciptakan data sintesis [19]. Dalam penelitian ini, pengujian dilakukan dengan membandingkan performa model SVM pada data asli tanpa SMOTE dan data yang telah dioversampling menggunakan SMOTE.



Gambar 3. Jumlah sentimen sebelum SMOTE

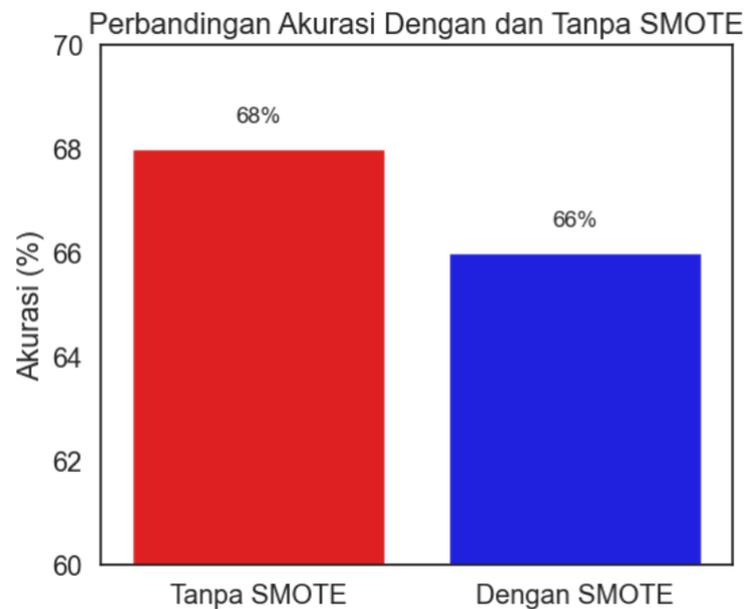


Gambar 4. Jumlah sentimen setelah SMOTE

Pada **Gambar 3** menampilkan distribusi jumlah data sebelum diterapkannya teknik SMOTE. Dari visualisasi tersebut, terlihat bahwa jumlah data sentimen positif jauh lebih sedikit dibandingkan dengan kelas negatif dan netral. Secara rinci, jumlah data sebelum oversampling adalah sebagai

berikut: 504 untuk kelas negatif, 538 untuk kelas netral, dan hanya 177 untuk kelas positif. Perbedaan ini menunjukkan adanya ketidakseimbangan kelas yang dapat menyebabkan model cenderung lebih dominan dalam mengklasifikasikan kelas dengan jumlah data yang lebih besar.

Sementara itu, **Gambar 4** menampilkan hasil distribusi data setelah dilakukan oversampling menggunakan SMOTE. Dari grafik tersebut, terlihat bahwa jumlah sampel pada setiap kelas telah diseimbangkan sehingga masing-masing kelas memiliki 538 sampel. Teknik SMOTE bekerja dengan mensintesis sampel baru berdasarkan kedekatan data dalam ruang vektor, sehingga kelas positif yang awalnya minoritas kini memiliki jumlah data yang setara dengan kelas lainnya. Dengan penerapan SMOTE, distribusi data menjadi lebih merata, yang diharapkan dapat meningkatkan performa model klasifikasi.



Gambar 5. Perbandingan akurasi dengan dan tanpa SMOTE

Pada **Gambar 5**, ditampilkan perbandingan akurasi model sebelum dan sesudah diterapkannya teknik SMOTE. Dari hasil yang diperoleh, akurasi model sebelum SMOTE mencapai 68%, sedangkan setelah diterapkan SMOTE, akurasi model mengalami sedikit penurunan menjadi 66%. Penurunan akurasi ini dapat disebabkan oleh beberapa faktor. Salah satunya adalah karena SMOTE bekerja dengan mensintesis sampel baru pada kelas minoritas berdasarkan interpolasi data yang sudah ada, sehingga data tambahan yang dihasilkan mungkin tidak sepenuhnya mencerminkan pola asli dari data sesungguhnya. Akibatnya, model dapat mengalami *overfitting* atau justru kehilangan beberapa pola penting dari data aslinya. Meskipun akurasi mengalami sedikit penurunan, penerapan SMOTE tetap memberikan manfaat, terutama dalam mengatasi ketidakseimbangan kelas. Model yang dilatih dengan data yang lebih seimbang diharapkan dapat lebih adil dalam mengklasifikasikan setiap kelas, terutama pada kelas minoritas yang sebelumnya kurang terwakili.

5 Kesimpulan

Penelitian ini menunjukkan bahwa model SVM mampu melakukan klasifikasi sentimen dengan baik, di mana tuning parameter TF-IDF dengan $\text{min_df} = 2$ dan $\text{max_features} = 2000$ menghasilkan F1-score tertinggi sebesar 68%. Kernel linear terbukti paling stabil dibandingkan kernel lainnya, sedangkan penerapan SMOTE berhasil menyeimbangkan data tetapi sedikit menurunkan akurasi dari 68% menjadi 66%. Meskipun demikian, SMOTE tetap membantu dalam mengatasi ketidakseimbangan kelas. Untuk penelitian selanjutnya, disarankan eksplorasi metode *balancing* lain seperti ADASYN, penerapan *word embedding* seperti Word2Vec atau BERT, serta perbandingan dengan model lain seperti Random Forest untuk meningkatkan performa klasifikasi. Dengan pengembangan lebih lanjut, diharapkan penelitian ini dapat memberikan hasil yang lebih akurat dan relevan dalam menganalisis sentimen publik terhadap kebijakan pendidikan di Indonesia.

Referensi

- [1] CNN, “Tingkat Kepuasan Publik atas Kinerja Menteri dari berbagai Survei,” *CNN Indonesia*, 2024. <https://www.cnnindonesia.com/nasional/20240110114905-617-1047602/tingkat-kepuasan-publik-atas-kinerja-menteri-dari-berbagai-survei>.
- [2] R. Deswandi Yahya, S. Adi Wibowo, and N. Vendyansyah, “Analisis Sentimen untuk Deteksi Ujaran Kebencian pada Media Sosial terkait Pemilu 2024 menggunakan Metode *Support Vector Machine*,” *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 2, pp. 1182–1189, 2024, doi: 10.36040/jati.v8i2.9076*.
- [3] Suswadi and M. Erkamim, “*Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine*,” *Ilk. J. Ilm., vol. 15, no. 3, pp. 427–435, 2023, doi: 10.33096/ilkom.v15i3.1610.427-435*.
- [4] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *EduTic - Sci. J. Informatics Educ., vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779*.
- [5] H. C. Husada and A. S. Paramita, “Analisis Sentimen pada Maskapai Penerbangan di Platform Twitter menggunakan Algoritma Support Vector Machine (SVM),” *Teknika, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311*.
- [6] N. T. Luchia *et al.*, “Analisis Sentimen Pengguna Twitter terhadap Aplikasi TikTok menggunakan Algoritma Naïve Bayes Clasifier,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy., no. September 2016, pp. 100–104, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas*.
- [7] G. R. Ramadhan and C. A. Sugianto, “Analisis Sentimen Ulasan Aplikasi Dana di Google Play Store menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform., vol. 8, pp. 9849–9857, 2024, doi: 10.36040/jati.v7i6.8178*.
- [8] R. W. Hardian, P. E. Prasetyo, U. Khaira, and T. Suratno, “Analisis Sentiment Kuliah Daring Di Media Sosial Twitter Selama Pandemi Covid-19 menggunakan Algoritma Sentistrength,” *MALCOM Indones. J. Mach. Learn. Comput. SCI., vol. 1, no. 2, pp. 138–143, 2021, doi: 10.57152/malcom.v1i2.15*.
- [9] A. E. Budiman and A. Widjaja, “Analisis Pengaruh Teks Preprocessing terhadap Deteksi Plagiarisme pada Dokumen Tugas Akhir,” *J. Tek. Inform. dan Sist. Inf., vol. 6, no. 3, pp. 475–488, 2020, doi: 10.28932/jutisi.v6i3.2892*.
- [10] V. W. D. Thomas and F. Rumaisa, “Analisis Sentimen Ulasan Hotel Bahasa Indonesia menggunakan Support Vector Machine dan TF-IDF,” *J. Media Inform. Budidarma, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218*.
- [11] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen berbasis Aspek pada Review Female Daily menggunakan TF-IDF dan Naïve Bayes,” *J. Media Inform. Budidarma, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845*.
- [12] A. Pralabaika, “Analisa Sentimen Transisi Kendaraan Konvensional ke Listrik dengan menerapkan Algoritma Text Mining dan Term Frequency Inverse Document Frequency (TF-IDF),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 7, pp. 25–33, 2024, doi: 10.30865/komik.v6i1.7918*.
- [13] M. P. Pulungan, A. Purnomo, and A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *J. Teknol. Inf. dan Ilmu Komput., vol. 10, no. 7, pp. 1493–1502, 2023, doi: 10.25126/jtiik.1077989*.
- [14] R. W. Pratiwi, S. F. H, D. Dairoh, D. I. Af'idah, Q. R. A, and A. G. F, “Analisis Sentimen pada Review Skincare Female Daily menggunakan Metode Support Vector Machine (SVM),” *J. Informatics, Inf. Syst. Softw. Eng. Appl., vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387*.
- [15] D. Anggraini and T. Sutabri, “Pengembangan Aplikasi Penyaringan Spam e-Mail menggunakan Teknik Machine Learning dengan Metode Support Vector Machines,” *IJM Indones. J. Multidiscip., vol. 2, pp. 106–114, 2024, [Online]. Available: https://journal.csspublishing/index.php/ijm*.
- [16] E. Suryati, Styawati, and A. A. Aldino, “Analisis Sentimen Transportasi Online menggunakan Ekstraksi Fitur Model Word2vec Text Embedding dan Algoritma Support Vector Machine

<http://sistemasi.ftik.unisi.ac.id>

- (SVM),” *J. Teknol. Dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>.
- [17] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, “Prediksi Penyakit Diabetes menggunakan *Algoritma Support Vector Machine (SVM)*,” *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [18] F. Darmawan, M. Joe, Y. I. Kurniawan, and L. Afuan, “Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada *Twitter* menggunakan *Support Vector Machine*,” *J. Eksplora Inform.*, vol. 13, no. 1, pp. 24–36, 2023, doi: 10.30864/eksplora.v13i1.854.
- [19] J. T. Kumalasari and A. Merdekawati, “Analisis Sentimen terhadap Program Merdeka Belajar Kampus Merdeka pada Sosial Media *Twitter* menggunakan *K-Means Clustering*, *Support Vector Machine (SVM)* dan *Syntethic Minority Oversampling Technique (SMOTE)*,” *SATIN Sains dan Teknol. Inf.*, vol. 9, no. 1, pp. 1–12, 2023.