

Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X

Elsa Triningsih*, M Afdal, Inggih Permana, Nesdi Evrilyan Rozanda

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia

Email: ^{1,*}12150321835@students.uin-suska.ac.id, ²m.afdal@uin-suska.ac.id, ³inggihermana@uin-suska.ac.id,

⁴nesdi.er@uin-suska.ac.id

Email Penulis Korespondensi: 12150321835@students.uin-suska.ac.id.

Submitted: 26/12/2024; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Pemerintah telah meluncurkan Program Makan Bergizi Gratis sebagai bagian dari upaya strategis untuk menurunkan angka stunting di Indonesia. Namun, program ini memunculkan berbagai kontroversi di kalangan masyarakat, terutama terkait besarnya alokasi anggaran yang dianggap memberatkan serta dampaknya terhadap sektor pendidikan dan stabilitas keuangan negara. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program tersebut dengan memanfaatkan data dari platform media sosial X (Twitter) sebanyak 2.400 data. Sentimen publik diklasifikasikan menjadi tiga kategori, yaitu positif, negatif, dan netral, menggunakan dua algoritma Machine Learning, yaitu Support Vector Machine (SVM) dan Random Forest. Selain itu, teknik SMOTE digunakan untuk menangani ketidakseimbangan data dalam proses pelatihan model. Hasil analisis menunjukkan bahwa sentimen negatif mendominasi sebesar 46%, dengan isu utama yang disoroti adalah tingginya alokasi anggaran dan dampaknya terhadap pendidikan. Dalam hal kinerja, algoritma SVM dengan SMOTE menghasilkan akurasi tertinggi sebesar 85,74%, mengungguli algoritma Random Forest yang hanya mencapai akurasi 81,53%.

Kata Kunci: Makan Bergizi Gratis; Analisis Sentimen; Machine Learning; Support Vector Machine; Random Forest

Abstract—The government has launched the Free Nutritious Meal Program as part of a strategic effort to reduce stunting in Indonesia. However, the program has generated a lot of controversy among the public, especially regarding the large budget allocation that is considered burdensome and its impact on the education sector and the country's financial stability. This study aims to analyze public sentiment towards the program by utilizing data from social media platform X (Twitter) as much as 2,400 data. Public sentiment is classified into three categories, namely positive, negative, and neutral, using two machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest. In addition, the SMOTE technique is used to handle data imbalance in the model training process. The analysis results showed that negative sentiments dominated at 46%, with the main issue highlighted being the high budget allocation and its impact on education. In terms of performance, the SVM algorithm with SMOTE produced the highest accuracy of 85.74%, outperforming the Random Forest algorithm which only achieved 81.53% accuracy.

Keywords: Free Nutritious Meal; Sentiment Analysis; Machine Learning; Support Vector Machine; Random Forest

1. PENDAHULUAN

Stunting merupakan masalah gizi yang menjadi fokus pemerintah Indonesia. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI) tahun 2023, target penurunan stunting pada tahun 2024 adalah sebesar 14%, sementara angka stunting pada tahun 2022 masih mencapai 21,6% [1]. Berbagai upaya penanganan stunting telah dilakukan oleh pemerintah Indonesia salah satunya adalah Program Makan Bergizi Gratis atau sebelumnya dikenal dengan Makan Siang Gratis. Program ini diusung oleh Presiden dan Wakil Presiden 2024 yang bertujuan untuk memberikan bantuan gizi berupa makan siang serta susu gratis di sekolah, pesantren, dan ibu hamil sebagai langkah pengentasan stunting [2].

Namun, program ini juga menuai kontroversi terutama pada sosial media seperti X terkait dengan besarnya anggaran yang dibutuhkan. Adapun besaran anggaran yang diperkirakan mencapai Rp450 Triliun. Selain itu, perhatian ini juga tertuju pada rencana penggunaan APBN pendidikan dan dana BOS, yang ditentang oleh Federasi Serikat Guru Indonesia (FSGI) [3]. Hal ini tentunya memunculkan kekhawatiran masyarakat mengenai potensi dampaknya terhadap biaya pendidikan, gaji guru, dan stabilitas keuangan negara. Meskipun program ini menawarkan manfaat potensial, tantangan besar dalam implementasinya menimbulkan keraguan apakah program ini dapat berjalan secara berkelanjutan dan memberikan manfaat jangka panjang [4]. Dengan berbagai isu yang muncul ini semakin memperkuat polarisasi opini publik, untuk itu analisis sentimen diperlukan untuk memberikan pemahaman bagaimana pandangan masyarakat terhadap kebijakan ini.

Analisis sentimen adalah proses komputasi untuk mengevaluasi pendapat dan emosi individu terhadap entitas, peristiwa, dan atribut yang terkait [5]. Tujuan utamanya adalah mengidentifikasi polaritas teks dalam dokumen untuk menilai apakah opini yang disampaikan bersifat negatif, positif, atau netral [6]. Dalam hal ini masyarakat memanfaatkan platform media sosial untuk mengekspresikan perasaan dan sentimen mereka tentang suatu situasi. Oleh karena itu, penting untuk menganalisis sentimen publik dan dinamikanya guna mengungkap wawasan tentang isu-isu terkini [7]. Data pada penelitian ini berasal dari sosial media X yaitu platform yang memungkinkan pengguna untuk berdiskusi topik terkini. Selain menjadi wadah ekspresi pengguna, X juga dianggap sumber data akurat untuk memahami pandangan masyarakat [8].

Dalam melakukan klasifikasi sentimen, penelitian ini menggunakan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *Random Forest*. Algoritma SVM dikenal efektif untuk klasifikasi [9]. Algoritma ini memisahkan

data menjadi dua atau lebih kelas dengan mencari hyperplane [10]. Sementara itu, Random Forest digunakan karena memiliki keunggulan dalam menghasilkan akurasi tinggi [11]. *Random Forest* ialah algoritma klasifikasi yang menggabungkan sejumlah pohon keputusan yang dilatih pada data yang dibagi secara acak [12].

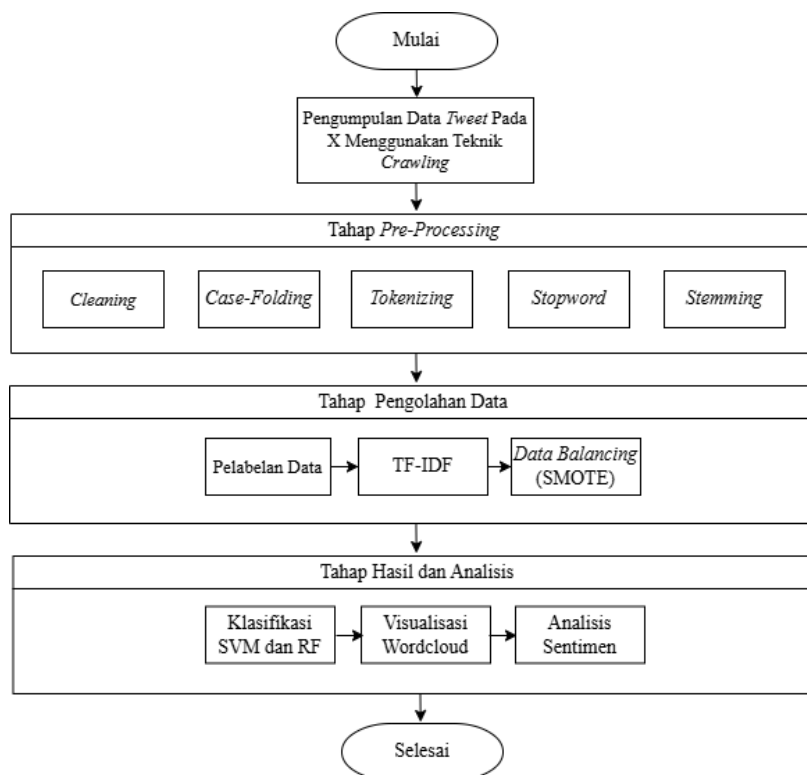
Adapun pada penelitian sebelumnya telah dilakukan oleh Ramadhani dkk (2024) untuk analisis sentimen makan siang gratis menggunakan algoritma SVM memiliki akurasi 75% sebagai akurasi terbaik dalam perbandingan 2 algoritma pada penelitiannya [13]. Penelitian yang dilakukan oleh Purwanti & Sugiyono (2024) dalam analisis sentimen makan siang gratis menggunakan algoritma SVM Hasil klasifikasi ini menunjukkan bahwa algoritma SVM dapat mengolah data dengan efektif dan menghasilkan akurasi tinggi sebesar 94% [14]. Penelitian oleh Sari & Suryono (2024) pada analisis sentimen *Metaverse* Komparasi kedua model dengan metode optimasi SMOTE menghasilkan akurasi tinggi, dengan 90% untuk SVM dan 91% untuk algoritma *Random Forest*. Hal ini menunjukkan bahwa, algoritma SVM dapat memprediksi sentimen lebih akurat dan memiliki tingkat kesalahan yang lebih rendah dibandingkan algoritma *Random Forest*. [15]. Penelitian oleh Fitri (2024) dalam analisis sentimen masyarakat terhadap pinjaman online yang dihasilkan algoritma Random Forest lebih unggul dalam menganalisis sentimen masyarakat terhadap pinjaman online, dengan akurasi 100%, dibandingkan SVM yang mencapai 99%. [16]. Penelitian yang dilakukan oleh Purba dkk (2024) dalam analisis sentimen debat calon presiden yang dihasilkan kedua algoritma tersebut memiliki kinerja yang lebih unggul dibandingkan algoritma lainnya [17].

Tujuan dari penelitian ini ialah menganalisis sentimen publik terhadap kebijakan Makan Bergizi Gratis guna memahami persepsi masyarakat apakah positif, netral atau negatif terhadap program makan bergizi gratis melalui pendekatan algoritma *machine learning*. Untuk mencapai tujuan tersebut, penelitian ini membandingkan kinerja dua algoritma klasifikasi, yaitu *Support Vector Machine* dan *Random Forest* dalam menganalisis sentimen opini masyarakat, khususnya pada platform X. Hasil analisis penelitian ini diharapkan dapat memberikan pemahaman dan gambaran terkait pandangan masyarakat terhadap program makan bergizi gratis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk mencapai tujuan penelitian, tahapan metodologi yang diterapkan dalam penelitian ini terdapat pada Gambar 1.



Gambar 1. Metode Penelitian

2.2 Tahap Pengumpulan data

Penelitian ini berfokus pada reaksi masyarakat di media sosial X. Data yang dikumpulkan berupa *tweet* dengan *keyword* dan berisi pendapat, komentar, atau informasi yang membahas dan berkaitan langsung dengan pelaksanaan Program Makan Bergizi Gratis pada platform X. Data yang digunakan ialah selama tahun 2024 pada periode 1 Januari 2024 hingga 1 Desember 2024 sebanyak 2400 *tweet* dengan menggunakan teknik *Crawling*.

2.2 Tahap Pre-Processing

Setelah data dikumpulkan maka selanjutnya masuk pada tahapan *pre-processing* dimulai dari proses *cleaning* untuk menghilangkan emoji, simbol dan tanda baca yang tidak penting dalam teks, *case-folding* mengubah huruf dalam teks menjadi *lowercase* atau huruf kecil agar lebih seragam, *tokenizing* memecah kalimat menjadi bagian-bagian kata, *stopword* menghilangkan setiap kata dari teks yang dianggap tidak ada makna, dan *stemming* yaitu menghapus imbuhan pada kata untuk mendapatkan bentuk dasarnya.

2.2 Tahap Pengolahan Data

Tahapan ini merupakan proses mengelola, memahami, dan mengkategorikan data teks. Adapun tahapannya sebagai berikut:

a. Pelabelan Data

Proses pelabelan data menjadi positif, negatif dan netral yang dilakukan manual oleh pakar pada data yang sebelumnya telah dilakukan proses *pre-processing*.

b. TF-IDF

Tahap ini digunakan untuk memberikan bobot pada setiap istilah dalam dokumen. Perhitungan bobot ini dilakukan dengan memanfaatkan modul *sklearn* dalam bahasa pemrograman *Python*

c. Data Balancing

proses *data balancing* dilakukan dengan menggunakan SMOTE yaitu metode *oversampling* untuk menyeimbangkan distribusi data antar kelas karena efektivitas analisis sentimen juga dipengaruhi oleh data set yang digunakan, yang dapat mempengaruhi cara penanganannya [19].

2.2 Tahap Hasil dan Analisis

a. Klasifikasi Menggunakan *Support Vector Machine* dan *Random Forest*

Proses klasifikasi menggunakan algoritma SVM dan Random forest dilakukan untuk mengidentifikasi sentimen positif, negatif dan netral. Dalam hal ini model dilatih menggunakan data latih dan dievaluasi menggunakan data uji. SVM telah banyak diterapkan dan terbukti efektif dalam analisis sentimen dan tugas penambangan teks [20]. Salah satu keunggulan SVM adalah kemampuannya untuk menentukan *hyperplane* optimal yang dapat memisahkan dua atau lebih kelas. *Hyperplane* yang memisahkan dua kelas secara optimal diperoleh melalui proses dengan memaksimalkan margin dan mengidentifikasi support vectors, yaitu titik data yang paling dekat dengan *hyperplane*. [22]. Dalam hal ini berbagai kernel digunakan seperti *linear*, *RBF*, *polynomial*, dan *sigmoid* untuk mengubah level data dimensi yang berbeda tergantung dari kernel yang digunakan. Persamaan dari masing-masing kernel *Support Vector Machine* disajikan Tabel 1 [23]:

Tabel 1. Rumus Kernel pada SVM

Kernel	Persamaan
<i>Linear</i>	$K(x_i, x) = x^T x$
<i>Polynomial</i>	$K(x_i, x) = (\gamma \cdot x^T x + r)^p, \gamma > 0$
<i>RBF</i>	$K(x_i, x) = \exp(-\gamma x_i - x ^2), \gamma > 0$
<i>Sigmoid</i>	$K(x_i, x) = \tanh(\gamma x^T x + r)$

Sementara, Random Forest bekerja dengan membangun beberapa pohon keputusan dari sampel data dan subset fitur yang dipilih secara acak. Hasil akhir diperoleh dengan menggabungkan prediksi semua pohon, menggunakan mayoritas suara untuk klasifikasi, sehingga meningkatkan akurasi dan mengurangi overfitting [24]. Klasifikasi *Random Forest* dipilih terutama karena kemampuannya yang luar biasa dalam menangani kompleksitas data berdimensi tinggi, termasuk data dengan banyak fitur dan variabel saling terkait sehingga menjadikannya pilihan yang populer dalam melakukan klasifikasi [25]. Dalam masalah klasifikasi, setiap pohon menghasilkan prediksi, dan kelas dengan jumlah suara terbanyak dipilih sebagai output akhir. Proses ini dijelaskan menggunakan rumus voting pada persamaan 1:

$$\hat{y} = \text{mode} \{T_1(x), T_2(x), \dots, T_n(x)\} \quad (1)$$

Pada klasifikasi, prediksi akhir (\hat{y}) ditentukan melalui voting mayoritas dari prediksi semua decision tree. Setiap $T_i(x)$ merepresentasikan prediksi decision tree ke- i untuk input, dan (n) menunjukkan jumlah total decision tree. Prediksi akhir adalah kelas yang paling sering dipilih (modus) oleh seluruh decision tree.

Dalam penerapan parameter kedua model, penelitian ini menggunakan Gridserch sebagai tuning parameter untuk menghasilkan parameter yang paling optimal untuk menghasilkan akurasi yang tinggi.

b. Visualisasi *Wordcloud*

Hasil klasifikasi sentimen kemudian divisualisasikan menggunakan *library Wordcloud*. *Word Cloud* adalah representasi data yang menampilkan kumpulan kata-kata penting berdasarkan frekuensi kemunculannya dalam data.

c. Analisis Sentimen

Setelah data divisualisasikan data kemudian dianalisis berdasarkan kata yang sering muncul untuk mengidentifikasi opini publik terhadap program sehingga dapat memberikan gambaran yang komprehensif mengenai persepsi masyarakat terhadap program makan bergizi gratis.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dikumpulkan pada penelitian ini berasal dari komentar atau *tweet* masyarakat tentang program makan bergizi gratis yang terdapat pada platform X. Proses pengambilan data dilakukan dengan menggunakan teknik pengambilan data otomatis yaitu dengan teknik *Crawling* pada sosial media X, yang diimplementasikan dengan bahasa pemrograman Python. Tweet yang diambil dengan menggunakan hastag #Makanbergizigratis dan #Makansianggratis. Total jumlah data yang dikumpulkan adalah 2400 komentar. Adapun data yang diambil ialah data pada tahun 2024 (Januari - Desember 2024).

3.2 Proses Pre-Processing

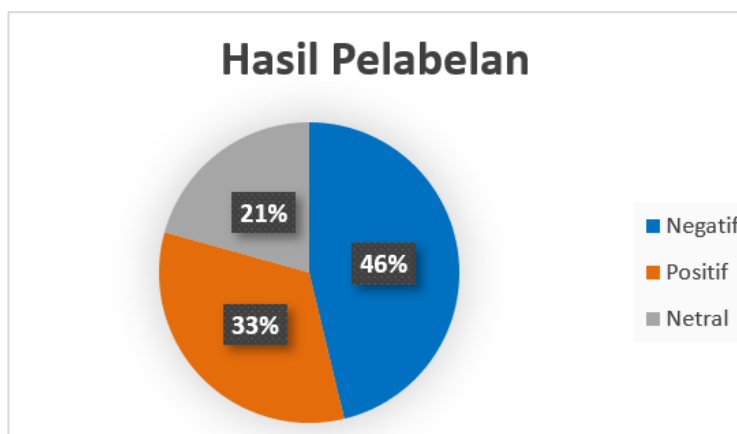
Setelah data dikumpulkan, dilakukan tahapan *pre-processing* yang meliputi *cleaning* untuk menghilangkan karakter atau simbol tidak relevan, *case-folding* untuk mengubah teks menjadi huruf kecil, proses *tokenizing* untuk menguraikan teks ke bentuk satuan kata, *stopword removal* untuk menghapus kata-kata minim makna, dan *stemming* untuk mengembalikan kata ke bentuk dasar.

Table 2. Hasil Pre-Processing Data

Proses	Hasil
Data awal	Mari Dukung Program Unggulan Prabowo - Gibran untuk mewujudkan masyarakat Indonesia yang lebih sejahtera #PrabowoSubianto #PrabowoGibran #MakanBergiziGratis #ProgramUnggulanPrabowo #IndonesiaEmas2045 https://t.co/AbRZi50gKF
Cleaning	Mari Dukung Program Unggulan Prabowo Gibran untuk mewujudkan masyarakat Indonesia yang lebih sejahtera
Case Folding	mari dukung program unggulan prabowo gibran untuk mewujudkan masyarakat indonesia yang lebih sejahtera
Tokenizing	[mari,dukung,program,unggulan,prabowo,gibran,untuk,mewujudkan,masyarakat,indonesia,yang,lebih,sejahtera]
Stopword	[mari,dukung,program,unggulan,prabowo,gibran,mewujudkan,masyarakat,indonesia,sejahtera]
Stemming	mari dukung program unggul prabowo gibran wujud masyarakat indonesia sejahtera

3.1 Proses Pelabelan

Dari hasil pelabelan manual oleh pakar sebanyak 2400 data, didapatkan pada komentar pada sosial media X memiliki 1110 sentimen negatif, 794 sentimen positif, dan 496 sentimen netral. persentase keberhasilan pelabelan setiap kategori dapat dilihat secara rinci pada Gambar 2.



Gambar 2. Persentasi Hasil Pelabelan

Analisis pada Gambar 2. memperlihatkan sentimen tertinggi ada pada sentimen negatif yaitu sebesar 46%, sedangkan sentimen positif sebesar 33%, dan sentimen terendah pada netral sebesar 21%.

3.2 TF-IDF

Tahap pembobotan TF-IDF untuk mendapatkan bobot dari kata. Hasil yang diperoleh setelah penerapan proses ini terdapat pada Tabel 3.

Tabel 3. Pembobotan TF-IDF

No.	anak	gizi	gratis	makan	program	...	Zakat
1	0.0	0.07	0.0	0.0	0.14	...	0.0
2	0.0	0.11	0.06	0.06	0.11	...	0.0
3	0.0	0.07	0.08	0.08	0.16	...	0.0
...
500	0.0	0.0	0.08	0.08	0.0	...	0.0

3.3 Data Balancing (SMOTE)

Pada penelitian ini, teknik SMOTE diterapkan untuk menyeimbangkan distribusi kelas dalam dataset dengan cara menghasilkan data sintesis untuk kelas yang kurang terwakili, dengan tujuan untuk meningkatkan kinerja algoritma klasifikasi. Hasil dari penerapan SMOTE terdapat pada Tabel 4.

Tabel 4. Hasil Penerapan SMOTE

Penerapan	Negatif	Positif	Netral	Total Data
Tanpa SMOTE	1110	794	496	2400
SMOTE	1110	1110	1110	3330

3.4 Klasifikasi Algoritma SVM dan *Random Forest*

3.4.1 Pembagian Data

Setelah dataset dilakukan proses *pre-processing* data, pelabelan, dan perhitungan TF-IDF, tahapan selanjutnya ialah melakukan proses klasifikasi menggunakan model *SVM* dan *Random Forest*. Sebelum tahapan klasifikasi dilakukan, Data dipisahkan menjadi dua bagian, yaitu data latih dan data uji, menggunakan metode *hold-out validation*. Dalam penelitian ini, pembagian data dilakukan dengan proporsi 80:20, yang berarti 80% data dimanfaatkan sebagai data pelatihan untuk pembelajaran model, Sedangkan 20% dari data digunakan untuk data uji dalam mengevaluasi performa model.

3.4.2 Hasil Algoritma Support Vector Machine

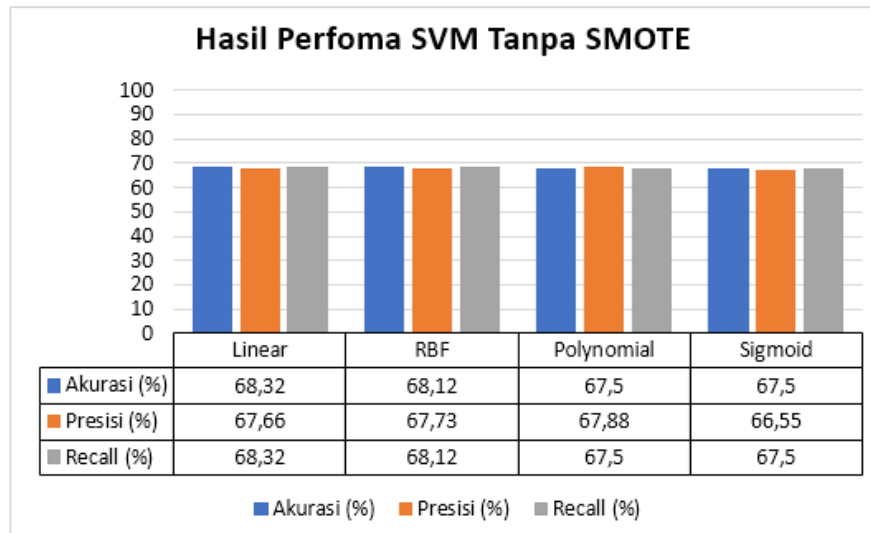
Langkah selanjutnya adalah mengevaluasi kinerja model Support Vector Machine (SVM) yang telah dilatih dengan menggunakan dataset uji. Evaluasi ini melibatkan berbagai konfigurasi kernel, seperti kernel *linear*, RBF, *polynomial*, dan *sigmoid*, untuk menentukan konfigurasi yang memberikan hasil terbaik dan optimal. Evaluasi dilakukan pada dua skenario. Pertama, dengan dataset asli tanpa menggunakan teknik SMOTE, dan kedua, dengan dataset yang telah diperbaiki dengan SMOTE untuk mengatasi masalah ketidakseimbangan kelas. Parameter yang digunakan adalah Cost (C) dengan nilai (0.01, 0.1, 1, 10, 100), gamma (γ) dengan nilai (0.001, 0.01, 0.1, 1), dan d (degree) dengan nilai (2, 3, 4, 5, 8) Untuk menilai kinerja model pada setiap konfigurasi kernel, berbagai metrik evaluasi seperti *accuracy*, *precision*, dan *recall* digunakan untuk memberikan pemahaman yang menyeluruh mengenai kinerja model. Hasil dari matrik evaluasi untuk setiap konfigurasi kernel, baik yang diuji tanpa SMOTE maupun dengan SMOTE, dapat terdapat pada Tabel 5.

Tabel 5. Perfoma Algoritma SVM

Kategori	Kernel	Cost (C)	Gamma (γ)	Degree (d)	Akurasi	Presisi	Recall
Tanpa SMOTE	Linear	1	-	-	68,32%	67,66%	68,32%
	RBF	10	1	-	68,12%	67,73%	68,12%
	Polynomial	10	1	2	67,50%	67,88%	67,50%
	Sigmoid	10	0,1	-	67,50%	66,55%	67,50%
Dengan SMOTE	Linear	10	-	-	78,68%	78,78%	79,68%
	RBF	10	1	-	85,74%	85,92%	85,74%
	Polynomial	1	1	2	80,03%	80,32%	80,03%
	Sigmoid	100	0,1	-	78,68%	78,78%	78,68%

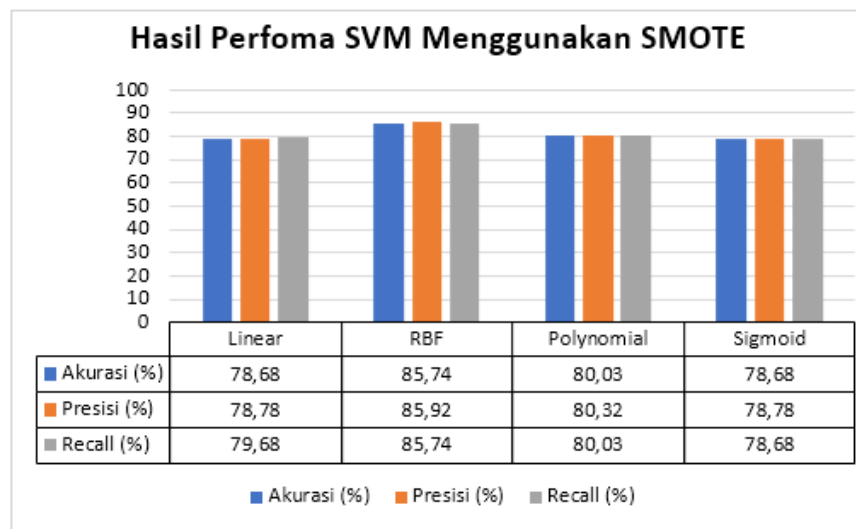
Berdasarkan tabel performa SVM, penggunaan SMOTE secara signifikan meningkatkan akurasi, presisi, dan recall untuk semua kernel. Tanpa SMOTE, performa tertinggi dicapai oleh kernel *Linear* dengan akurasi dan recall 68,32% dan presisi 67,66%. Setelah menggunakan SMOTE, kernel RBF menjadi yang terbaik dengan akurasi dan recall sebesar 85,74% serta presisi 85,92%, menunjukkan kemampuannya menangani data yang tidak seimbang. kernel *Linear* dan *Sigmoid* juga mengalami peningkatan performa ke 78,68%, sementara *Polynomial* mencapai akurasi 80,03% dengan presisi 80,32%. Secara keseluruhan, SMOTE berhasil memperbaiki performa SVM, dengan RBF

kernel sebagai pilihan optimal. Adapun hasil proyeksi perbandingan akurasi pada masing masing kernel tanpa SMOTE terdapat pada Gambar 3.



Gambar 3. Hasil Perfoma SVM Tanpa SMOTE

Sementara untuk hasil performa SVM dengan menggunakan SMOTE dapat terdapat pada Gambar 4.



Gambar 4. Hasil Perfoma SVM Menggunakan SMOTE

3.4.3 Hasil Algoritma Random Forest

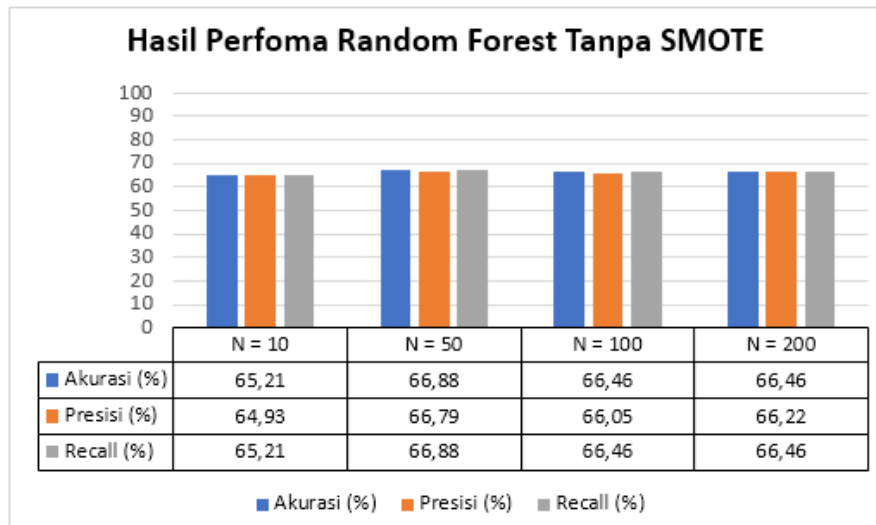
Langkah berikutnya adalah mengevaluasi kinerja model *Random Forest*. Evaluasi ini bertujuan untuk mengukur kinerja model dengan memanfaatkan konfigurasi, seperti jumlah pohon keputusan ($n_{estimators}$). Parameter yang digunakan adalah max_depth dengan nilai (10, 20, 30, 40, None), $min_samples_split$ dengan nilai (2, 5, 10, 20), min_sample_leaf dengan nilai (1, 2, 4, 8), dan $bootstrap$ dengan nilai (true, false). Evaluasi juga dilakukan pada dua skenario yaitu tanpa SMOTE dan Menggunakan SMOTE serta menggunakan matriks evaluasi, seperti tingkat *accuracy*, *precision*, dan *recall* untuk mengukur performa *Random Forest*. Hasil kinerja Random Forest, baik tanpa maupun dengan penerapan SMOTE pada setiap nilai N ($n_{estimator}$), terdapat pada Tabel 6.

Tabel 6. Hasil Perfoma *Random Forest*

Kategori	Nilai N	max_depth	min_samples_split	min_sample_leaf	bootstrap	Akurasi	Presisi	Recall
Tanpa SMOTE	N = 10	30	2	1	false	65,21%	64,93%	65,21%
	N = 50	none	10	2	false	66,88%	66,79%	66,88%
	N = 100	none	10	1	false	66,46%	66,05%	66,46%
	N = 200	none	5	1	false	66,46%	66,22%	66,46%
Dengan SMOTE	N = 10	40	5	4	true	79,43%	75,69%	79,43%
	N = 50	none	5	1	true	80,03%	80,32%	80,03%

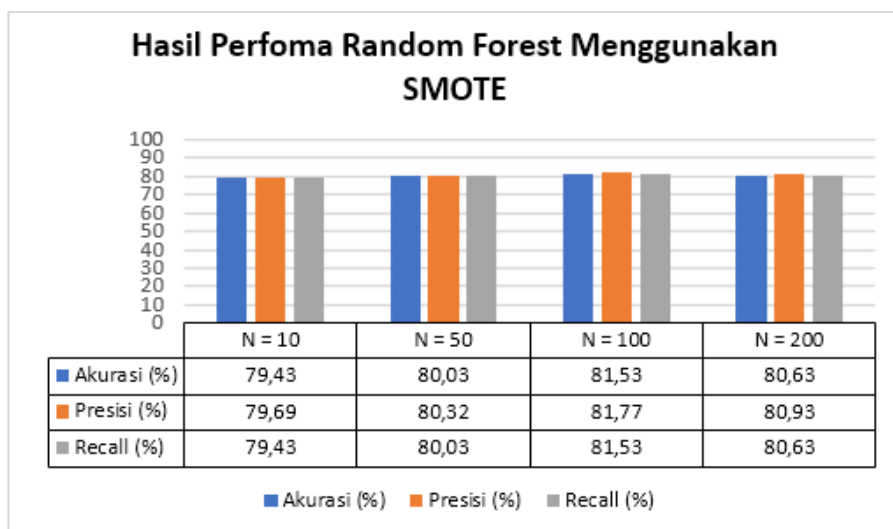
Kategori	Nilai N	max_depth	min_samples_split	min_sample_leaf	bootstrap	Akurasi	Presisi	Recall
	N = 100	none	2	1	false	81,53%	81,77%	80,93%
	N = 200	none	5	1	false	80,63%	79,51%	80,63%

Proyeksi perbandingan performa dalam grafik diagram batang pada masing masing nilai N tanpa menggunakan SMOTE terdapat pada Gambar 5.



Gambar 5. Perfoma *Random Forest* Tanpa Menggunakan SMOTE

Adapun hasil perfoma *Random Forest* dengan menerapkan SMOTE terdapat pada Gambar 6.

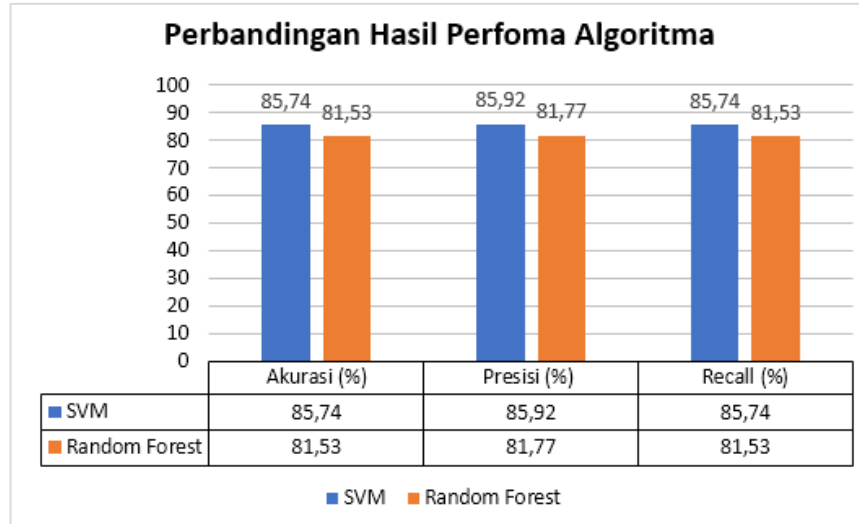


Gambar 6. Perfoma *Random Forest* Menggunakan SMOTE

Kedua gambar tersebut menunjukkan peningkatan performa model setelah menggunakan teknik SMOTE, dengan nilai terbaik tercapai pada N = 100 dengan akurasi 81,53%, presisi 81,77%, recall 81,53%. Peningkatan jumlah n_estimators setelah N = 100 tidak memberikan peningkatan yang terlalu signifikan, sehingga nilai N = 100 dapat dianggap optimal. Hal ini menunjukkan bahwa SMOTE efektif dalam mengatasi ketidakseimbangan kelas pada data, sehingga model lebih mampu mengenali pola data, terutama untuk kelas minoritas, serta mengurangi kesalahan prediksi.

3.4.4 Hasil Perbandingan Klasifikasi

Setelah melalui proses klasifikasi menggunakan model SVM dan *Random Forest*, diperoleh performa terbaik pada *accuracy*, *precision*, dan *recall* pada tiap model. Hasil dari perbandingan kedua algoritma menunjukkan bahwa model SVM memiliki nilai akurasi dan recall sebesar 85,74% dan presisi 85,92%. Sementara itu, model *Random Forest* hanya mencapai akurasi dan *recall* sebesar 81,53%, dan presisi 81,77%. Perbandingan hasil model klasifikasi dalam diagram batang terdapat pada Gambar 7.



Gambar 7. Perbandingan Hasil Algoritma

Hasil pada grafik di atas menunjukkan model SVM mempunyai accuracy yang lebih tinggi dari model Random forest. Tingkat akurasi ini menunjukkan seberapa akurat model dalam mengklasifikasikan data sentimen dengan benar. Pada nilai precision, SVM juga memiliki nilai yang lebih tinggi dibanding Random Forest, yang mengindikasikan bahwa model tersebut lebih andal dalam mengklasifikasikan data sentimen. Sementara itu, pada matriks recall menunjukkan SVM mempunyai nilai yang lebih tinggi daripada Rndom forst. Berdasarkan rata-rata recall yang diperoleh, dapat disimpulkan bahwa kedua model mampu mengklasifikasikan data uji dengan baik.

3.5 Visualisasi Data Sentimen

Visualisasi data bertujuan untuk mengidentifikasi topik yang sering dibahas oleh pengguna X. Dengan memanfaatkan teknik ini, kata yang tersembunyi dalam ribuan komentar dapat diungkap secara lebih jelas dan ringkas sehingga dapat menghasilkan informasi penting terhadap kebijakan Makan Begizi Gratis. Visualisasi dari 150 kata yang paling sering muncul dari komentar X terhadap Program Makan Bergizi Gratis pada tiap kelas sentimen terdapat pada Gambar 8, 9, dan 10.



Gambar 8. Hasil Visualisasi Sentimen Positif



Gambar 9. Hasil Visualisasi Sentimen Negatif

SMOTE untuk mengatasi ketidakseimbangan data terbukti efektif dalam mengelola data secara optimal. Hasil pengujian menunjukkan bahwa SVM dengan kernel RBF memberikan performa terbaik dengan akurasi tertinggi yaitu 85,74%, sementara *Random Forest* hanya mencapai akurasi 81,53%.

REFERENCES

- [1] Kementerian Kesehatan RI, “Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%.” Accessed: Dec. 24, 2024. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230125/3142280/prevalensi-stunting-di-indonesia-turun-ke-216-dari-244/>
- [2] Tundo and D. N. Rachmawati, “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Terhadap Program Makan Siang Gratis,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 2925–2939, Sep. 2024, doi: 10.35870/jimik.v5i3.978.
- [3] N. A. Rahmawati, S. A. Prasetyo, and M. W. Ramadhani, “Memetakan Visi Prabowo Gibran Pada Masa Kampanye Dalam Prespektif Pembangunan,” *WISSEN: Jurnal Ilmu Sosial dan Humaniora*, vol. 2, no. 3, pp. 97–120, Jun. 2024, doi: 10.62383/wissen.v2i3.176.
- [4] P. A. Maharani, A. R. Namira, and T. V. Chairunnisa, “Peran Makan Siang Gratis Dalam Janji Kampanye Prabowo Gibran Dan Realisasinya,” *Journal Of Law And Social Society*, vol. 1, no. 1, pp. 1–10, Jun. 2024, doi: 10.70656/jolasos.v1i1.79.
- [5] C. Steven and W. Wella, “The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter,” *IJNMT (International Journal of New Media Technology)*, vol. 7, no. 2, pp. 102–110, Dec. 2020, doi: 10.31937/ijnmt.v7i2.1732.
- [6] F. R. B. Kahi, A. Talakua, and R. Reynaldi, “Analisis Sentimen Masyarakat Di Twitter Terhadap Pemerintahan Anies Baswedan Menggunakan Metode Naive Bayes Classifier,” *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 324–336, Apr. 2024, doi: 10.33395/jmp.v13i1.13636.
- [7] M. Iqbal, M. Afdal, and R. Novita, “Implementasi Algoritma Support Vector Machine Untuk Analisa Sentimen Data Ulasan Aplikasi Pinjaman Online di Google Play Store,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1244–1252, Jul. 2024, doi: 10.57152/malcom.v4i4.1435.
- [8] A. Sitanggang, Y. Umidah, Y. Umidah, R. I. Adam, and R. I. Adam, “Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naive Bayes,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4902.
- [9] S. D. Wahyuni and R. H. Kusumodestoni, “Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting,” *Bulletin of Information Technology (BIT)*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
- [10] A. M. Siregar, “Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM),” *Faktor Exacta*, vol. 16, no. 3, Oct. 2023, doi: 10.30998/faktorexacta.v16i3.16703.
- [11] E. Fitri, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *Jurnal Transformatika*, vol. 18, no. 1, p. 71, Jul. 2020, doi: 10.26623/transformatika.v18i1.2317.
- [12] A. Wandani, “Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 651–665, Sep. 2021.
- [13] 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 Naive Bayes Classifier dan Support Vector Machine,” *Technology and Science (BITS)*, vol. 6, no. 3, 2024, doi: 10.47065/bits.v6i3.6188.
- [14] Z. Purwanti and Sugiyono, “Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 3065–3079, Sep. 2024, doi: 10.35870/jimik.v5i3.1001.
- [15] P. K. Sari and R. R. Suryono, “Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse,” *Jurnal Mnemonic*, vol. 7, no. 1, pp. 31–39, Feb. 2024, doi: 10.36040/mnemonic.v7i1.8977.
- [16] D. A. Fitri, “Komparasi Algoritma Random Forest Classifier Dan Support Vector Machine Untuk Sentimen Masyarakat Terhadap Pinjaman Online Di Media Sosial,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2018–2029, 2024, doi: 10.29100/jupi.v9i4.5608.
- [17] R. W. Purba, A. A. Waskita, and M. Makshun, “Analisis Sentimen Opini Debat Calon Presiden Dengan Menggunakan Classifier Machine Learning (Studi Kasus : Pada Data Twitter 2024),” *Infotech: Journal of Technology Information*, vol. 10, no. 2, pp. 221–232, Nov. 2024, doi: 10.37365/jti.v10i2.300.
- [18] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjee.v5i1.16830.
- [19] W. Widayat, “Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 1018, Jul. 2021, doi: 10.30865/mib.v5i3.3111.
- [20] E. Miranda, R. A. Elias, T. M. Kibitiah, and A. Permana, “Indonesia China Trade Relations, Social Media and Sentiment Analysis: Insight from Text Mining Technique,” in *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, IEEE, Oct. 2021, pp. 334–339. doi: 10.1109/ICCSAI53272.2021.9609735.
- [21] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, “Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF,” in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/ICDABI51230.2020.9325685.
- [22] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *Journal of Information System Management (JOISM)*, vol. 3, no. 1, pp. 1–7, Jan. 2021, doi: 10.24076/JOISM.2021v3i1.341.
- [23] 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, Feb. 2021, doi: 10.34148/teknika.v10i1.311.



- [24] M. K. Suryadi, R. Herteno, S. W. Saputro, M. R. Faisal, and R. A. Nugroho, "Comparative Study of Various Hyperparameter Tuning on Random Forest Classification With SMOTE and Feature Selection Using Genetic Algorithm in Software Defect Prediction," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 137–147, Mar. 2024, doi: 10.35882/jeeemi.v6i2.375.
- [25] Y. F. Zamzam, T. H. Saragih, R. Herteno, Muliadi, D. T. Nugrahadi, and P.-H. Huynh, "Comparison of CatBoost and Random Forest Methods for Lung Cancer Classification using Hyperparameter Tuning Bayesian Optimization-based," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 125–136, Mar. 2024, doi: 10.35882/jeeemi.v6i2.382.