OPTIMASI ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) MENGGUNAKAN TEKNIK SMOTE UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP DAMPAK PENGGUNAAN *ARTIFICIAL INTELLIGENCE* (AI)

Khurotul Aen, Fitri Ayuning Tyas, Hidayatur Rakhmawati
Teknik Informatika, Universitas Muhammadiyah Brebes
Jl. Pangeran Diponegoro Grengseng No.184, Taraban, Kec. Paguyangan,
Kabupaten Brebes, Jawa Tengah, Indonesia
kurotaen05@gmail.com

ABSTRAK

Perkembangan teknologi Artificial Intelligence (AI) memicu beragam opini publik, terutama di media sosial. Analisis sentimen menjadi pendekatan penting untuk memahami persepsi masyarakat, namun sering menghadapi tantangan berupa ketidakseimbangan data antar kategori sentimen. Penelitian ini bertujuan mengoptimalkan algoritma *Support Vector Machine* (SVM) dalam klasifikasi sentimen terhadap opini publik mengenai AI dengan menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Metode yang digunakan adalah eksperimen, data diambil dari media sosial X dan diklasifikasikan ke dalam tiga kategori: positif, negatif, dan netral. Tahapan preprocessing meliputi *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Pembobotan kata dilakukan menggunakan TF-IDF, sedangkan klasifikasi menggunakan SVM dengan kernel *Linear*, RBF, dan *Polynomial*. Validasi dilakukan menggunakan *K-Fold Cross Validation* dan pengujian model menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa SMOTE berhasil meningkatkan akurasi SVM, terutama pada kernel RBF dan *Polynomial*. Rata-rata akurasi kernel RBF meningkat dari 76,56% menjadi 81,22%, dan kernel *Polynomial* dari 71,89% menjadi 78,00%. *Kernel* Linear tetap stabil pada kisaran 80%. Teknik SMOTE terbukti efektif dalam meningkatkan akurasi dan mengatasi ketidakseimbangan data dalam analisis sentimen terhadap penggunaan AI.

Kata kunci : *Artificial Intelligence*, Analisis Sentimen, SMOTE, *Support Vector Machine*

1. PENDAHULUAN

Di era digital, *artificial intelligence* AI) berkembang pesat dan semakin berpengaruh dalam berbagai aspek kehidupan. AI memberikan manfaat, seperti membantu pendidikan [1], tetapi juga menimbulkan risiko, seperti penyalahgunaan teknologi *deepfake* untuk menyebarkan informasi palsu [2]. Survei SAS dan IDP Asia/Pasifik (2008) mencatat peningkatan adopsi AI di Asia Tenggara, dari 8% menjadi 14% [3]. Perkembangan ini memicu diskusi luas di media sosial, terutama di platform X (*Twitter*), tempat publik mengungkapkan berbagai opini mengenai manfaat dan risiko AI.

Tingginya volume unggahan membuat pemahaman terhadap opini publik menjadi tantangan. Setiap pengguna memiliki sudut pandang yang berbeda, mulai dari dukungan penuh terhadap perkembangan AI hingga kekhawatiran atas dampaknya. Dalam hal ini, analisis sentimen menjadi metode efektif untuk mengklasifikasikan opini tersebut. Analisis sentimen merupakan metode untuk mengukur opini individu terhadap suatu topik, produk, layanan, atau peristiwa tertentu berdasarkan tingkat kesepakatan yang mereka ungkapkan [4]. Analisis sentimen berperan dalam membangun sistem yang dapat mengevaluasi, mengenali, serta mengekspresikan pendapat dalam bentuk teks [5]. Untuk menangani data dalam jumlah besar dari media sosial, *text mining* digunakan sebagai pendekatan sistematis yang mendukung klasifikasi sentimen secara lebih akurat.

Text mining merupakan salah satu teknik yang dapat dimanfaatkan untuk mengklasifikasikan teks dalam analisis sentimen [6]. Salah satu algoritma yang umum digunakan adalah *Support Vector Machine* (SVM), yang mampu menganalisis area yang membedakan setiap kategori polaritas dalam teks, yang kemudian diklasifikasikan ke dalam sentimen positif, netral dan negatif [7]. SVM bekerja dengan menentukan *hyperplane* terbaik untuk memisahkan data, baik dalam ruang dua dimensi maupun dimensi lebih tinggi [8]. Dalam implementasinya, SVM dapat menggunakan berbagai jenis kernel, seperti *linear*, *polynomial*, *radial basis function* (RBF), dan *sigmoid* [9]. Kemampuan SVM dalam klasifikasi teks membuatnya menjadi algoritma utama dalam penelitian analisis sentimen, termasuk untuk memahami persepsi publik terhadap AI.

Beberapa penelitian menunjukkan bahwa SVM unggul dibandingkan algoritma lain. Dalam penelitian yang dilakukan Bhatara [10], SVM mencapai akurasi 85%, lebih tinggi dibandingkan *Naïve Bayes* yang hanya memperoleh 83%, dengan peningkatan akurasi setelah penerapan SMOTE. Pada penelitian yang dilakukan Puspita [11], SVM menunjukkan akurasi tertinggi sebesar 99% dibandingkan *Decision Tree* (98%) dan *Naïve Bayes* (91%) dalam analisis *tweet* terkait kebijakan makan gratis, dengan SMOTE digunakan untuk menyeimbangkan distribusi kelas. Sementara itu, penelitian oleh Iskandar dan Nataliani [12] menunjukkan bahwa SVM memperoleh akurasi sebesar 96,43%, lebih unggul dibandingkan *Naïve*

Bayes (83,54%) dan k-NN (59,68%), serta menunjukkan performa yang lebih optimal setelah penerapan SMOTE dalam mengatasi ketidakseimbangan data.

Berdasarkan tiga penelitian sebelumnya yang telah dilakukan, algoritma SVM terbukti efektif dalam menganalisis opini publik karena memiliki ketahanan tinggi, kemampuan generalisasi yang baik, serta akurasi yang stabil, bahkan dengan data latih terbatas[6]. Keunggulan ini menjadikannya pilihan tepat dalam analisis sentimen terhadap dampak penggunaan AI. Meski demikian, SVM masih menghadapi tantangan, terutama dalam menangani ketidakseimbangan data [13]. Ketidakseimbangan jumlah sampel antar kelas membuat model cenderung bias terhadap kelas mayoritas dan menurunkan akurasi [14]. Untuk mengatasi hal ini, penelitian ini menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) guna menyeimbangkan distribusi kelas sebelum klasifikasi. SMOTE meningkatkan jumlah sampel minoritas secara sintetis agar model dapat belajar lebih optimal dan memberikan hasil klasifikasi yang lebih akurat dan seimbang.

Berdasarkan latar belakang di atas, peneliti tertarik untuk meneliti "Optimasi Algoritma *Support Vector Machine* (SVM) dengan SMOTE untuk Analisis Sentimen terhadap Dampak Penggunaan *Artificial Intelligence* (AI)." Penelitian ini bertujuan untuk mengoptimalkan SVM dengan *Synthetic Minority Over-sampling Technique* (SMOTE) guna menangani ketidakseimbangan data.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian sebelumnya yang dijadikan referensi dalam penelitian ini adalah sebagai berikut.

Tabel 1. Penelitian Terdahulu

No	Penulis	Judul	Hasil Penelitian
1	Wahyu Adi Prabowo, Fitriani Azizah	*Sentiment Analysis for Detecting Cyberbullying Using* TF-IDF *and* SVM (2021)	Penelitian ini berhasil memperoleh akurasi 93%, presisi 95%, serta *recall* 97%.
2	Arya Damar Pratama, Hendry	Analisis Sentimen Masyarakat terhadap Penggunaan *Chatgpt* Menggunakan Metode SVM (2024)	Hasil penelitian menunjukkan bahwa metode SVM menghasilkan akurasi sebesar 87,6%.
3	Emi Yuspita, Ryan Randy Suryono	Perbandingan Berbagai Metode Klasifikasi Teks untuk Sentimen Kebijakan Makan Gratis di Indonesia (2024)	Hasil penelitian menunjukkan SVM mencapai akurasi 99% setelah SMOTE, *Decision Tree* 98%, dan *Naïve Bayes* turun dari 92% ke 91%.
4	Jessica Widyadhana Iskandar, Yessica Nataliani	Perbandingan *Naïve Bayes*, SVM, dan k-NN untuk Analisis Sentimen *Gadget* Berbasis Aspek (2021)	Hasil penelitian ini SVM memiliki akurasi tertinggi yaitu 96.43%, melampaui *Naïve Bayes* sebesar 83.54% dan k-NN sebesar 59.68%.
5	Farhan Nufairi, Nunik Pratiwi, Fariz Herlando	Analisis Sentimen pada Ulasan Aplikasi *Threads* di *Google Play Store* Menggunakan Algoritma SVM (2024)	Hasil penelitian memperlihatkan bahwa algoritma SVM memperoleh akurasi sebesar 88%.

2.2. Landasan Teori

a. Analisis Sentimen

Analisis sentimen merupakan teknik pemrosesan teks yang digunakan untuk memahami makna dari suatu pernyataan atau opini dalam sebuah teks [15]. [16]. Tujuan dari analisis ini adalah mengekstraksi atribut dalam teks atau komentar untuk memahami respons yang terkandung di dalamnya, kemudian mengklasifikasikan ke dalam kategori positif, negatif, atau netral [8].

b. *Artificial Intelligence*

Istilah *Artificial Intelligence* pertama kali diperkenalkan pada tahun 1956 oleh John McCarthy, seorang profesor di Dartmouth College yang berlokasi di Hanover, New Hampshire, Amerika Serikat. Menurut Kusumadewi (2003), "Kecerdasan buatan atau *artificial intelligence* merupakan salah satu bagian ilmu komputer yang membuat agar mesin (komputer) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia".

c. *Support Vector Machine* (SVM)

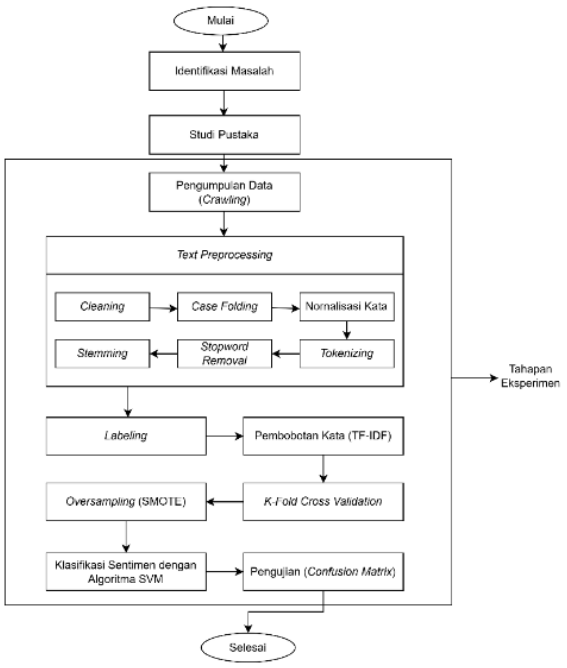
SVM adalah metode yang digunakan untuk menganalisis data dan mengenali pola dalam tugas klasifikasi serta regresi. Algoritma ini berfungsi menentukan *hyperplane* optimal yang memisahkan dua kelas data dengan cara mengurangi kesalahan klasifikasi dan memperbesar margin geometrisnya [17]. Model SVM dapat diterapkan dengan berbagai jenis kernel, di antaranya *linear, polynomial, Radial Basis Function* (RBF), dan *sigmoid* [18].

d. SMOTE (*Synthetic Minority Over-sampling Technique)*

Menurut [20] SMOTE merupakan teknik yang digunakan untuk menyeimbangkan dataset dengan menyamakan jumlah sampel pada setiap kategori. SMOTE bekerja dengan melakukan *oversampling* pada kelas minoritas dengan cara menghasilkan data sintetis sehingga jumlahnya setara dengan kelas mayoritas. Teknik ini membantu mengatasi ketidakseimbangan data dengan menambahkan sampel baru secara acak berdasarkan pola yang ada dalam dataset.

3. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen untuk menganalisis sentimen publik terhadap dampak penggunaan AI. Metode eksperimen merupakan pendekatan sistematis yang digunakan untuk menganalisis hubungan antara sebab dan akibat [21]. Adapun tahapan dalam pelaksanaan penelitian ini ditunjukkan pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah
Tahap ini berfokus pada perumusan masalah untuk memahami permasalahan secara menyeluruh dan menemukan solusi terbaik.

b. Studi Pustaka
Pada tahap ini, peneliti mengumpulkan sumber relevan untuk mendukung penyelesaian masalah dan memperkuat penelitian.

c. Pengumpulan Data
Pada tahap ini, data dikumpulkan melalui teknik *crawling*, yaitu proses pengambilan data secara otomatis dari *platform* X menggunakan program komputer atau bot (*crawler*) [22]. Sebanyak 3.020 *tweet* berhasil dikumpulkan dengan kata kunci "Penggunaan AI", "Dampak AI", "Penggunaan Kecerdasan Buatan", dan "Dampak Kecerdasan Buatan".

d. *Text Pre-processing*
Text pre-processing merupakan tahap penting dalam pengolahan data yang bertujuan untuk membersihkan data sebelum digunakan dalam penelitian [23]. Pada penelitian ini, tahap *text pre-processing* terdiri dari beberapa proses yaitu:
- *Cleaning* merupakan proses menyederhanakan data teks dengan menghapus informasi yang tidak relevan atau tidak memiliki makna [24].
- *Case folding* merupakan tahap untuk mengonversi semua kata dengan huruf kapital menjadi huruf kecil [23].
- Normalisasi kata merupakan proses mengubah kata tidak baku menjadi bentuk baku sesuai dengan standar KBBI [25].
- *Tokenizing* adalah proses memecah sebuah kalimat menjadi kata-kata tunggal atau token [26].
- *Stopwords removal* digunakan untuk menghilangkan kata-kata yang tidak relevan dan tidak memiliki makna deskriptif, dengan memanfaatkan daftar *stopword* yang tersedia [7].
- *Stemming* adalah proses mengubah kata menjadi bentuk dasarnya dengan menghapus awalan dan akhiran [27].

e. *Labeling*
Pemberian label (*labeling*) adalah proses menandai respons *tweet* dalam dataset menggunakan *Python* dengan bantuan *library Transformers***.** Komentar dalam dataset diklasifikasikan menjadi tiga kategori, yaitu positif, negatif, dan netral.

f. Pembobotan Kata (TF-IDF)
Pada tahap ini, data yang telah diproses dan diberi label diberi bobot menggunakan metode TF-IDF. Teknik ini mengukur pentingnya suatu kata dalam dokumen relatif terhadap seluruh dataset, di mana kata yang sering muncul di satu *tweet* namun jarang di *tweet* lain akan memiliki bobot lebih tinggi, sehingga mendukung efektivitas klasifikasi sentimen [28].

g. *K-fold Cross Validation*
K-Fold Cross Validation adalah teknik evaluasi yang membagi dataset menjadi beberapa lipatan (*folds*) agar model diuji pada berbagai subset data [29]. Penelitian ini menggunakan K = 2–10 untuk mencegah *overfitting* dan memastikan generalisasi model.

h. Teknik *Oversampling* SMOTE
Setelah pembobotan kata, data yang tidak seimbang ditangani menggunakan SMOTE, yang menambah sampel sintetis pada kelas minoritas agar distribusi lebih seimbang [20]. SMOTE diterapkan setelah TF-IDF dan sebelum validasi model, dengan bantuan *library imbalanced-learn* di *Python*.

i. Klasifikasi SVM
Tahap ini merupakan proses utama, yaitu klasifikasi sentimen menggunakan algoritma SVM dengan tiga jenis kernel: *Linear*, RBF, dan *Polynomial* [8].

j. Pengujian Menggunakan *Confussion Matrix*
Evaluasi performa model dilakukan menggunakan *Confusion Matrix* untuk mengukur akurasi, presisi, *recall*, dan *f1-score* dari model SVM.

4. HASIL DAN PEMBAHASAN

4.1. *Crawling*

Hasil proses ini menghasilkan sebanyak 3.020 data *tweet* yang membahas topik terkait dampak penggunaan AI. Berikut ini hasil dari proses *crawling*:

Tabel 2. Hasil *Crawling*

Created_at	***Username***	***Tweet***
2025-03-26 04:09:18+00:00	urnightma rebebb	@burningsolas dampak dari Gibran nyuruh pake AI nih jadi tolol semua
2025-03-26 00:59:32+00:00	stwuerds	AI LU BAWA DAMPAK BAIK APA KE NEGARA? YG ADA MAKIN KEPECAH BELAH

4.2. *Text Pre-preprocessing*

Sebelum dianalisis, data terlebih dahulu dibersihkan dan disederhanakan melalui tahapan pra-pemrosesan teks. Langkah ini dilakukan untuk memastikan data dalam kondisi yang sesuai dan siap digunakan dalam proses analisis sentimen.

a. *Cleaning*

Pada tahap ini, data dibersihkan dengan menghapus elemen yang tidak relevan seperti URL, HTML, emoji, angka, simbol, dan *mention*.

Tabel 3. Hasil *Cleaning*

Sebelum *Cleaning*	**Sesudah *Cleaning***
@burningsolas dampak dari Gibran nyuruh pake AI nih jadi tolol semua	dampak dari Gibran nyuruh pake AI nih jadi tolol semua
AI LU BAWA DAMPAK BAIK APA KE NEGARA? YG ADA MAKIN KEPECAH BELAH	AI LU BAWA DAMPAK BAIK APA KE NEGARA YG ADA MAKIN KEPECAH BELAH

b. *Case Folding*

Pada tahap ini, dilakukan proses *case folding* untuk mengubah seluruh teks dalam *tweet* menjadi huruf kecil.

Tabel 4. Hasil *Case Folding*

Sebelum *Case Folding*	**Sesudah *Case Folding***
dampak dari Gibran nyuruh pake AI nih jadi tolol semua	dampak dari gibran nyuruh pake ai nih jadi tolol semua
AI LU BAWA DAMPAK BAIK APA KE NEGARA YG ADA MAKIN KEPECAH BELAH	ai lu bawa dampak baik apa ke negara yg ada makin kepecah belah

c. Normalisasi

Pada tahap ini, dilakukan proses normalisasi kata untuk mengganti kata-kata tidak baku atau *slang* dengan bentuk baku sesuai dengan KBBI.

Tabel 5. Hasil Normalisasi

Sebelum Normalisasi	**Sesudah Normalisasi**
dampak dari gibran nyuruh pake ai nih jadi tolol semua	dampak dari gibran memerintah pakai ai ini jadi tolol semua
ai lu bawa dampak baik apa ke negara yg ada makin kepecah belah	ai kamu bawa dampak baik apa ke negara yang ada semakin kepecah belah

d. *Tokenizing*

Pada tahap ini, dilakukan *tokenizing*, yaitu proses memecah teks menjadi unit kata atau token.

Tabel 6. Hasil *Tokenizing*

Sebelum *Tokenizing*	**Sesudah *Tokenizing***
dampak dari gibran memerintah pakai ai ini jadi tolol semua	[dampak, dari, gibran, memerintah, pakai, ai, ini, jadi, tolol, semua]
ai kamu bawa dampak baik apa ke negara yang ada semakin kepecah belah	[ai, kamu, bawa, dampak, baik, apa, ke, negara, yang, ada, semakin, kepecah, belah]

e. *Stopword Removal*

Pada tahap ini, dilakukan *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti *"yang", "dan", "di", "ke",* dan sebagainya.

Tabel 7. Hasil *Stopword Removal*

Sebelum *Stopword Removal*	**Sesudah *Stopword Removal***
[dampak, dari, gibran, memerintah, pakai, ai, ini, jadi, tolol, semua]	[dampak, gibran, memerintah, pakai, ai, tolol]
[ai, kamu, bawa, dampak, baik, apa, ke, negara, yang, ada, semakin, kepecah, belah]	[ai, bawa, dampak, negara, kepecah, belah]

f. *Stemming*

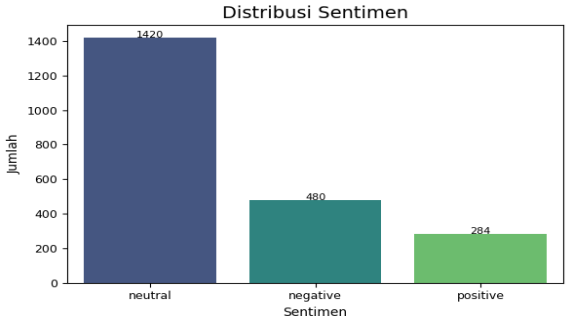
Pada tahap ini, dilakukan *stemming* untuk mengubah kata-kata dalam teks ke bentuk dasarnya.

Tabel 8. Hasil *Stemming*

Sebelum *Stemming*	**Sesudah *Stemming***
[dampak, gibran, memerintah, pakai, ai, tolol]	dampak gibran perintah pakai ai tolol
[ai, bawa, dampak, negara, kepecah, belah]	ai bawa dampak negara pecah belah

4.3. *Labeling*

Tahap ini menghasilkan klasifikasi sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Hasil pelabelan ini akan menjadi acuan dalam proses pelatihan model klasifikasi. Distribusi masing-masing kategori ditunjukkan pada Gambar 2.



Gambar 2. Hasil *Labeling*

4.4. TF-IDF

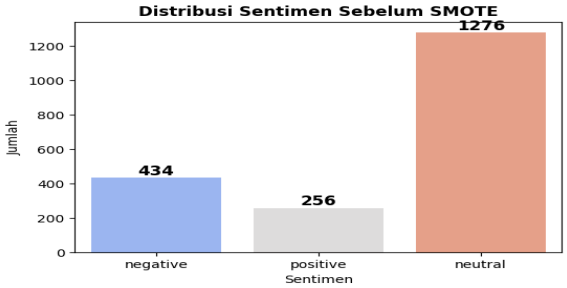
Pada tahap ini, pembobotan kata dilakukan menggunakan metode TF-IDF untuk mengukur pentingnya suatu kata dalam setiap dokumen. Hasil pembobotan ditampilkan pada Tabel 8 yang memuat 5 *term* teratas beserta peringkatnya.

Tabel 9. Hasil TF-IDF

Term	***Rank***
ai	0.064800
guna	0.050310
dampak	0.045494
buat	0.039542
cerdas	0.039440

4.5. *K-fold Cross Validation*

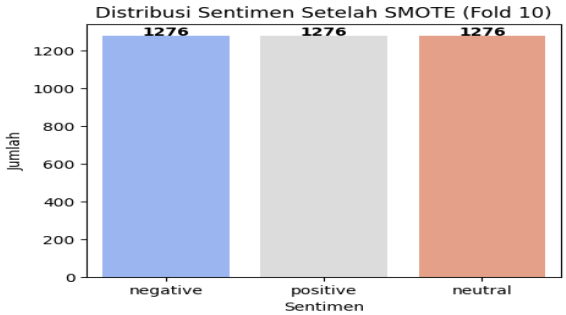
Validasi dilakukan menggunakan *K-Fold Cross Validation* dengan nilai K antara 2 hingga 10. Teknik ini membantu mengevaluasi performa model dan mencari nilai K yang optimal. Sebagai contoh, berikut adalah implementasi *K-Fold Cross Validation* dengan K = 10. Distribusi sentimen sebelum SMOTE ditunjukkan pada Gambar 3.



Gambar 3. Distribusi Sentimen Sebelum SMOTE

4.6. SMOTE

Setelah diterapkan SMOTE, jumlah data pada setiap kelas menjadi seimbang. Hal ini bertujuan agar model tidak condong ke salah satu kelas saat proses pelatihan. Distribusi sentimen setelah penyeimbangan dapat dilihat pada Gambar 4.



Gambar 4. Distribusi Sentimen Setelah SMOTE

4.7. Klasifikasi SVM

Tweet yang telah seimbang kemudian diklasifikasikan menggunakan algoritma SVM. Tiga jenis kernel digunakan dalam proses ini, yaitu *Linear*, RBF, dan *Polynomial*, untuk melihat perbandingan performa masing-masing dalam menentukan kategori sentimen. Proses klasifikasi dilakukan pada data uji setelah model dilatih dengan data latih yang telah diproses.

```
SVC
SVC(C=1, degree=2, kernel='poly', random_state=1)
```

Gambar 5. Klasifikasi SVM

4.8. Pengujian *Confussion Matrix*

Pada tahap ini, dilakukan pengujian terhadap performa algoritma SVM dalam mengklasifikasikan data ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Evaluasi dilakukan menggunakan *confusion matrix* dan dihitung metrik evaluasi berupa akurasi, presisi, *recall*, dan *f1-score*. Pengujian dilakukan dalam dua kondisi, yaitu sebelum dan sesudah penerapan teknik penyeimbangan data SMOTE, untuk melihat dampaknya terhadap performa model.

4.9. Kernel Linear

Tabel 10. Kernel *Linear* Sebelum SMOTE

Fold	**Akurasi**	***Negative***			***Neutral***			***Positive***		
		Presisi	***Recall***	***F1-score***	**Presisi**	***Recall***	***F1-score***	**Presisi**	***Recall***	***F1-score***
2	79%	77%	56%	65%	79%	97%	87%	87%	27%	42%
3	80%	76%	65%	70%	80%	96%	87%	86%	26%	40%
4	81%	78%	64%	70%	81%	97%	88%	86%	27%	41%
5	81%	79%	65%	71%	81%	97%	88%	94%	28%	43%
6	81%	77%	66%	71%	81%	97%	88%	93%	28%	43%
7	80%	76%	62%	68%	80%	97%	88%	100%	27%	42%
8	80%	76%	62%	68%	80%	96%	87%	100%	31%	48%
9	81%	78%	60%	68%	80%	97%	88%	100%	31%	48%
10	81%	80%	58%	67%	80%	97%	88%	100%	36%	53%

Tabel 11. *Linear* Sebelum SMOTE

Fold	Akurasi	*Negative*			*Neutral*			*Positive*		
		Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*
2	80%	74%	68%	71%	83%	92%	87%	63%	37%	47%
3	79%	67%	72%	70%	84%	89%	87%	63%	36%	46%
4	79%	72%	69%	70%	84%	91%	87%	61%	38%	47%
5	80%	74%	69%	71%	84%	92%	88%	64%	40%	49%
6	80%	72%	72%	72%	84%	91%	88%	70%	40%	51%
7	80%	75%	69%	72%	83%	92%	88%	68%	41%	52%
8	81%	74%	67%	70%	83%	92%	87%	76%	46%	57%
9	81%	72%	68%	70%	83%	92%	87%	79%	47%	59%
10	80%	76%	65%	70%	82%	92%	86%	72%	46%	57%

Kernel *Linear* menunjukkan akurasi yang stabil di kisaran 79%–81% sebelum dan sesudah SMOTE. Sebelum SMOTE, kelas *neutral* memiliki performa terbaik (presisi 80%, *recall* 97%, *F1-score* 87%), sementara kelas *negative* berada pada presisi 76–80%, *recall* 56–66%, dan *F1-score* 65–71%. Kelas *positive* masih rendah dengan *F1-score* hanya 40–53%. Setelah SMOTE diterapkan, terjadi peningkatan signifikan pada kelas *positive*, dengan presisi 61–79%, *recall* 36–47%, dan *F1-score* hingga 59%. Kelas *neutral* tetap stabil, sedangkan kelas *negative* sedikit menurun. Secara keseluruhan, SMOTE membantu menyeimbangkan prediksi antar kelas.

4.10. Kernel RBF

Tabel 12. RBF Sebelum SMOTE

Fold	Akurasi	*Negative*			*Neutral*			*Positive*		
		Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*
2	75%	53%	85%	65%	88%	81%	84%	73%	26%	38%
3	75%	53%	88%	66%	88%	80%	84%	86%	32%	46%
4	77%	56%	83%	67%	87%	84%	85%	84%	30%	44%
5	77%	58%	82%	68%	86%	84%	85%	78%	32%	45%
6	77%	58%	84%	68%	86%	84%	85%	78%	30%	43%
7	76%	56%	84%	67%	85%	83%	84%	92%	29%	44%
8	77%	58%	82%	68%	84%	84%	84%	100%	31%	48%
9	77%	59%	83%	69%	85%	85%	85%	100%	31%	48%
10	78%	60%	83%	70%	85%	85%	85%	100%	32%	49%

Tabel 13. RBF Setelah SMOTE

Fold	Akurasi	*Negative*			*Neutral*			*Positive*		
		Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*
2	78%	73%	62%	67%	80%	93%	86%	69%	31%	43%
3	81%	73%	74%	74%	84%	94%	88%	78%	29%	43%
4	82%	76%	73%	75%	83%	94%	88%	77%	32%	46%
5	81%	77%	72%	74%	83%	94%	88%	74%	35%	48%
6	81%	75%	72%	74%	83%	93%	88%	81%	36%	50%
7	82%	76%	74%	75%	83%	94%	88%	82%	34%	48%
8	81%	74%	70%	72%	83%	94%	88%	93%	37%	53%
9	83%	78%	74%	76%	83%	96%	89%	92%	34%	50%
10	82%	79%	71%	75%	82%	95%	88%	90%	32%	47%

Kernel RBF mengalami peningkatan performa setelah penerapan SMOTE. Sebelum SMOTE, akurasi berada di kisaran 75%–78%, dengan kelas *neutral* tampil stabil (presisi 84–88%, *recall* 80–85%, *F1-score* 73–86%). Kelas *negative* memiliki *recall* tinggi namun *F1-score* masih rendah (65–70%), dan kelas *positive* menunjukkan performa paling lemah (*F1-score* 38–49%). Setelah SMOTE, akurasi meningkat menjadi 78%–83%. Performa kelas *negative* membaik dengan *F1-score* hingga 76%, kelas *neutral* semakin konsisten (*F1-score* 77–93%), dan kelas *positive* meningkat dengan *F1-score* mencapai 53%. SMOTE terbukti efektif dalam menyeimbangkan prediksi, terutama untuk kelas *positive*

4.11. Kernel Polynomial

Tabel 14. *Polynomial* Sebelum SMOTE

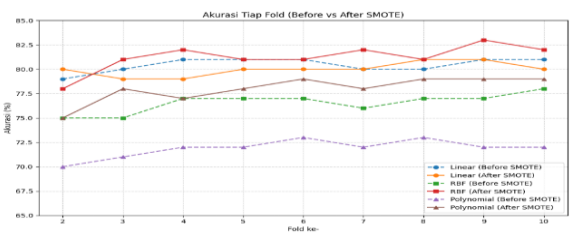
Fold	Akurasi	*Negative*			*Neutral*			*Positive*		
		Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*
2	70%	97%	12%	21%	68%	100%	81%	100%	19%	32%
3	71%	94%	20%	33%	70%	100%	82%	100%	17%	29%
4	72%	96%	22%	35%	70%	100%	82%	100%	17%	29%
5	72%	100%	20%	33%	70%	100%	82%	100%	19%	32%
6	73%	100%	25%	40%	71%	100%	83%	100%	19%	32%
7	72%	100%	24%	38%	70%	100%	83%	100%	17%	29%
8	73%	100%	22%	36%	71%	100%	83%	100%	23%	37%
9	72%	100%	21%	34%	70%	100%	82%	100%	22%	36%
10	72%	100%	19%	32%	70%	100%	83%	100%	25%	40%

Tabel 15. *Polynomial* Setelah SMOTE

Fold	Akurasi	*Negative*			*Neutral*			*Positive*		
		Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*	Presisi	*Recall*	*F1-score*
2	75%	87%	38%	53%	74%	97%	84%	76%	26%	39%
3	78%	81%	52%	63%	77%	97%	86%	89%	26%	41%
4	77%	83%	49%	62%	76%	97%	86%	82%	25%	39%
5	78%	85%	49%	62%	77%	98%	86%	90%	33%	49%
6	79%	87%	51%	65%	77%	98%	86%	88%	32%	47%
7	78%	85%	50%	63%	76%	98%	86%	92%	29%	44%
8	79%	85%	48%	62%	77%	98%	86%	100%	34%	51%
9	79%	84%	51%	64%	77%	98%	87%	100%	34%	51%
10	79%	88%	46%	60%	77%	99%	86%	100%	36%	53%

Kernel *Polynomial* menunjukkan peningkatan performa setelah penerapan SMOTE. Sebelum SMOTE, akurasi berada di kisaran 70%–73%, dengan kelas *neutral* tampil stabil (*F1-score* 81–83%), kelas *negative* memiliki presisi tinggi namun *recall* rendah (*F1-score* 21–40%), dan kelas *positive* mencatat *F1-score* terendah (29–40%). Setelah SMOTE, akurasi meningkat menjadi 75%–79%. Performa kelas *negative* menjadi lebih seimbang (*F1-score* hingga 65%), kelas *neutral* tetap tinggi dan stabil (*F1-score* 84–86%), dan kelas *positive* mengalami peningkatan (*F1-score* hingga 53%). Secara umum, SMOTE berhasil meningkatkan keseimbangan klasifikasi pada seluruh kelas.

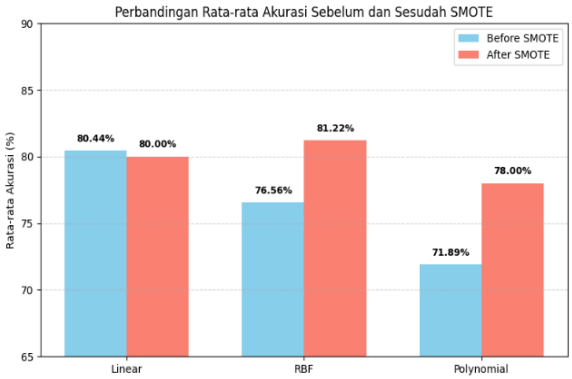
Untuk melihat bagaimana pola akurasi dari masing-masing kernel di setiap nilai K dalam proses *K-Fold Cross Validation*, ditampilkan grafik akurasi dari *fold* ke-2 hingga ke-10. Visualisasi ini memberikan gambaran apakah model mengalami peningkatan performa yang konsisten di tiap *fold* setelah proses penyeimbangan data dilakukan. Grafik ini juga membantu dalam mengevaluasi stabilitas dan kemampuan generalisasi model pada masing-masing kernel SVM.



Gambar 6. Akurasi Per-*fold*

Berdasarkan grafik akurasi per *fold*, terlihat bahwa kernel RBF mengalami peningkatan yang konsisten pada sebagian besar *fold*, dengan nilai akurasi yang cenderung meningkat atau stabil dari K yang lebih kecil hingga K yang lebih besar. Kernel *Polynomial* juga menunjukkan tren peningkatan yang cukup signifikan setelah SMOTE. Sementara kernel *Linear* cenderung stabil di seluruh *fold*, menandakan bahwa model ini cukup *robust*, meskipun tidak mengalami peningkatan sebesar kernel lainnya.

Untuk melihat perbandingan rata-rata akurasi antar kernel secara keseluruhan, ditampilkan Gambar 7 yang menyajikan rata-rata akurasi sebelum dan sesudah SMOTE.



Gambar 7. Rata-rata Akurasi

Gambar tersebut menunjukkan bahwa kernel RBF memiliki rata-rata akurasi tertinggi setelah SMOTE (81,22%), disusul oleh *Linear* (80,00%) dan *Polynomial* (78,00%). Hal ini mengindikasikan bahwa

SVM dengan kernel RBF paling optimal dalam menangani data yang telah diseimbangkan menggunakan SMOTE.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, penerapan SMOTE terbukti efektif dalam meningkatkan performa algoritma SVM untuk klasifikasi sentimen terhadap opini publik mengenai penggunaan AI. Kernel RBF menunjukkan peningkatan akurasi tertinggi dari 76,56% menjadi 81,22%, diikuti kernel *Polynomial* dari 71,89% menjadi 78,00%. Sementara itu, kernel *Linear* tetap stabil di kisaran 80%, menunjukkan kemampuannya dalam menangani data tidak seimbang tanpa teknik tambahan. Selain meningkatkan akurasi, SMOTE juga membantu model mengenali kelas minoritas dengan lebih baik.

DAFTAR PUSTAKA

[1] Nurdin, L. Jama, T. Z. Magnus, R. Priskila, and V. H. Pranatawijaya, "Analisis Sentimen Dampak Artificial Intelligence (AI) Untuk Pendidikan Pada X Menggunakan Naïve Bayes," *J. Inform. Upgris*, vol. 10, no. 1, pp. 15–19, 2024, doi: 10.26877/jiu.v10i1.18867.

[2] N. Bontridder and Y. Poullet, "The role of artificial intelligence in disinformation," *Data Policy*, vol. 3, no. 3, 2021, doi: 10.1017/dap.2021.20.

[3] S. Rahayu, J. J. Purnama, A. Hamid, and N. K. Hikmawati, "Analisis Sentimen AicoGPT (Generative Pre-trained Transformer) Menggunakan TF-IDF," *J. Buana Inform.*, vol. 14, no. 02, pp. 97–106, 2023, doi: 10.24002/jbi.v14i02.7039.

[4] R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, pp. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.

[5] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 11, 2023, doi: 10.31328/jointecs.v8i1.4265.

[6] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.

[7] J. Ono, Y. Anshori, Y. Y. Joefrie, and M. Y. S. Pusadan, "Analisis Sentimen Terhadap Presiden Terpilih Dimedia Sosial Twitter (X) Menggunakan Algoritma Support Vector Machine Jumaita," *Indones. J. Comput. Sci.*, vol. 13, no. 5, pp. 8321–8332, 2024, doi: 10.33022/ijcs.v13i5.4384.

[8] Herwinsyah and A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.

[9] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi)," *JSTIE (Jurnal Sarj. Tek. Inform.*, vol. 9, no. 2, p. 88, 2021, doi: 10.12928/jstie.v9i2.20181.

[10] D. W. Bhatara and R. R. Suryono, "Analisis Sentimen Aplikasi BCA Mobile Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 4, pp. 1907–1917, 2024, doi: 10.37373/tekno.v10i2.419.

[11] E. Puspita and R. R. Suryono, "Perbandingan Berbagai Metode Klasifikasi Teks Untuk Sentimen Kebijakan Makan Gratis di Indonesia Emi," *Indones. J. Comput. Sci.*, vol. 13, no. 5, pp. 8447–8457, 2024, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[12] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.

[13] T. S. Sabrila, Y. Azhar, and C. S. K. Aditya, "Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 10–19, 2022, doi: 10.14421/jiska.2022.7.1.10-19.

[14] R. Agus Dendi, Tri Sugiharto, and Khoiril Irfan, "Perbandingan Teknik Optimasi Grid Search dan Randomized Search dalam Meningkatkan Akurasi Metode Klasifikasi SVM Pada Sentimen Ulasan Perbandingan Teknik Optimasi Grid Search dan Randomized Search dalam Meningkatkan Akurasi Metode Klasifikasi SVM Pada Sen," *SKANIKA*, vol. 8, no. 1, pp. 13–22, 2024, doi: 10.36080/skanika.v8i1.3328.

[15] A. Fauzi and A. H. Yunial, "Analisis Sentimen US Airline Pada Media Sosial Menggunakan Perbandingan Algoritma Data Mining," *J. Edukasi dan Penelit. Inform.*, vol. 10, no. 2, p. 277, 2024, doi: 10.26418/jp.v10i2.76024.

[16] F. I. Wibowo and A. Febriandirza, "Analisis Sentimen Ulasan Pengguna Game Pubg Di Google Play Store Menggunakan Algoritma Naïve Bayes," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, p. 590, 2024, doi: 10.30865/json.v5i3.7264.

[17] W. A. Prabowo and F. Azizah, "Sentiment

Analysis for Detecting Cyberbullying Using TF-IDF and SVM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 11–12, 2020, doi: 10.29207/resti.v4i6.2753.

[18] A. Damayanti, H. L. Safitri, and R. Cahyadi, "Analisis Sentimen Tindakan Pemerintah Indonesia Dalam Penanganan Covid-19 Menggunakan Metode Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 327, 2022, doi: 10.30865/json.v4i2.5341.

[19] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol.*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.

[20] A. Saputra, M. Ezar, A. Rivan, P. S. Informatika, U. Multi, and D. Palembang, "Analisis Sentimen Publik Terhadap Juru Parkir Liar Menggunakan Naïve Bayes dan SMOTE," *J. Algoritm.*, vol. 5, no. 1, pp. 68–77, 2024, doi: 10.35957/algoritme.xxxx.

[21] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[22] S. Nauli, S. S. Berutu, H. Budiati, and F. Maedjaja, "Klasifikasi kalimat perundungan pada twitter menggunakan algoritma support vector machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 107–122, 2025.

[23] Mustasaruddin, E. Budianita, M. Fikry, and F. Yanto, "Klasifikasi Sentiment Review Aplikasi MyPertamina Menggunakan Word Embedding FastText dan SVM (Support Vector Machine)," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 526, 2023, doi: 10.30865/json.v4i3.5695.

[24] F. Nufairi, N. Pratiwi, and F. Herlando, "Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 339–248, 2024, doi: 10.29100/jipi.v9i1.4929.

[25] S. Nurhaliza, Y. Yusra, and M. Fikry, "Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga BBM dengan Metode Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, p. 586, 2023, doi: 10.30865/json.v4i4.6322.

[26] V. Fitriyana, L. Hakim, D. C. R. Novitasari, and A. H. Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 40–49, 2023, doi: 10.35957/jatisi.v9i4.3586.

[27] M. I. Tri Atmojo and E. Sinduningrum, "Analisis Sentimen Tentang Penggunaan Galon Bebas BPA di Indonesia Menggunakan Algoritma Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 394, 2023, doi: 10.30865/json.v5i2.7101.

[28] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.

[29] R. A. Firsttama, A. A. Arifiyanti, and D. S. Y. Kartika, "Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, pp. 282–285, 2024, doi: 10.47233/jteksis.v6i2.1263.