

IMPLEMENTASI SVM DAN SMOTE PADA ANALISIS SENTIMEN MEDIA SOSIAL X TERHADAP PELANTIKAN AGUS HARIMURTI YUDHOYONO

Nurul Fajriyah¹⁾, Nouval Trezandy Lapatta^{*2)}, Deny Wiria Nugraha³⁾, Rahmah Laila⁴⁾

1. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia
2. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia
3. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia
4. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia

Article Info

Kata Kunci: Pelantikan AHY; Menteri ATR; SVM; SMOTE; Analisis Sentimen.

Keywords: *AHY inauguration; Minister of ATR; SVM; SMOTE; Sentiment Analysis.*

Article history:

Received 11 November 2024

Revised 15 Desember 2024

Accepted 14 Januari 2025

Available online 1 Maret 2025

DOI :

<https://doi.org/10.29100/jupi.v10i2.6246>

* Corresponding author.

Nouval Trezandy Lapatta

E-mail address:

nouval@untad.ac.id

ABSTRAK

Pelantikan Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang/Badan Pertanahan Nasional (ATR/BPN) telah memicu berbagai reaksi publik yang terekam dalam media sosial X. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap pelantikan tersebut menggunakan algoritma *Support Vector Machine* (SVM) dan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Data yang digunakan dalam penelitian ini diambil dari komentar masyarakat di media sosial X, yang kemudian diolah untuk membedakan antara sentimen positif, negatif, dan netral. Dalam proses analisis, data awal yang diperoleh cenderung tidak seimbang, dengan jumlah data sentimen negatif yang lebih banyak dibandingkan dengan sentimen positif dan netral. Oleh karena itu, teknik SMOTE diterapkan untuk mengatasi ketidakseimbangan kelas dan meningkatkan performa model. Algoritma SVM kemudian digunakan untuk melakukan klasifikasi sentimen. Hasil penelitian menunjukkan bahwa model SVM yang diimbangi dengan SMOTE memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan sentimen publik dibandingkan dengan model tanpa SMOTE dengan akurasi sebesar 0.93, presisi sebesar 0.93 dan *recall* sebesar 0.93.

ABSTRACT

The inauguration of Agus Harimurti Yudhoyono as Minister of Agrarian and Spatial Planning/National Land Agency (ATR/BPN) has triggered various public reactions recorded on social media X. This research aims to analyze public sentiment towards the inauguration using the Support Vector Machine (SVM) algorithm and the Synthetic Minority Oversampling Tech-nique (SMOTE) technique. The data used in this research was taken from public comments on social media X, which were then processed to differentiate between positive, negative and neutral sentiment. In the analysis process, the initial data obtained tends to be unbalanced, with the amount of negative sentiment data being greater than positive and neutral sentiment. Therefore, the SMOTE technique is applied to overcome the weld imbalance and improve the model performance. The SVM algorithm is then used to perform sentiment classification. The research results show that the SVM model balanced with SMOTE has a high level of accuracy in classifying public sentiment compared to the model without SMOTE with an accuracy of 0,93, precision of 0,93 and recall of 0,93.

I. PENDAHULUAN

Presiden Joko Widodo dan Wakil Presiden Ma'ruf Amin memimpin kabinet yang dikenal sebagai Kabinet Indonesia Maju. Kabinet ini terdiri dari empat menteri koordinator dan tiga puluh menteri bidang, yang diumumkan pada tanggal 23 oktober 2019. Mereka dilantik sesuai dengan Surat Keputusan Presiden Republik Indonesia Nomor 113/P Tahun 2019 tentang Pembentukan Kementerian Negara dan Pengangkatan Negara dan Pengangkatan Menteri Negara Kabinet Indonesia Maju Periode 2019–2024[1]. Selama periode ini, Presiden Joko Widodo telah melakukan 6 kali perombakan menteri. Salah satunya yaitu perombakan menteri Agraria dan Tata Ruang/Kepala Badan Pertanahan Nasional (ATR/BPN).

Presiden Joko Widodo secara resmi melantik Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang/Kepala Badan Pertanahan Nasional (ATR/BPN) untuk sisa masa jabatan periode 2019-2024 di Istana Negara, Jakarta, pada Rabu, 21 Februari 2024. Pelantikan ini dilakukan berdasarkan Keputusan Presiden Republik Indonesia Nomor 34/P Tahun 2024 tentang Pemberhentian dan Pengangkatan Menteri Negara Kabinet Indonesia Maju. Presiden Jokowi juga melantik Hadi Tjahjanto sebagai Menteri Koordinator Bidang Politik, Hukum, dan Keamanan (Menko Polhukam), menggantikan Mahfud MD. Langkah pelantikan Ahy, pemimpin Demokrat yang selama hampir satu dekade berada di luar pemerintahan, dianggap sebagai upaya untuk menghalau serangan lawan politik terkait wacana hak angket di DPR. Hal ini akhirnya mendapat sorotan dan menimbulkan pro dan kontra dari kalangan masyarakat di media sosial.

Media sosial merupakan alat dan *platform* yang memungkinkan orang berinteraksi, berbagi informasi, dan menjalin hubungan secara online[2]. Media sosial mencakup berbagai jenis komunikasi digital yang memungkinkan pengguna untuk membuat dan membagikan konten, serta berpartisipasi dalam percakapan interaktif dan kolaboratif. Twitter atau yang saat ini telah berganti nama menjadi X merupakan salah satu media penghubung yang diminati seluruh masyarakat di dunia. Dibuktikan dengan meningkatnya jumlah pengguna X yang tercatat di seluruh dunia salah satunya Indonesia. X digunakan untuk menyampaikan informasi berupa komentar, kritik, maupun saran terhadap hal-hal yang menjadi pembahasan hangat di kalangan masyarakat saat ini[3].

Analisis sentimen merupakan teknik yang mengevaluasi sentimen, opini, penilaian, dan emosi terhadap suatu produk, layanan, peristiwa, atau atribut lainnya[4]. Metode yang dapat digunakan untuk analisis sentimen adalah algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*. Pemilihan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) dalam analisis sentimen didasarkan pada keunggulan dan keterbatasan masing-masing. SVM dipilih karena kemampuannya dalam generalisasi, efektivitas pada data berdimensi tinggi, dan fleksibilitas dengan berbagai jenis kernel[5]. SVM juga unggul pada data tidak seimbang, terutama dengan teknik seperti SMOTE. Namun, SVM memiliki keterbatasan dalam hal waktu komputasi, pemilihan parameter yang tepat, dan interpretabilitas. *Naïve Bayes* dipilih karena kesederhanaan, kecepatan dalam pelatihan dan prediksi, serta kemampuannya untuk bekerja baik pada dataset kecil dan fitur yang tidak relevan[6]. Meskipun *Naïve Bayes* membuat asumsi independensi fitur yang jarang terjadi dalam data teks, performanya bisa lebih rendah pada data kompleks atau tidak seimbang. Kombinasi kedua algoritma ini memungkinkan analisis yang komprehensif dan fleksibel, dengan SVM memberikan performa yang baik pada data kompleks dan NB menawarkan solusi yang cepat dan efisien.

Algoritma *Support Vector Machine* adalah bagian dari *Supervised Learning* yang melakukan prediksi kelas berdasarkan pola yang dipelajari selama *training*. Untuk melakukan klasifikasi, algoritma ini menggunakan *hyperplane* sebagai pemisah antara kelas-kelas yang berbeda [7]. *Hyperplane* yang optimal adalah yang memiliki *margin* terbesar, karena *margin* yang lebih besar biasanya berhubungan dengan tingkat kesalahan generalisasi yang lebih rendah.[8].

Penelitian berjudul “Analisis Sentimen Twitter Terhadap Menteri Indonesia Dengan Algoritma *Support Vector Machine* dan *Naïve Bayes*” menggunakan dataset sebanyak 3.905 data. Teknik yang dilakukan dalam penelitian yaitu *transform case*, *tokenize*, dan *stopword removal*. Hasil pengujian pada penerapan algoritma SVM didapatkan akurasi sebesar 89,60%, *recall* sebesar 90,91%, dan *precision* sebesar 97,64%. Sedangkan untuk penerapan algoritma *naïve bayes* didapatkan hasil yang lebih rendah dengan akurasi sebesar 85,74%, *Recall* sebesar 85,74%, dan *precision* sebesar 100,00%[1].

Penelitian yang berjudul “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma *Support Vector Machine* (SVM)” melibatkan beberapa tahapan seperti *preprocessing* data, pembobotan kata dengan TF-IDF, penerapan algoritma SVM, dan analisis hasil klasifikasi. Hasil terbaik dari klasifikasi diperoleh dengan menggunakan kernel RBF, dengan parameter $C=10$ dan $\gamma=1$, menghasilkan nilai akurasi sebesar 84,37% dan 80,41% dengan menggunakan 10 *fold cross validation*[9].

Penelitian oleh Herwinsyah dan Arita Witanti yang berjudul “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma *Support Vector Machine* (SVM)” menggunakan perbandingan data 80:20 dengan data uji sebanyak 942 data dan 3766 data latih, hasil prediksi untuk data uji yaitu *f1-score* sebesar 0,93, akurasi sebesar 0,88, presisi sebesar 0,88 dan *recall* sebesar 0,99. Hasilnya menunjukkan bahwa dari 4078 data *tweet*, terdapat 2525 sentimen positif, 771 sentimen negatif, dan 1912 sentimen netral yang memperoleh akurasi sebesar 73,6% [10].

Penelitian berjudul “Analisis Sentimen Terhadap Inflasi Pasca Covid-19 Berdasarkan Twitter dengan Metode Klasifikasi *K-Nearest Neighbor* dan *Support Vector Machine*” yang menggunakan data sebanyak 5989 *tweet* yang kemudian dikategorikan menjadi sentiment positif, negatif, dan netral. Hasil yang didapatkan adalah metode *Support Vector Machine* (SVM) mendapatkan akurasi yang lebih baik yaitu 79% sedangkan metode *K-Nearest*

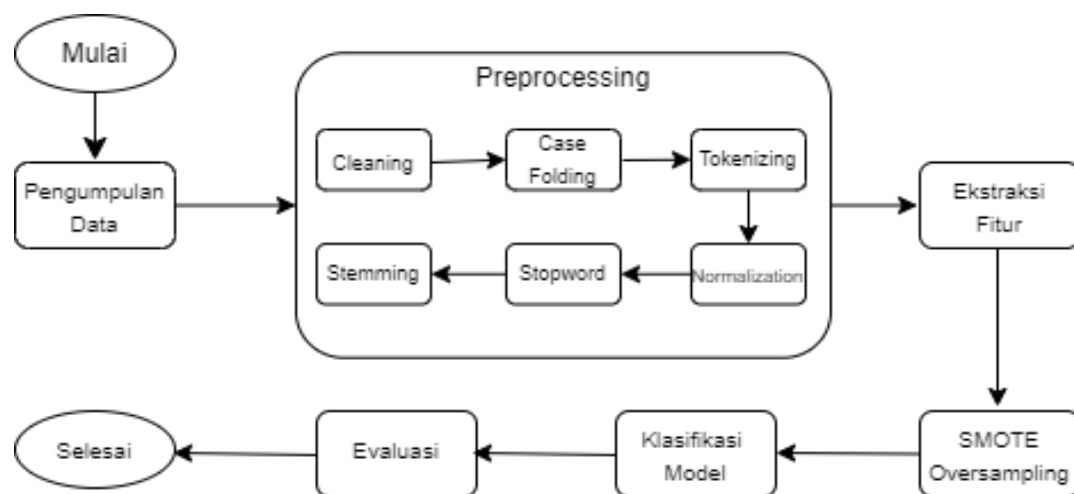
Neighbor (K-NN) mendapat akurasi sebesar 54%[11].

Penelitian berjudul “Analisis Sentimen Terhadap Kinerja Menteri Kesehatan Indonesia Selama Pandemi Covid-19” melibatkan serangkaian proses *preprocessing*, termasuk transformasi huruf, penghapusan URL, penghapusan anotasi, penghapusan *hashtag*, tokenisasi, dan penghapusan *stopword*. Setelah itu, teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data, dan validasi dilakukan menggunakan *k-fold Cross Validation*. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki akurasi tertinggi dalam mengklasifikasikan komentar positif dan negatif terkait analisis sentimen terhadap Kementerian Kesehatan, dengan nilai akurasi 66,45%, sedangkan algoritma NB memiliki akurasi 72,57%. [12].

Penelitian ini akan melengkapi dan memperluas temuan sebelumnya dengan berfokus pada analisis sentimen publik terhadap pelantikan Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang Republik Indonesia menggunakan algoritma SVM. Studi sebelumnya telah menyelidiki SVM dalam berbagai konteks, seperti vaksinasi COVID-19, kinerja menteri, dan sentimen terhadap inflasi pasca-COVID-19. Namun, penelitian ini menerapkan teknik SMOTE untuk menangani data tidak seimbang. Oleh karena itu, penelitian ini tidak hanya meningkatkan pengetahuan tentang kinerja SVM dalam konteks yang berbeda, tetapi juga menambah literatur dengan melakukan analisis menyeluruh terhadap sentimen politik di media sosial. Hasil penelitian ini memungkinkan penelitian lebih lanjut tentang cara terbaik untuk mengoptimalkan teknik analisis sentimen dalam berbagai konteks sosial dan politik lainnya. Selain itu, hasil ini membantu pemahaman lebih baik tentang persepsi publik terhadap perubahan politik, memberikan dasar untuk pengembangan kebijakan, dan memperkuat strategi komunikasi pemerintah di media sosial.

II. METODE PENELITIAN

Penelitian ini bertujuan untuk menangani sentimen publik pasca pelantikan Agus Harimurti Yudhoyono (AHY) sebagai Menteri Agraria dan Tata Ruang/Badan Pertanahan Nasional (ATR/BPN) dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Synthetic Minority Oversampling Technique* (SMOTE). Dalam penelitian ini, dilakukan berbagai tahap, mulai dari pengumpulan data komentar atau *tweet* terkait pelantikan tersebut. Data yang telah terkumpul kemudian diproses melalui tahap *preprocessing* untuk membersihkan dan menyederhanakan data. Selanjutnya, dilakukan *label encoding* untuk mengubah label teks menjadi numerik. Setelah itu, dilakukan ekstraksi fitur untuk mendapatkan representasi data yang relevan. Teknik *oversampling* dengan SMOTE digunakan untuk menangani ketidakseimbangan data. Setelah data siap, dilakukan klasifikasi menggunakan model SVM. Terakhir, dilakukan evaluasi untuk mengukur kinerja model yang telah dibangun. Penelitian ini diharapkan mampu memberikan wawasan tentang perkembangan sentimen publik dan menyediakan analisis yang lebih akurat mengenai pandangan masyarakat terhadap pelantikan AHY sebagai Menteri ATR/BPN.



Gambar. 1. Tahap Penelitian

A. Pengumpulan Data

Data dikumpulkan dari komentar masyarakat di media sosial X terkait dengan pelantikan Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang/Badan Pertanahan Negara. Pengumpulan data dilakukan dengan menggunakan teknik *crawl* yang memanfaatkan *tools tweet-harvest*. *Tools* ini memungkinkan pengumpulan data secara otomatis dengan menggunakan kata kunci spesifik yang relevan dengan topik penelitian.

[illegible]

TABEL II
 HASIL CASE FOLDING DATA

<i>Cleaning</i>	<i>Case folding</i>
pertanyaan buat mas mentri ATR mas AHY apakah masih jadi mentri ATR atau pindah posisi jamentri lain Saya berharap mas men-tri ATR masih di posisi mentri ATR soal nya masih banyak mavia tanah yg harus di berantas	pertanyaan buat mas mentri atr mas ahy apakah masih jadi mentri atr atau pindah posisi jadi mentri lain saya berharap mas mentri atr masih di posisi mentri atr soal nya masih banyak mavia tanah yg harus di berantas

3) Tokenizing

Tokenisasi adalah proses penting dalam pengolahan teks yang bertujuan untuk memecah teks menjadi unit-unit lebih kecil yang disebut token. Token ini dapat berupa kata, frasa, atau simbol lain yang dianggap relevan untuk analisis. Melalui proses tokenisasi, teks dipecah menjadi potongan-potongan kata sehingga setiap elemen teks dapat dianalisis secara individual. Tahap ini sangat penting karena memberikan dasar untuk analisis yang lebih rinci dan akurat terhadap setiap elemen dalam teks, serta membantu dalam identifikasi dan pemrosesan informasi yang terkandung di dalamnya[17].

TABEL III
 HASIL TOKENIZING DATA

<i>Case Folding</i>	<i>Tokenizing</i>
pertanyaan buat mas mentri atr mas ahy apakah masih jadi mentri atr atau pindah posisi jadi mentri lain saya berharap mas mentri atr masih di posisi mentri atr soal nya masih banyak mavia tanah yg harus di berantas	pertanyaan,buat,mas,mentri,atr,mas,ahy,apakah,masih,jadi,mentri,atr,atau,pindah,posisi,jadi,mentri,lain,saya,berharap,mas,mentri,atr,masih,di,posisi,mentri,atr,soal,nya,masih,banyak,mavia,tanah,yg,harus, di, berantas

4) Normalization

Normalization adalah tahap krusial dalam pengolahan teks di mana data yang mengandung kata-kata tidak baku atau bentuk variasi lainnya diubah menjadi bentuk baku yang sesuai dengan standar kamus bahasa Indonesia[18]. Proses ini bertujuan untuk menyelaraskan semua kata dalam data agar mengikuti standar linguistik yang konsisten. Dengan melakukan normalisasi, variasi dalam penulisan kata-kata dapat diubah menjadi bentuk yang seragam, sehingga memungkinkan analisis yang lebih akurat dan dapat diandalkan.

TABEL IV
 HASIL NORMALIZATION DATA

<i>Tokenizing</i>	<i>Normalization</i>
pertanyaan,buat,mas,mentri,atr,mas,ahy,apakah,masih,jadi,mentri,atr,atau,pindah,posisi,jadi,mentri,lain,saya,berharap,mas,mentri,atr,masih,di,posisi,mentri,atr,soal,nya,masih,banyak,mavia,tanah,yg,harus, di,berantas	pertanyaan,buat,mas,mentri,atr,mas,ahy,apakah,masih,jadi,mentri,atr,atau,pindah,posisi,jadi,mentri,lain,saya,berharap,mas,mentri,atr,masih,di,posisi,mentri,atr,soal,nya,masih,banyak,mafia,tanah,yang, harus,di,berantas

5) Stopword

Stopword adalah istilah yang merujuk pada kata-kata yang tidak memiliki makna semantik yang signifikan dan tidak memberikan informasi yang relevan dengan konteks kasus yang sedang diteliti[19]. Kata-kata ini biasanya mencakup istilah-istilah umum seperti preposisi, konjungsi, dan artikel yang sering muncul dalam teks namun tidak memberikan kontribusi penting terhadap pemahaman atau analisis. Dengan menghapus kata-kata yang dianggap tidak memberikan informasi substansial, analisis dapat lebih fokus pada kata-kata yang lebih bermakna dan informatif, yang pada gilirannya meningkatkan kualitas hasil analisis dan relevansi informasi yang diperoleh.

TABEL V
 HASIL STOPWORD DATA

<i>Normalization</i>	<i>Stopword</i>
pertanyaan,buat,mas,mentri,atr,mas,ahy,apakah,masih,jadi,mentri,atr,atau,pindah,posisi,jadi,mentri,lain,saya,berharap,mas,mentri,atr,masih,di,posisi,mentri,atr,soal,nya,masih,banyak,mafia,tanah,yang, harus,di,berantas	pertanyaan,buat,mas,mentri,atr,mas,ahy,jadi,mentri,atr,pindah,posisi,jadi,mentri,berharap,mas,mentri,atr,posisi,mentri,atr,soal,nya,banyak,mafia,tanah, berantas

6) Stemming

Stemming adalah proses penting dalam pengolahan teks yang bertujuan untuk menghapus imbuhan dari sebuah kata sehingga kata tersebut dikembalikan ke bentuk dasarnya[20]. Proses ini melibatkan penghilangan awalan, akhiran, dan infiks dari kata-kata sehingga kata-kata yang memiliki variasi bentuk dapat disederhanakan menjadi bentuk dasarnya yang konsisten. Stemming mempermudah analisis teks dengan memungkinkan fokus pada bentuk kata yang konsisten. Hal ini membantu dalam meningkatkan efisiensi analisis teks dengan mengurangi

kompleksitas variabilitas kata dan memungkinkan identifikasi pola atau informasi yang relevan lebih mudah dilakukan

TABEL VI
HASIL STEMMING DATA

Stopword	Stemming
pertanyaan,buat,mas,menteri,atr,mas,ahy,jadi,menteri,atr,pindah,posisi,jadi,menteri,berharap,mas,menteri,atr,posisi,menteri,atr,soal,nya,banyak,mafia,tanah,berantas	tanya,buat,mas,menteri,atr,mas,ahy,jadi,menteri,atr,pindah,posisi,jadi,menteri,harap,mas,menteri,atr,posisi,menteri,atr,soal,nya,banyak,mafia,tanah,berantas

C. Label Encoder

Label encoder adalah sebuah teknik dalam analisis data dan pembelajaran mesin yang digunakan untuk mengubah nilai-nilai kategori atau label menjadi nilai numerik. Hal ini penting karena sebagian besar algoritma pembelajaran mesin membutuhkan *input* berupa angka, bukan teks atau kata-kata untuk melakukan proses pelatihan. Dalam konteks klasifikasi, label-label yang digunakan untuk mengidentifikasi kategori atau hasil (misalnya negatif, netral, dan positif dalam analisis sentimen) perlu diubah menjadi bentuk numerik agar dapat diproses oleh algoritma. *Label Encoder* melakukan ini dengan cara memberikan label numerik secara berurutan pada setiap nilai unik dalam kolom kategori. Misalnya, jika kita memiliki label kategori A, B, dan C, *Label Encoder* akan mengubahnya menjadi nilai numerik seperti 0, 1, dan 2.

D. Ekstraksi Fitur TF-IDF

Proses pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) mengubah data teks menjadi format numerik dengan memberikan bobot pada setiap kata atau fitur. TF-IDF adalah ukuran statistik yang menilai pentingnya sebuah kata dalam suatu dokumen. TF mengukur seberapa sering kata muncul dalam dokumen tertentu, yang mencerminkan kepentingannya dalam dokumen tersebut. Sementara itu, DF mengukur frekuensi kemunculan kata di seluruh dokumen, yang menunjukkan seberapa umum kata tersebut. IDF adalah kebalikan dari DF, memberikan bobot lebih tinggi pada kata-kata yang jarang muncul di seluruh dokumen. Bobot kata bervariasi tergantung pada frekuensi kemunculannya dalam dokumen tertentu.[21]. Tabel 7 menampilkan 10 kata dengan skor tertinggi pada TF-IDF

TABEL VII
EKSTRAKSI FITUR MENGGUNAKAN TF-IDF

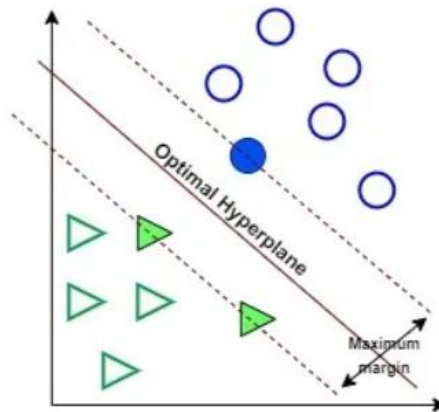
Fitur	Skor TF-IDF
jadi	0.074595
menteri	0.073639
ahy	0.071485
mas	0.034078
enggak	0.033649
lantik	0.029246
pak	0.027037
kalau	0.022204
atr	0.021563
jokowi	0.020918

E. SMOTE Oversampling

SMOTE (*Synthetic Minority Oversampling Technique*) adalah metode yang sangat populer dan efektif untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. Metode ini bekerja dengan menghasilkan sampel sintesis baru dari kelas minoritas untuk meningkatkan proporsi kelas tersebut dalam dataset[22]. SMOTE tidak hanya menyeimbangkan jumlah instance antara kelas mayoritas dan minoritas, tetapi juga meningkatkan kinerja metode klasifikasi secara keseluruhan. Proses ini melibatkan pembuatan *instance* baru yang merepresentasikan variasi dalam kelas minoritas, sehingga memperkaya data latih dan memungkinkan model klasifikasi untuk belajar dari representasi yang lebih seimbang dan mendalam dari seluruh kelas. Tujuan utama dari penerapan SMOTE adalah untuk meningkatkan akurasi dan efektivitas metode klasifikasi dengan mengurangi bias yang sering kali terjadi akibat ketidakseimbangan kelas

F. Klasifikasi Model

Support Vector Machine (SVM) adalah algoritma yang dirancang untuk menemukan *hyperplane* optimal yang dapat memisahkan dua kelas. *Hyperplane* ini berfungsi sebagai pembatas antara kelas-kelas tersebut. SVM berfokus pada pemaksimalan jarak antara data pelatihan dan batas keputusan. Algoritma ini memiliki keunggulan dalam bekerja dengan baik pada berbagai ukuran dataset, baik kecil maupun besar, serta dapat menangani data dengan banyak atribut dengan mudah. Meskipun awalnya hanya mendukung klasifikasi biner, SVM sekarang telah diperluas untuk mendukung klasifikasi *multiclass*[23].



Gambar. 3. Ilustrasi SVM

Tidak semua data dapat dipisahkan dengan mudah oleh sebuah *hyperplane* linier. Untuk mengatasi masalah ini, SVM dapat memanfaatkan berbagai fungsi kernel yang memungkinkan pemetaan data ke ruang fitur yang lebih tinggi, sehingga memungkinkan pemisahan yang lebih kompleks. Terdapat beberapa kernel di antaranya linear, RBF, polinomial dan sigmoid.

1) Linier

Fungsi kernel linier adalah jenis kernel dalam SVM yang digunakan untuk memetakan data ke ruang fitur yang lebih tinggi tanpa melakukan transformasi nonlinier. Fungsi ini tetap memetakan data ke dimensi fitur yang sama tanpa perubahan struktur data. Meskipun fungsinya sederhana, kernel linier sering kali cukup efektif untuk klasifikasi data yang bersifat linier[24].

2) RBF

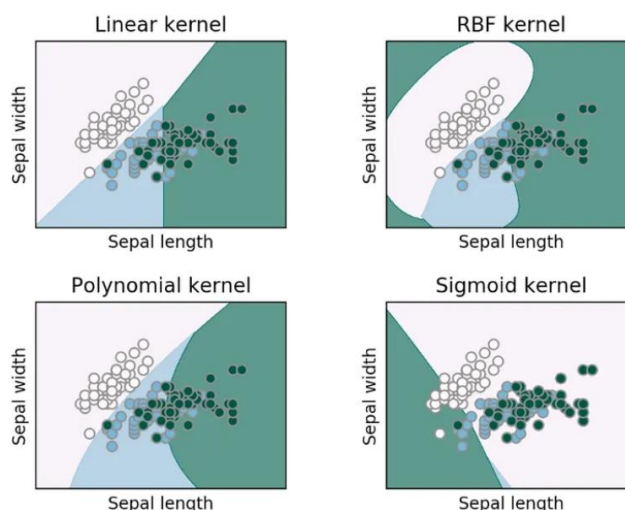
Kernel radial basis, juga dikenal sebagai kernel Gaussian atau RBF, memanfaatkan parameter γ untuk mengatur sebaran kernel. Penyetelan parameter ini sangat memengaruhi kinerja kernel, jadi harus dilakukan dengan hati-hati. Fungsi eksponensial dapat menjadi linier jika parameter ini terlalu tinggi. Ini akan mengurangi kekuatan kernel nonlinier dalam proyeksi dimensi yang lebih besar. Tidak ada mekanisme regularisasi dalam kernel RBF, dan batas keputusan yang dibuat olehnya sangat rentan terhadap gangguan pada data pelatihan. Oleh karena itu, karena nilai parameter γ sangat memengaruhi kinerja SVM, nilai parameter harus dipilih dengan hati-hati.[24].

3) Polynomial

Kernel polinomial dapat digunakan untuk memetakan data ke ruang dimensi yang lebih tinggi, baik dalam konteks hard margin maupun soft margin, ketika menghadapi pola nonlinier. Ini berguna karena gangguan data atau representasi fitur yang kurang optimal sering kali membuat pemisahan data secara linier menjadi sulit[24]

4) Sigmoid

Dalam SVM, fungsi kernel sigmoid memungkinkan transformasi data menjadi bentuk nonlinier dengan menggunakan kurva sigmoid. Ini sangat berguna dalam situasi di mana data tidak dapat dipisahkan secara linier dalam dimensi fitur yang ada. SVM dapat menghasilkan klasifikasi yang efektif untuk data yang kompleks dan nonlinier dengan menyesuaikan parameter gamma dan konstanta c . [24].



Gambar. 4. Macam-macam Kernel

G. Evaluasi

Confusion matrix adalah teknik yang digunakan dalam *data mining* dan *machine learning* untuk mengevaluasi akurasi prediksi model terhadap label data. Teknik ini sering dipakai untuk menilai model klasifikasi, di mana model harus memprediksi label data berdasarkan atribut yang tersedia. Tabel *confusion matrix* mencakup baris untuk label aktual dan kolom untuk label yang diprediksi oleh model. Tabel ini juga menunjukkan frekuensi di mana model memprediksi label dengan benar atau salah[25].

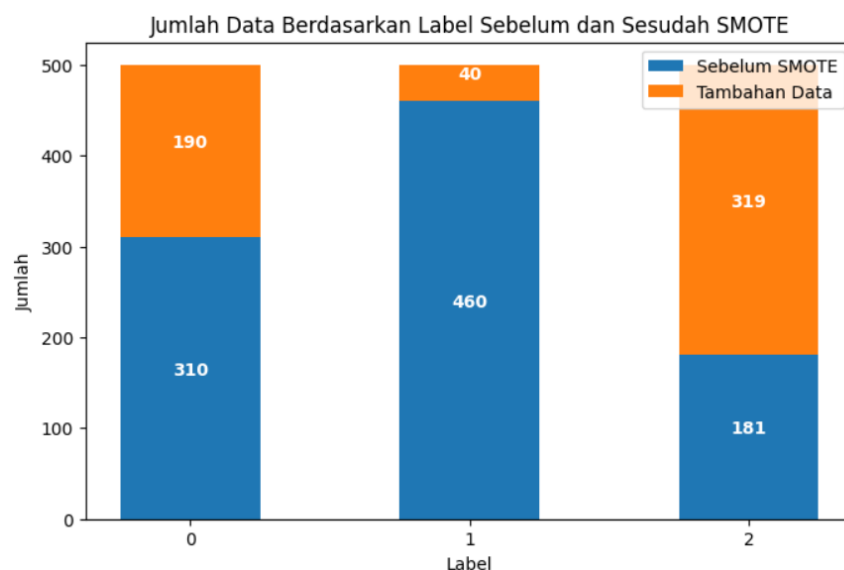
		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar. 5. Tabel *Confussion Matrix*

Berdasarkan gambar 3, *False Positive* (FP) adalah data positif yang salah diprediksi sebagai negatif, sedangkan *False Negative* (FN) adalah data negatif yang juga diprediksi salah oleh model. *True Positive* (TP) adalah data positif yang diprediksi dengan akurat oleh model, sementara *True Negative* (TN) adalah data negatif yang juga diprediksi dengan benar. Evaluasi metrik ini sangat penting dalam menilai kinerja model klasifikasi, karena memberikan wawasan tentang akurasi dan kemampuan model dalam membedakan antara kategori yang berbeda.

III. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data yang diambil dari media sosial X sebanyak 951 data yang kemudian akan diberi label secara manual. Didapatkan data sentimen positif berjumlah 310 data, data sentimen negatif berjumlah 460 data, dan data sentimen netral berjumlah 181 data. Terjadi ketidakseimbangan kelas pada data, yang memungkinkan terjadinya beberapa dampak, diantaranya *overfitting* pada kelas mayoritas dan penurunan nilai akurasi. Untuk mengatasi hal ini, diterapkan teknik SMOTE untuk meningkatkan jumlah sampel dalam kelas minoritas. Setelah teknik SMOTE diterapkan, jumlah data dalam dataset menjadi seimbang untuk setiap kategori sentimen (positif, negatif, dan netral), dengan total 500 data untuk masing-masing kategori. Langkah ini sangat penting untuk mengatasi masalah ketidakseimbangan kelas yang dapat memengaruhi kinerja model dalam analisis sentimen. Dengan keseimbangan ini, model dapat lebih mudah menemukan pola-pola penting dari ketiga kategori sentimen, yang pada akhirnya akan menghasilkan pola-pola penting dalam analisis sentimen. Gambar 6 menampilkan perbandingan data sebelum dan setelah dilakukan SMOTE



Gambar. 6. Perbandingan Data Sebelum dan Setelah SMOTE

Data yang telah melalui tahap-tahap *preprocessing* akan diklasifikasikan menggunakan algoritma SVM. Klasifikasi dilakukan dengan menggunakan empat kernel yaitu linear, RBF, Polinomial, dan sigmoid. Perbedaan dari keempat kernel tersebut adalah kernel linear digunakan untuk pemisahan linear sederhana tanpa transformasi data, kernel RBF mampu menangani data non-linear dengan memetakan ke ruang fungsi berdimensi tinggi, kernel sigmoid cocok untuk pola non-linear dengan fungsi aktivasi sigmoid, sedangkan kernel polinomial memetakan data ke ruang fitur lebih tinggi dengan fungsi polynomial untuk menangani hubungan non-linear. Yang akan dibandingkan dalam pengujian ini adalah nilai *accuracy*, *F1-score*, *precision*, *recall* serta akan dilakukan evaluasi model menggunakan *confussion matrix* dan *AUC Score*. Setelah menggunakan SMOTE, terlihat bahwa kinerja model SVM menjadi lebih baik, akurasi pada tiap kernel meningkat. Hasil ini menunjukkan bahwa SMOTE dapat meningkatkan kinerja model dalam menangani ketidakseimbangan kelas. Tabel 8 menunjukkan *confussion matrix* sebelum penerapan SMOTE, sedangkan Tabel 9 menunjukkan *confussion matrix* setelah penerapan SMOTE.

TABEL VIII
 HASIL *CONFUSION MATRIX* SEBELUM SMOTE

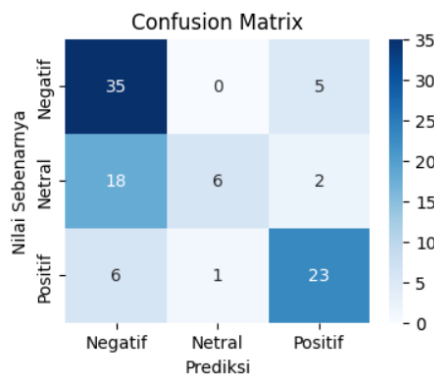
Kernel	Akurasi	Presisi	Recall	F1-Score
RBF	0.67	0.73	0.62	0.67
Linear	0.65	0.70	0.60	0.64
Poly	0.65	0.70	0.60	0.64
Sigmoid	0.61	0.66	0.56	0.60

TABEL IX
 HASIL *CONFUSION MATRIX* SETELAH SMOTE

Kernel	Akurasi	Presisi	Recall	F1-Score
RBF	0.93	0.93	0.93	0.92
Linear	0.89	0.88	0.87	0.87
Poly	0.93	0.92	0.92	0.91
Sigmoid	0.81	0.81	0.81	0.80

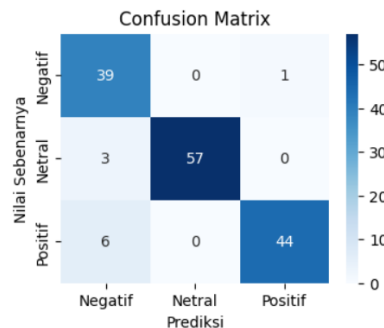
Setelah penerapan SMOTE, kernel RBF dan polinomial menunjukkan akurasi 93% dan nilai *F1-score* yang lebih tinggi. Kernel RBF unggul dalam menangani data non-linear dengan memetakan ke ruang berdimensi lebih tinggi, sehingga efektif untuk data analisis sentimen yang kompleks dan memiliki pola non-linear. SMOTE membantu mengatasi ketidakseimbangan data, memungkinkan RBF memanfaatkan distribusi data yang lebih merata. Kernel polinomial juga menunjukkan hasil yang baik karena mampu menangani interaksi kompleks antar fitur dengan memetakan data ke ruang fitur yang lebih besar. Sebaliknya, kernel linear dan sigmoid kurang efektif dalam menangani data non-linear. Secara keseluruhan, kernel RBF dan polinomial lebih baik untuk analisis sentimen publik terkait pelantikan Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang karena kemampuan mereka dalam menangani hubungan non-linear dan memetakan data ke dimensi yang lebih tinggi, memungkinkan pemisahan kelas yang lebih akurat.

Tabel *confusion matrix* digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan prediksi yang dibuat oleh model dengan nilai yang sebenarnya dalam data pengujian. Tabel ini menampilkan jumlah *instance* untuk setiap kombinasi prediksi dan nilai aktual, memungkinkan identifikasi prediksi yang benar dan salah. Dalam penelitian ini, digunakan matriks 3x3 untuk melakukan evaluasi, setiap sel dalam matriks menunjukkan jumlah *instance* yang sesuai dengan kombinasi kelas prediksi dan kelas aktual. Hasilnya dapat dilihat pada gambar berikut



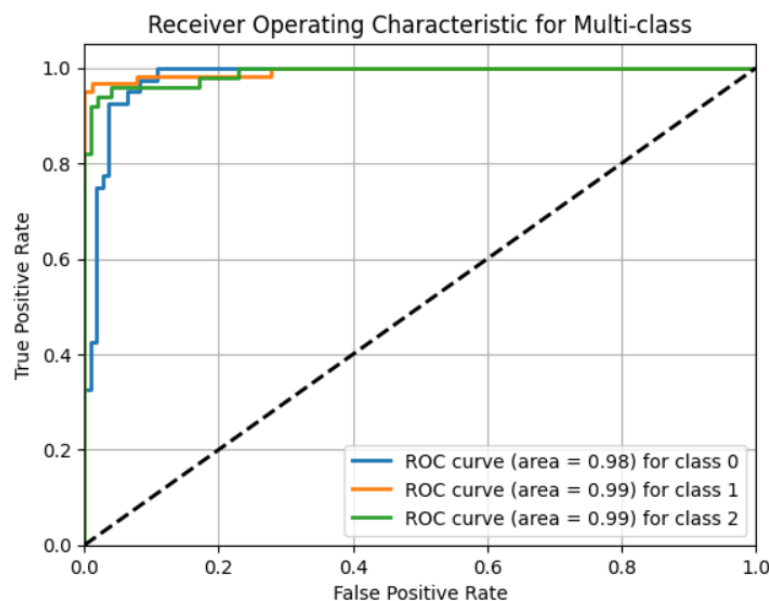
Gambar. 7. Matriks Prediksi model SVM Sebelum SMOTE

Pada gambar di atas ditampilkan *confusion matrix* untuk kernel RBF yang mencapai akurasi tertinggi sebelum penerapan SMOTE pada data. Ada 35 data kelas negatif yang diprediksi dengan benar, tidak ada data kelas negatif yang diprediksi sebagai kelas netral, dan 5 data kelas negatif yang diprediksi sebagai kelas positif. Selain itu, ada 6 data kelas netral yang diprediksi dengan benar, 18 data kelas netral yang diprediksi sebagai kelas negatif, dan 2 data kelas netral yang diprediksi sebagai kelas positif. Terakhir, ada 23 data kelas positif yang diprediksi dengan benar, 6 data kelas positif yang diprediksi sebagai kelas negative dan 1 data kelas negatif diprediksi netral.



Gambar. 8. Matris Prediksi Model SVM Setelah SMOTE

Gambar 8 menunjukkan *confusion matrix* untuk kernel RBF yang mencapai akurasi tertinggi setelah penerapan metode SMOTE pada data. Dalam *confusion matrix* ini, ada 39 data kelas negatif yang diprediksi dengan benar, tidak ada data kelas negatif yang diprediksi sebagai kelas netral, dan 1 data kelas negatif yang diprediksi sebagai kelas positif. Selain itu, ada 57 data kelas netral yang diprediksi dengan benar, 3 data kelas netral yang diprediksi sebagai kelas negatif, dan tidak ada data kelas netral yang diprediksi sebagai kelas positif. Terakhir, ada 44 data kelas positif yang diprediksi dengan benar, 6 data kelas positif yang diprediksi sebagai kelas negatif, dan tidak ada data kelas positif yang diprediksi sebagai kelas netral. Untuk pengujian model dengan menggunakan *AUC Score* dapat dilihat pada kurva ROC pada gambar 9 berikut.



Gambar. 9. Kurva ROC

Berdasarkan kurva ROC yang ditampilkan pada gambar di atas, nilai *AUC Score* untuk model adalah 0,98, yang diperoleh dari rata-rata nilai AUC di setiap kelas. Nilai AUC (*Area Under the Curve*) dan kurva ROC (*Receiver Operating Characteristic*) adalah metrik evaluasi yang penting dalam analisis kinerja model klasifikasi. Kurva ROC menggambarkan hubungan antara TPR (*True Positive Rate*) dan FPR (*False Positive Rate*) pada berbagai *threshold* klasifikasi. Nilai AUC berkisar antara 0 hingga 1, dengan interpretasi sebagai berikut:

- $AUC = 0.5$ menunjukkan bahwa model memiliki kinerja yang sama dengan tebak-tebakan acak
- $0.5 < AUC < 0.7$ mengindikasikan kinerja yang kurang baik.
- $0.7 \leq AUC < 0.8$ menunjukkan kinerja yang cukup baik.
- $0.8 \leq AUC < 0.9$ menunjukkan kinerja yang sangat baik.

- $AUC \geq 0.9$ menunjukkan kinerja yang sangat baik (Excellent Classification).

Nilai AUC sebesar 0.98, seperti yang diperoleh dalam penelitian ini, menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas-kelas yang berbeda. Ini berarti bahwa model mampu mengidentifikasi dengan sangat akurat mana instance yang benar-benar positif dan mana yang benar-benar negatif, dengan sedikit kesalahan klasifikasi.

Penelitian ini menunjukkan bahwa penerapan teknik SMOTE secara signifikan meningkatkan performa model SVM, dengan akurasi mencapai 93% pada kernel RBF dan Polinomial. Ini melampaui hasil penelitian sebelumnya, seperti yang dilaporkan oleh Herwingsyah dan Witanti yang mencatat akurasi 88% pada data tidak seimbang[10]. Sementara itu, Husada dan Paramita melaporkan akurasi 84,37% dengan kernel RBF[9], dan Puspitasari dkk. mencatat akurasi 79% untuk analisis sentimen inflasi pasca COVID-19[11]. Hasil penelitian ini menegaskan efektivitas teknik SMOTE dalam mengatasi ketidakseimbangan kelas dan mengoptimalkan performa SVM, terutama pada kernel RBF dan Polinomial yang menunjukkan hasil terbaik

Untuk penelitian lebih lanjut tentang analisis sentimen, metode tambahan dapat dipikirkan untuk meningkatkan hasil dan mengatasi masalah saat ini. Pertama, mempelajari algoritma dan teknik tambahan seperti *Random Forest*, *Gradient Boosting*, atau model *Neural Networks* seperti LSTM dan BERT dapat menawarkan perbandingan kinerja dan mungkin meningkatkan akurasi. Selain itu, kualitas data dapat diperbaiki dan ketidakseimbangan kelas dapat diatasi dengan lebih efektif dengan penggunaan metode balancing alternatif seperti ADASYN.

IV. KESIMPULAN

Setelah dilakukan penelitian dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Synthetic Minority Oversampling Technique* (SMOTE), dapat ditarik kesimpulan bahwa model awal memiliki akurasi yang relatif rendah sebelum penerapan SMOTE pada data. Hal ini menunjukkan bahwa model mengalami kesulitan dalam mengklasifikasikan data dengan tepat akibat ketidakseimbangan kelas yang signifikan. Namun, setelah SMOTE diterapkan, terjadi peningkatan yang sangat baik dalam kinerja model untuk melakukan analisis sentiment terhadap pelantikan Agus Harimurti Yudhoyono sebagai Menteri Agraria dan Tata Ruang/Kepala Badan Pertanahan Nasional dengan menggunakan kernel *Radial Basis Function* (RBF). Didapatkan akurasi sebesar 0,93, yang menunjukkan bahwa sebagian besar prediksi yang dibuat oleh model adalah benar. Selain itu, presisi dan *recall* yang masing-masing sebesar 0,93 mengindikasikan bahwa model mampu mengidentifikasi *instance* positif dengan baik tanpa terlalu banyak kesalahan. *AUC score* yang mencapai 0,98 mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas-kelas yang berbeda, dan dapat dikategorikan sebagai *Excellent Classification*. Dengan demikian, dapat disimpulkan bahwa penerapan SMOTE berhasil meningkatkan performa model SVM secara signifikan, membuatnya lebih andal dan efektif dalam melakukan analisis sentiment terhadap data yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] S. N. Nugraha, R. Pebrianto, A. Latif, and M. R. Firdaus, "Analisis Sentimen Twitter Terhadap Menteri Indonesia Dengan Algoritma Support Vector Machine Dan Naive Bayes," *E-Link J. Tek. Elektro dan Inform.*, vol. 17, no. 1, p. 1, 2022, doi: 10.30587/e-link.v17i1.3965.
- [2] A. Handayani and I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.
- [3] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Educit - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/educit.v7i1.8779.
- [4] G. Sanjaya and K. M. Lhaksmiana, "Lexicon Based .," vol. 7, no. 3, pp. 9698–9710, 2020.
- [5] I. G. Harsemadi, "Perbandingan Kinerja Algoritma K-NN dan SVM dalam Sistem Klasifikasi Genre Musik Gamelan Bali," *INFORMATICS Educ. Prof. J. Informatics*, vol. 8, no. 1, pp. 1–10, 2023.
- [6] "sitasi revisi.pdf."
- [7] N. Yolanda, I. H. Santi, and D. F. H. Permadi, "Analisis Sentimen Analisis Sentimen Popularitas Aplikasi Moodle dan Edmodo Menggunakan Algoritma Support Vector Machine," *J. Algoritm.*, vol. 3, no. 1, pp. 48–59, 2022, doi: 10.35957/algoritme.v3i1.3313.
- [8] F. F. Abdulloh and I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 13, no. 3, p. 141, 2021, doi: 10.22303/csrid.13.3.2021.141-148.
- [9] 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.
- [10] H. Syah 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.
- [11] R. Puspitasari, Y. Findawati, M. A. Rosid, P. S. Informatika, and U. M. Sidoarjo, "Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Analisis Sentimen Terhadap Inflasi Pasca Covid-19 Berdasarkan Twitter Dengan Metode Klasifikasi K-Nearest Neighbor Dan," vol. 4, no. 4, pp. 1–11, 2023.
- [12] T. Rivanie, R. Pebrianto, T. Hidayat, A. Bayhaqy, W. Gata, and H. B. Novitasari, "Analisis Sentimen Terhadap Kinerja Menteri Kesehatan Indonesia Selama Pandemi Covid-19," *J. Inform.*, vol. 21, no. 1, pp. 1–13, 2021, doi: 10.30873/ji.v21i1.2864.
- [13] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 8, no. April, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.
- [14] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm," *J.*

- Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.
- [15] K. S. Putri, I. R. Setiawan, and A. Pambudi, “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier,” *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.
- [16] S. Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [17] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [18] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, “Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 423–431, 2022, doi: 10.32672/jnkti.v5i3.4382.
- [19] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [20] A. Santosa, I. Purnamasari, and Mayasari Rini, “Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM,” *J. Sains Komput. Inform.*, vol. 6, pp. 81–93, 2022.
- [21] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [22] M. T. Mixue, “Jurnal Indonesia : Manajemen Informatika dan Komunikasi PERBANDINGAN IMPLEMENTASI METODE SMOTE PADA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM ANALISIS SENTIMEN OPINI Jurnal Indonesia : Manajemen Informatika dan Komunikasi,” vol. 4, no. 3, pp. 849–855, 2023.
- [23] F. Abdusyukur, “Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34010/komputa.v12i1.9418.
- [24] S. D. Wahyuni and R. H. Kusumodestoni, “Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting,” vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
- [25] F. Septianingrum, J. H. Jaman, and U. Enri, “Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1431, 2021, doi: 10.30865/mib.v5i4.3260.