

Perbandingan Metode Naïve Bayes dan Support Vector Machine Dalam Analisis Sentimen Terhadap Tokoh Publik

Ardiyansah^{1*}, Parjito²

¹Fakultas Teknik dan Ilmu Komputer, Sistem informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

²Fakultas Teknik dan Ilmu Komputer, Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}Ardiy016@gmail.com, ²djito@teknokrat.ac.id

Email Penulis Korespondensi: ardiy016@email.com

Abstrak—Keberadaan Twitter atau sekarang diganti dengan Nama X telah digunakan secara luas oleh berbagai lapisan masyarakat dalam beberapa tahun terakhir. Media sosial X adalah salah satu media yang merepresentasikan tanggapan masyarakat terhadap tokoh publik. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini masyarakat Indonesia mengenai tokoh publik Luhut Binsar Pandjaitan di media sosial X. Data yang digunakan sebanyak 4008 data terkait topik tersebut yang diperoleh melalui teknik web scraping. Penelitian ini membandingkan performa dua algoritma klasifikasi populer dalam analisis sentimen, yaitu Naïve Bayes dan Support Vector Machine (SVM). Sebelum perbandingan, dilakukan optimasi SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan jumlah data minoritas dan mayoritas agar kedua algoritma dapat belajar dengan lebih baik dari setiap kelas sentimen. Hasil komparasi menunjukkan algoritma Naïve Bayes menghasilkan akurasi 95%, sedangkan SVM menghasilkan akurasi 99%, presisi 99%, recall 100%, dan F1-Score 99%. Evaluasi performa juga dilakukan dengan menganalisis confusion matrix dari masing-masing algoritma. Dapat disimpulkan bahwa SVM memiliki performa terbaik dalam mengklasifikasikan sentimen positif dan negatif dengan lebih akurat dibandingkan Naïve Bayes untuk kasus analisis sentimen terhadap tokoh publik Luhut Binsar Pandjaitan ini. Oleh karena itu, algoritma SVM dapat menjadi pilihan yang lebih baik untuk analisis sentimen terhadap tokoh publik. Penelitian ini berkontribusi dalam pemahaman opini masyarakat tentang kinerja Luhut selama menjabat sebagai Menteri Koordinator Bidang Maritim dan Investasi Indonesia.

Kata Kunci: Analisis Sentimen; Tokoh Publik; Media Sosial; Naïve Bayes; Support Vector Machine

Abstract—The existence of Twitter, or now replaced with Name X, has been widely used by various levels of society in recent years. And social media X is one of the media that represents public responses to public figures. This study aims to perform sentiment analysis on the opinions of the Indonesian public regarding the public figure Luhut Binsar Pandjaitan on social media X. The data used is 4008 data related to the topic which was obtained through web scraping techniques. This study compares the performance of two popular classification algorithms in sentiment analysis, namely Naïve Bayes and Support Vector Machine (SVM). Before the comparison, SMOTE (Synthetic Minority Over-sampling Technique) optimization was carried out to balance the number of minority and majority data so that both algorithms could learn better from each sentiment class. The results of the comparison show that the Naïve Bayes algorithm produces an accuracy of 95%, while the SVM produces an accuracy of 99%, precision 99%, recall 100%, and F1-Score 99%. Performance evaluation was also carried out by analyzing the confusion matrix of each algorithm. It can be concluded that SVM has the best performance in classifying positive and negative sentiments more accurately than Naïve Bayes for the case of sentiment analysis towards the public figure Luhut Binsar Pandjaitan. Therefore, the SVM algorithm can be a better choice for sentiment analysis towards public figures. This research contributes to the understanding of public opinion about Luhut's performance while serving as the Coordinating Minister for Maritime Affairs and Investment of Indonesia.

Keywords: Sentiment Analysis; Public Figure; Social Media; Naive Bayes; Support Vector Machine

1. PENDAHULUAN

Twitter merupakan salah satu media sosial terpopuler yang berperan sebagai wadah komunikasi di masyarakat[1]. Twitter merupakan salah satu media sosial paling populer di Indonesia dan Indonesia merupakan negara dengan pengguna Twitter kelima teraktif di dunia[2]. Pengguna Twitter membuat posting yang disebut tweet, yang berisi hingga 140 karakter dan memungkinkan mereka untuk berinteraksi dan berbagi pendapat atau opini satu sama lain.

Twitter dapat menampilkan berbagai pendapat, termasuk pendapat tentang produk, layanan, tokoh publik, dan masalah politik, tergantung pada konteksnya. Tujuan utama analisis sentimen Twitter adalah untuk memahami opini publik terhadap suatu topik, produk, layanan, atau entitas dengan menganalisis konten tweet yang terkait. Ini bisa berguna bagi perusahaan, organisasi, dan bahkan pemerintah untuk memantau sentimen publik dan merespons dengan tepat. dan dalam penelitian ini mengambil beberapa isu tentang tokoh publik yang terkenal di Indonesia yaitu Luhut Binsar Pandjaitan. Luhut Binsar Pandjaitan adalah seorang pejabat pemerintah dan pengusaha Indonesia yang saat ini menjabat sebagai Menteri Koordinator Bidang Kemaritiman dan Investasi dalam Kabinet Indonesia Maju dan Dalam masa jabatannya luhut binsar pandjaitan menuai berbagai macam komentar dan, mulai dari pujian, kritik dan saran.

Selama masa jabatannya, ada banyak masyarakat yang mendukung atau menentang keputusan yang ia buat. maka dari itu dengan banyaknya pengguna twitter menyampaikan pendapat tentang beliau, dapat dimanfaatkan untuk mencari informasi dari perbedaan pendapat positif dan negatif tersebut. Analisis sentimen dilakukan dengan tujuan untuk menemukan polaritas dalam teks atau opini yang terkandung dalam dokumen yang bersifat positif atau negatif. Media sosial Twitter dipilih karena sangat populer di kalangan pengguna internet saat ini.[3]. Metode penelitian yg bisa digunakan untuk menganalisis pendapat dari tweet pengguna twitter adalah analisis sentimen.

Analisis sentimen adalah teknik analisis data yang digunakan untuk mengidentifikasi dan mengevaluasi sentimen, opini, dan emosi yang terkandung dalam teks yang tidak terstruktur, seperti komentar di media sosial, Twitter adalah salah satu media sosial yang sering digunakan untuk mengungkapkan pendapat dan sentimen[4], proses analisis data

sentimen dapat menggunakan machine learning[5]. Dalam penelitian ini Algoritma Naïve Bayes akan dibandingkan dengan Support Vector Machine untuk memilih algoritma yang memiliki hasil klasifikasi terbaik[6]. Algoritma Naive Bayes adalah teknik klasifikasi yang mencari nilai probabilitas tertinggi dengan menggunakan teorema Bayes. Kelebihan algoritma ini adalah mudah digunakan untuk menangani data yang sangat besar[7].sedangkan Untuk mengklasifikasikan teks dalam situasi multi kelas, algoritma Support Vector Machine (SVM) dapat menggunakan kernel untuk melakukan klasifikasi biner. Algoritma SVM dapat menemukan bidang pemisah terbaik dalam ruang input. Bidang pemisah ini digunakan untuk memisahkan satu kelas dari kelas lain[8].

Berdasarkan penelitian sebelumnya analisis sentimen ini pernah dilakukan pada beberapa penelitian.penelitian sebelumnya dilakukan oleh Joang Ipmawati1, Saifulloh, Kusnawi pada tahun 2024 tentang Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan AlgoritmaSupport Vector Machine dari hasil penelitian ini metode SVM digunakan untuk menganalisis sentimen ulasan dari pengunjung tempat wisata di Yogyakarta yang diperoleh melalui Google Maps. Hasil analisis tersebut menunjukkan tingkat akurasi sebesar 87%, presisi sebesar 90%, recall sebesar 65%, dan nilai f1 sebesar 75% terhadap seluruh dataset dengan pengaturan random state 10. Secara rata-rata, akurasi keseluruhan mencapai 83,8%.[9].Penelitian selanjutnya dilakukan oleh debby alita dan rb ali sodikin tentang Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine hasil penelitian ini metode naive bayes lebih tinggi dari pada support vector machine ,metode naive bayes memiliki nilai accuracy 72.88%, precision 43.49% dan recall 54.95%,sedangkan metode support vector machine memiliki Accuracy 77.88%, Precision 75.00%, dan Recall 7.70%.[10].dan penelitian selanjutnya yang dilakukan oleh Ade Dwi Dayani, Yuhandri, Gunadi Widi Nurcahyo yang berjudul Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine dan hasil penelitian tersebut nilai precision sebesar 0,5156 atau 51,56%, recall sebesar 0,4560 atau 45,60%, f1-score 0,4661 atau 46,61%[11]. Selanjutnya penelitian ini dilakukan oleh bintang adhiyaksa maulana,rifqi arul fauzi,rissa ilmia agustin,siti alia azhaar,tatang rohana yang berjudul komparasi algoritma naive bayes dan svm untuk analisis sentimen twitter korupsi bansos beras masa pandemi hasil pene;itian ini nilai accuracy Support vector machine lebih tinggi dari pada naive bayes support vector machine memiliki accurecy 66,67% sedangkan naive bayes 60,61%.[12].dan penelitian selanjutnya yang dilakukan oleh Ni Luh Putu Merawati , Ahmad Zuli Amrullah , Ismarmiaty yang berjudul Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation hasil dari penelitian Hasil pengujian menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasi sentimen wisatawan dengan baik yang ditunjukkan dengan pengukuran hasil kinerja algoritma melalui 4 parameter yaitu nilai akurasi sebesar 92%, presisi 100%, recall 83,84% dan Spesifisitas 100%. Sedangkan pemodelan topik menggunakan algoritma LDA menghasilkan topik terbaik untuk kelas positif pada 8 topik dengan nilai koherensi 0,613 kemudian untuk kelas negatif topik terbaik pada 12 topik dengan nilai koherensi 0,528.[13].

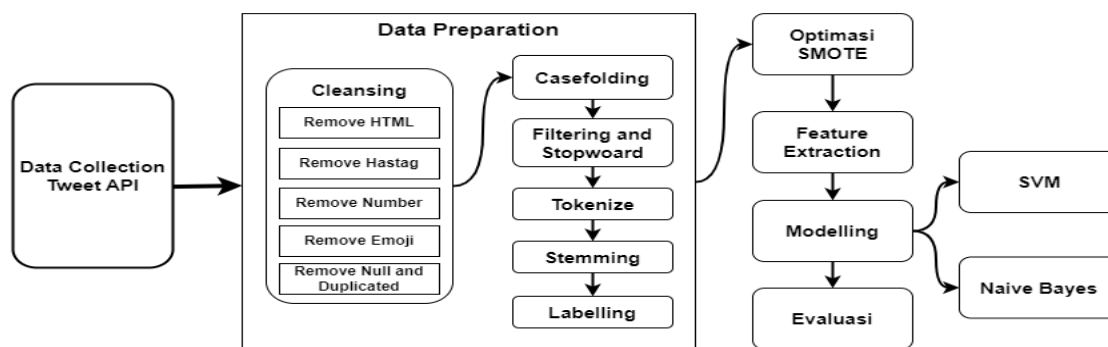
Pada penelitian terdahulu yang dijadikan refrensi dan acuan untuk menulis jurnal ini,dimana peneliti sebelumnya juga sudah melakukan komperasi metode naive bayes dan support vector machine dan ada yang menggunakan metode naive bayes dan sebagai acuan untuk melakukan analisis sentimen. Tujuan dari penelitian ini adalah untuk mengetahui opini masyarakat tentang luhut binsar pandjaitan, apakah masyarakat memiliki sudut pandang yang baik atau sebaliknya tentang luhut binsar pandjaitan. Penelitian ini mengacu pada penelitian sebelumnya yang telah dilakukan. Tujuan dari penelitian ini adalah untuk membandingkan dua metode klasifikasi, yaitu Naive Bayes dan support vector machine, untuk menentukan yang terbaik. Selain itu, penelitian ini juga melakukan optimasi menggunakan Synthetic Minority Over-sampling Technique (SMOTE) untuk meningkatkan kinerja setiap klasifikasi yang dilatih.

Penelitian ini diharapkan memberikan kontribusi dalam pemahaman opini masyarakat tentang tokoh publik luhut binsar pandjaitan. dengan mengetahui opini masyarakat diharapkan menambahkan wawasan bagaimana kinerja luhut binsar pandjaitan selama menjabat sebagai mentri. Dengan mempertimbangkan elemen seperti akurasi dan confusion matrix, analisis ini diharapkan membantu penelitian lanjutan dalam menentukan klasifikasi terbaik antara Naive Bayes dan support vector machine Setiap algoritma juga dioptimasi menggunakan SMOTE untuk mendapatkan kinerja terbaik.

2. METODOLOGI PENELITIAN

2.1 Tahapan penelitian

Ada beberapa tahapan untuk mendapatkan hasil yang terbaik. Tahapan yang dilakukan adalah pengumpulan data atau data collection, Data Preparation , Optimasi SMOTE, Feature Extraction, Modeling, Evaluasi. Alur penelitian analisis sentimen yang digambarkan dimulai dengan pengumpulan data melalui Tweet API, dilanjutkan dengan tahap Data Preparation yang mencakup Cleansing (menghapus HTML, hashtag, angka, emoji, serta data null dan duplikat) dan pemrosesan teks (casefolding, filtering stopwords, tokenisasi, stemming, dan labelling). Setelah itu, dilakukan Optimasi SMOTE untuk menyeimbangkan kelas data, diikuti oleh Feature Extraction untuk mendapatkan fitur-fitur penting dari teks. Proses berlanjut ke tahap Modelling untuk membangun model machine learning, yang kemudian dievaluasi kinerjanya. Penelitian ini menggunakan dua algoritma klasifikasi, yaitu SVM (Support Vector Machine) dan Naive Bayes. Penerapan metode yang menggunakan dua algoritma berfungsi untuk melakukan perbandingan sehingga dapat mengetahui algoritma yang memiliki performa terbaik.dan selanjutnya melakukan evaluasi dari setiap klasifikasi yang di hasilkan. Keseluruhan metodologi analisis sentimen yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

2.2 Data Collection

Data collection dalam analisis sentimen merujuk pada proses mengumpulkan data teks yang akan dianalisis, seperti ulasan produk, komentar di media sosial, atau artikel berita[14]. Dan dalam penelitian ini tahap Mengumpulkan data teks menggunakan teknik web scraping, dan ketika pengambilan data hanya dapat mengambil 200 tweet dalam satu kali crawling, dan dalam pengambilan data membutuhkan auth token akun dari media sosial twitter. Dan kata kunci dalam pemrosesan pengambilan data menggunakan kata kunci “luhut binsar pandjaitan”.

2.3 Data preparation

Data preparation atau persiapan data adalah proses yang penting dalam analisis sentimen sebelum melakukan pemrosesan dan analisis lebih lanjut pada data teks[15]. Karena data yang diperoleh dari media sosial seperti Twitter penuh dengan kata-kata tidak baku dan tidak struktur, tahap praproses ini sangat penting untuk analisis sentimen. dan Tahap Penelitian ini meliputi *cleansing Case folding, filltering dan stopword, tokenize ,stemming ,labelling*[16].

2.4 Optimasi Smote

SMOTE (Synthetic Minority Over-sampling Technique) adalah teknik yang digunakan dalam pembelajaran mesin (machine learning) untuk menangani masalah ketidakseimbangan kelas (class imbalance) pada set data. Masalah ini terjadi ketika jumlah data pada kelas minoritas (misalnya kelas positif) jauh lebih sedikit dibandingkan dengan kelas mayoritas (misalnya kelas negatif).

Teknik SMOTE bekerja dengan menciptakan data sintesis untuk kelas minoritas berdasarkan contoh yang sudah ada. Hal ini dilakukan dengan cara:

- Memilih sampel acak dari kelas minoritas.
- Menghitung k tetangga terdekat (k-nearest neighbors) untuk setiap sampel yang dipilih.
- Membuat data sintesis dengan menginterpolasi fitur-fitur dari sampel yang dipilih dan tetangganya.

Dengan cara ini, SMOTE menghasilkan data baru untuk kelas minoritas sehingga proporsi antara kelas minoritas dan mayoritas menjadi lebih seimbang. Hal ini dapat meningkatkan performa model machine learning yang sensitif terhadap ketidakseimbangan kelas, seperti dalam kasus klasifikasi biner atau multi-kelas.

2.5 Feature Extraction (TF-IDF)

Future Extraction atau Ekstraksi fitur adalah salah satu teknik yang sering digunakan untuk mengurangi kompleksitas data di mana data yang besar diubah menjadi fitur – fitur yang lebih sederhana namun tetap memiliki nilai informasi kinerja model dapat secara signifikan ditingkatkan dengan penambahan fitur ini[17]. TF-IDF terdiri dari dua komponen utama:

- Term Frequency (TF): Ini adalah frekuensi kemunculan sebuah kata (term) dalam sebuah dokumen. Semakin sering sebuah kata muncul dalam dokumen, semakin tinggi nilai TF-nya.
- Inverse Document Frequency (IDF): Ini adalah ukuran seberapa penting sebuah kata dalam seluruh kumpulan dokumen. Kata-kata yang jarang muncul di seluruh dokumen akan memiliki nilai IDF yang lebih tinggi, karena dianggap lebih penting/khas.

$$\text{Rumus TF-IDF: } \text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \quad (1)$$

Dimana:

- t : adalah kata/term
- d : adalah dokumen
- D : adalah kumpulan dokumen

Tujuan utama TF-IDF adalah untuk menghitung bobot setiap kata dalam dokumen, sehingga kata-kata yang sering muncul dalam dokumen tertentu namun jarang di dokumen lain akan memiliki bobot yang lebih tinggi. Ini membantu mengidentifikasi kata-kata yang paling relevan dan penting dalam suatu dokumen.

2.6 Klasifikasi Model

Dalam analisis sentimen, klasifikasi model merujuk pada penggunaan algoritma machine learning untuk mengklasifikasikan teks atau kalimat ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Dua algoritma klasifikasi yang umum digunakan dalam analisis sentimen adalah Support Vector Machine (SVM) dan Naive Bayes.

2.7 Support Vector Machine

Support Vector Machine (SVM) adalah salah satu algoritma supervised learning yang populer digunakan untuk klasifikasi dan regresi. SVM bekerja dengan menemukan hyperplane (garis pemisah) terbaik yang memisahkan kelas-kelas data dalam ruang multidimensi. Terlepas dari fakta bahwa SVM dasar hanya dapat mengklasifikasikan dua kelas, kasus yang biasanya terdiri dari lebih dari dua kelas kemudian dikembangkan untuk menangani masalah non-linear. Memetakan data ke dalam ruang besar dengan trik kernel. Support Vector Machine (SVM) dikenal sebagai teknik prediksi untuk klasifikasi dan regresi[18].

$$f(x) = \sum a_i y_i K(x, x') + b \quad \forall i = 1 \quad (2)$$

Persamaan $f(x) = \sum \alpha_i y_i K(x, x') + b$ merupakan fungsi keputusan SVM untuk klasifikasi biner, di mana α_i adalah bobot untuk setiap vektor support x_i , y_i adalah label kelas dari x_i , $K(x, x')$ adalah fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi agar dapat dipisahkan secara linear, b adalah bias, dan \sum merupakan penjumlahan atas semua vektor support. Fungsi ini digunakan untuk memprediksi kelas data baru x dengan mengevaluasi $f(x)$, jika $f(x) > 0$ maka x diklasifikasikan ke kelas positif, jika tidak ke kelas negatif. Hanya vektor support yang berkontribusi dalam penjumlahan berbobot, dan pemilihan fungsi kernel yang tepat sangat penting untuk memisahkan data tidak linier menjadi linier di ruang fitur yang lebih tinggi.

2.8 Naïve Bayes Classifier

Naive Bayes adalah classifier probabilistik, yang berarti ia memprediksi kemungkinan sebuah sampel data untuk menjadi anggota kategori tertentu. Naive Bayes sendiri dikemukakan oleh Thomas Bayes yang merupakan ilmuwan dari Inggris. Pada proses pengklasifikasiannya Naive Bayes memprediksi dengan memanfaatkan pengalaman di masa lalu sehingga dikenal Teorema Bayes[19]. Naive Bayes mampu mengklasifikasikan opini secara akurat, termasuk opini yang terdiri dari beberapa kalimat positif dan negatif.[20]. Keuntungan menggunakan metode ini adalah bahwa menentukan parameter yang dibutuhkan dalam proses klasifikasi hanya membutuhkan sedikit data pelatihan.[21]. Berikut adalah formula umum untuk menghitung probabilitas posterior menggunakan Naive Bayes:

$$P(c|x) = (P(x|c) * P(c)) / P(x) \quad (3)$$

Formula $P(c|x) = (P(x|c) * P(c)) / P(x)$ digunakan untuk menghitung probabilitas kelas c , diberikan data x . $P(c|x)$ adalah probabilitas posterior, $P(x|c)$ adalah probabilitas likelihood dengan asumsi fitur saling independen, $P(c)$ adalah probabilitas prior, dan $P(x)$ adalah bukti. Kelas dengan probabilitas posterior tertinggi dipilih sebagai prediksi untuk data baru berdasarkan formula Naive Bayes ini.

2.9 Evaluasi

Pada tahap evaluasi dilaksanakan dengan tujuan menilai kemampuan model dalam melakukan prediksi sentimen secara akurat. Kinerja prediksi diukur menggunakan sejumlah metrik, antara lain Confusion Matrix, akurasi, presisi, recall, dan F1-Score. Metrik-metrik tersebut memberikan gambaran sejauh mana model dapat mengklasifikasikan sentimen dengan tepat dan meminimalkan kesalahan prediksi. Confusion matrix merupakan alat evaluasi yang diperoleh dengan membandingkan hasil prediksi dari sebuah model dengan data aktual atau sebenarnya. Keluaran dari confusion matrix berbentuk tabel matriks yang menunjukkan hubungan antara nilai-nilai prediksi dan nilai-nilai sebenarnya. Dengan kata lain, confusion matrix menyajikan visualisasi perbandingan antara apa yang diprediksi oleh model dengan apa yang sebenarnya terjadi dalam data, sehingga kita dapat menganalisis seberapa akurat performa model dalam melakukan klasifikasi atau prediksi[22]. Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model dengan data aktual atau ground truth. Berikut gambaran dari confusion matrix

Tabel 1. Confusion Matrix

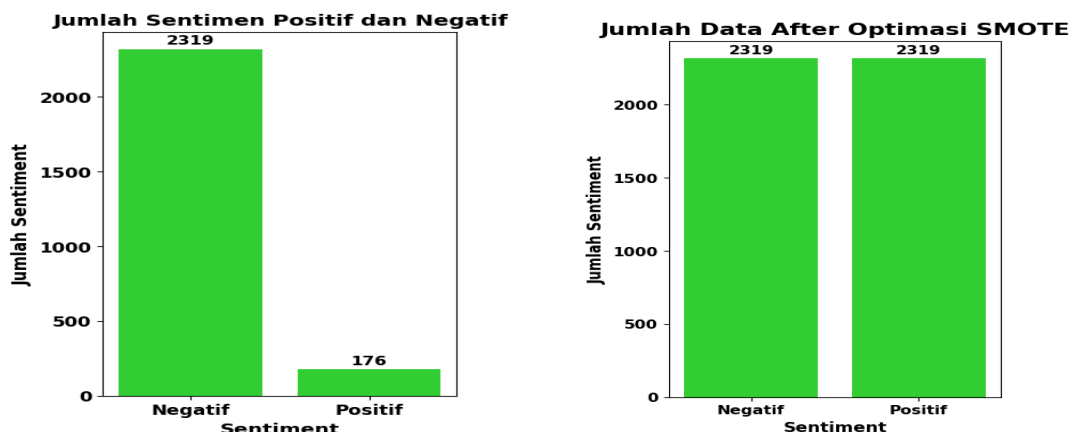
Nilai Aktual	Nilai Prediksi Positif	Nilai Prediksi Negatif
Negative	TP	FN
Positive	FP	TN

Pada tabel 1 di atas True Positif (TP) adalah Jumlah data positif yang diprediksi dengan benar sebagai positif. Sedangkan False Negatif (FN) itu jumlah data positif yang diprediksi salah sebagai negatif. Dan False Positif (FP) Jumlah data negatif yang diprediksi salah sebagai positif dan True Negatif (TN) adalah Jumlah data negatif yang diprediksi dengan benar sebagai negatif. Dengan menganalisis confusion matrix, kita dapat menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk menilai performa model klasifikasi secara lebih komprehensif.

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Akurasi

Setelah melakukan preprocessing data yang semua terkumpul mendapatkan data 4008 data, dan penelitian ini menghasilkan data sentimen positif 176 dan dengan sentimen negatif 2319. karena jumlah data untuk kedua sentimen tidak seimbang, model algoritma naïve bayes dan support vector machine cenderung mendominasi pembelajaran pada kata kata negatif. Hal ini menyebabkan akurasi dan performa untuk kelas negatif lebih tinggi. Dengan adanya tidak keseimbangan data tersebut, eksperimen ini berfokus pada nilai akurasi untuk mengevaluasi model algoritma. untuk mengatasi ketidak seimbangan data, penelitian ini menggunakan optimasi SMOTE agar model algoritma tidak hanya belajar pada satu sentimen saja. Ketika optimasi SMOTE diterapkan, jumlah data minority akan sama dengan data majority. Hasil optimasi SMOTE bisa di lihat pada gambar 2



Gambar 2. Optimasi SMOTE

Pada gambar 2 tersebut diperlihatkan perbandingan antara data sebelum dilakukan optimasi SMOTE dan sesudah dilakukan optimasi SMOTE. dan data yang sudah dilakukan optimasi SMOTE data dengan sentimen positif dan negatif akan sama, Dan memiliki data sebanyak 2319 untuk sentimen positif dan negatif. Dengan optimasi SMOTE ini, model algoritma akan lebih seimbang dalam mempelajari setiap sentimen karena tidak ada lagi data yang jumlahnya sedikit (minority) atau data dengan jumlah sentimen tertinggi (majoritas). Setelah data menjadi seimbang, perlu dilakukan proses pelatihan (training) dan pengujian (testing) agar model algoritma dapat belajar dari data baru yang telah dioptimasi dengan SMOTE.

3.2 Tahap pengujian

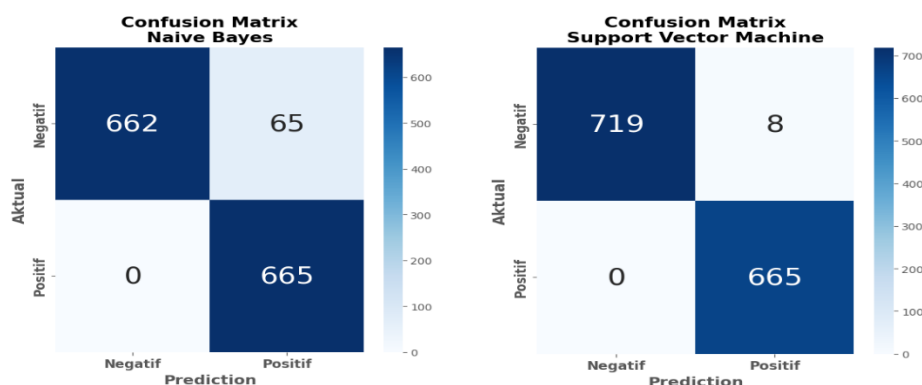
Dalam penelitian ini, setiap model algoritma Naïve Bayes dan Support vector machine menggunakan 70% data untuk pelatihan dan 30% data untuk pengujian. Data yang digunakan dalam penerapan model algoritma ini menggunakan data setelah dilakukan optimasi SMOTE. Dengan demikian, hasil dari setiap algoritma akan dibandingkan untuk mencari model dengan performa terbaik. Informasi lebih detail mengenai hasil dari model algoritma dapat dilihat pada gambar 3

Tabel 2. Modelling Naïve Bayes Dan Support Vector Machine

Modeling		accuracy	precision	recall	F1-Score
Naive bayes	Negatif	95%	100%	91%	95%
	Positif		91%	100%	95%
Logistic regression	Negatif	99%	99%	99%	99%
	Positif		99%	99%	99%

Pada tabel tersebut Berdasarkan eksperimen dengan menggunakan 70% data untuk pelatihan dan 30% data untuk pengujian sesudah dilakukan optimasi SMOTE, ditemukan bahwa baik model algoritma Naïve Bayes maupun support vector machine menghasilkan nilai akurasi yang cukup tinggi. Secara spesifik, model Naïve Bayes mencapai akurasi sebesar 95%, sedangkan model Support vector machine memperoleh akurasi 99%. Namun, dalam penelitian ini, nilai akurasi saja belum cukup untuk dijadikan patokan dalam mengevaluasi kemampuan suatu model dalam membedakan sentimen positif dan negatif secara akurat. Oleh karena itu, diperlukan analisis lebih lanjut terhadap metrik kinerja lain seperti presisi, recall, dan F1-Score untuk mendapatkan penilaian yang lebih komprehensif. Setelah menerapkan optimasi SMOTE pada model, terlihat adanya perbandingan nilai akurasi untuk kedua algoritma yang diuji. Pada algoritma Naïve Bayes, akurasi 95% dan sedangkan untuk support vector machine 99%. Untuk kasus sentimen negatif, presisi naive bayes dan support vector machine sama sama di angka 100% dan untuk recall naive bayes berada di 91% dan support vector machine 99%, dan F1-Score dari naive bayes 95% dan support vector machine 99%.

Dalam rangka menentukan model algoritma terbaik, peneliti melakukan eksperimen dengan membandingkan nilai-nilai pada confusion matrix dari kedua algoritma yang digunakan. Hasil perbandingan confusion matrix untuk masing-masing model algoritma dapat dilihat secara visual pada Gambar 4 yang disajikan.



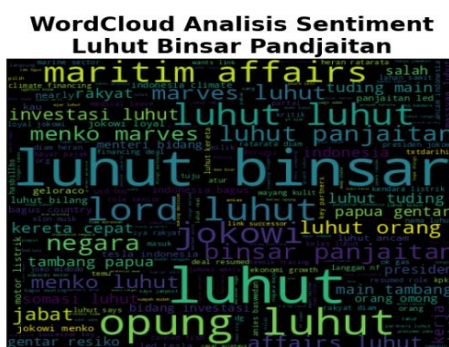
Gambar 3. Confusion matrix

Pada penelitian ini, setelah dilakukan optimasi, setiap algoritma menjadi lebih baik dalam memprediksi makna dari setiap kata. Berdasarkan gambar, algoritma Support Vector Machine (SVM) berhasil mengklasifikasikan 719 data sebagai True Negatif, sedangkan algoritma Naive Bayes menghasilkan 662 data True Negatif. Perbandingan ini menunjukkan bahwa algoritma SVM memiliki akurasi yang lebih tinggi dibandingkan Naive Bayes, dengan akurasi SVM mencapai 99%.

Dalam percobaan kedua, kedua algoritma yang digunakan menunjukkan kinerja yang baik dalam hal akurasi (precision) dan kelengkapan (recall) pada klasifikasi sentimen, baik untuk kelas sentimen positif maupun negatif. Untuk menilai algoritma mana yang terbaik, F1-score digunakan sebagai acuan untuk menentukan performa algoritma. F1-score dihitung berdasarkan precision dan recall. Ketika dibandingkan, algoritma SVM memiliki nilai F1-score sebesar 99%, menunjukkan bahwa algoritma ini memiliki performa yang sangat baik.

3.3 Visualisasi Data

Dalam penelitian ini, eksperimen dilakukan dengan memanfaatkan visualisasi wordcloud sebagai alat untuk menganalisis kumpulan data teks. Visualisasi data wordcloud (awan kata) dalam analisis sentimen sebuah representasi visual yang menampilkan kata-kata yang sering muncul dalam sekumpulan teks atau data teks. Setiap kata ditampilkan dengan ukuran font yang berbeda-beda, di mana kata yang lebih besar menunjukkan kata yang muncul lebih sering dibandingkan dengan kata yang lebih kecil dan hasil dari penelitian ini hasil wordcloud untuk topik luhut binsar pandjaitan bisa dilihat pada gambar 5 berikut ini



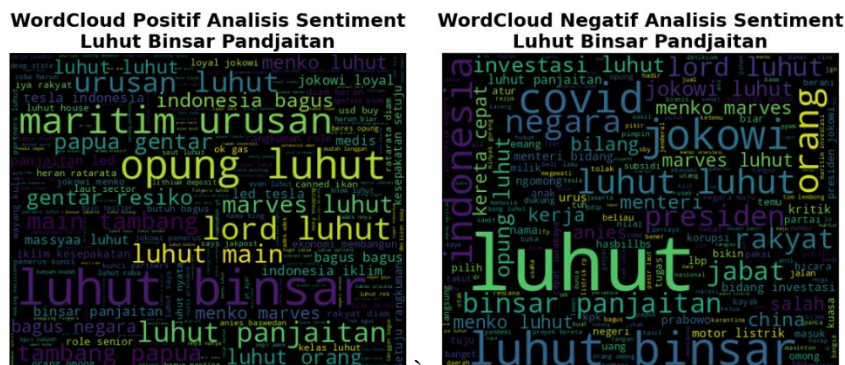
Gambar 4. Wordcloud Luhut binsar pandjaitan

Berdasarkan wordcloud tersebut, dapat dianalisis bahwa fokus utamanya adalah Luhut Binsar Pandjaitan, yang merupakan sosok sentral dalam analisis sentimen ini. Kata "Luhut", "Binsar", dan "Pاندجائتان" muncul dengan ukuran besar dan frekuensi tinggi, mengindikasikan bahwa analisis ini berpusat pada individu tersebut.

Beberapa kata kunci lain yang menonjol adalah "maritim", "menko", "marves", "investasi", dan "jokowi". Ini mengisyaratkan bahwa analisis sentimen ini terkait dengan peran Luhut Binsar Pandjaitan sebagai Menteri Koordinator Bidang Maritim dan Investasi di bawah pemerintahan Presiden Jokowi. Kata-kata seperti "orang", "negara", "rakyat", dan "papua" juga hadir, mungkin mengindikasikan adanya diskusi mengenai dampak kebijakan dan kinerja Luhut terhadap masyarakat Indonesia, termasuk di wilayah Papua. Selain itu, terdapat kata-kata yang berkonotasi netral seperti "bilang", "tuding", "resiko", dan "jabat", yang mungkin merujuk pada pernyataan, tuduhan, atau isu-isu terkait dengan posisi dan tanggung jawab Luhut.

Secara keseluruhan, wordcloud ini mencerminkan analisis sentimen yang komprehensif terhadap sosok Luhut Binsar Pandjaitan, terutama dalam konteks perannya sebagai Menko Maritim dan Investasi, serta dampaknya terhadap

berbagai aspek di Indonesia. Dan dalam penelitian ini menunjukkan masyarakat sering memberikan opini tentang luhut binsar dengan kata-kata tersebut. dan untuk melihat opini tersebut maka dilakukan analisis. dan berikut ini hasil wordcloud dari opini masyarakat termasuk positif atau negatif mengenai luhut binsar pandjaitan dan bisa dilihat pada gambar 6 berikut.

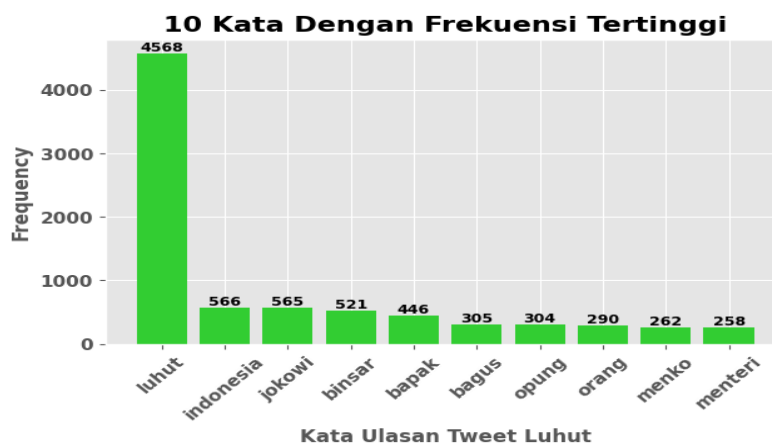


Gambar 6. Wordcloud sentimen positif dan negatif

Berdasarkan wordcloud tersebut, terlihat bahwa kata-kata yang paling menonjol dan berulang adalah "Luhut", "Menko Maryves", "Indonesia", "Urutan", "Maritim", dan "Panjaaitan". Dari kata-kata ini, dapat dianalisis bahwa wordcloud tersebut membahas tentang sosok Luhut Binsar Pandjaitan yang menjabat sebagai Menteri Koordinator Bidang Maritim dan Investasi Indonesia. Kata "Luhut" yang sangat dominan menunjukkan bahwa wordcloud ini berfokus pada pribadi beliau. Kata "Menko Maryves" mengonfirmasi jabatan beliau sebagai Menteri Koordinator. Sementara kata "Maritim", "Indonesia", dan "Urutan" mengindikasikan bahwa topik utama adalah terkait dengan urusan maritim dan investasi di Indonesia yang menjadi tanggung jawab beliau. Secara keseluruhan, wordcloud ini tampaknya dihasilkan dari analisis sentimen positif terhadap kinerja dan peran Luhut Binsar Pandjaitan sebagai Menko Maritim dan Investasi, dengan kata-kata yang disorot menunjukkan apresiasi terhadap upayanya dalam memajukan sektor maritim dan investasi di Indonesia.

Berdasarkan wordcloud sentimen negatif, terlihat bahwa kata-kata negatif yang menonjol adalah "Covid", "negara", "jokowi", "kerja", "rakyat", "jabat", "presiden", dan "china". Kata-kata ini mengindikasikan adanya sentimen negatif yang terkait dengan kinerja Luhut Binsar Pandjaitan dalam penanganan pandemi Covid-19 dan isu-isu lain yang berhubungan dengan pemerintahan. Kehadiran kata "Covid" menunjukkan bahwa wordcloud ini mungkin membahas kritikan terhadap penanganan pandemi yang mungkin dianggap kurang efektif. Sementara kata "negara", "jokowi", "presiden", dan "rakyat" mengisyaratkan bahwa kritikan tersebut juga terkait dengan kebijakan pemerintah dan dampaknya terhadap masyarakat. Kata "kerja" mungkin mengindikasikan ketidakpuasan terhadap kinerja beliau dalam menjalankan tugas sebagai Menko Maritim dan Investasi. Sementara "jabat" dapat merujuk pada keraguan atau kritik terhadap kapasitas beliau dalam mengemban jabatan tersebut. Kata "china" juga muncul dengan cukup menonjol, yang mungkin berkaitan dengan isu-isu terkait hubungan Indonesia dengan China atau investasi China di Indonesia yang menjadi tanggung jawab Luhut Binsar Pandjaitan. Secara keseluruhan, wordcloud ini tampaknya dihasilkan dari analisis sentimen negatif terhadap kinerja dan tindakan Luhut Binsar Pandjaitan, terutama dalam penanganan pandemi Covid-19, kebijakan pemerintah, dan isu-isu terkait investasi dan hubungan dengan China.

Frekuensi kata menunjukkan seberapa sering suatu kata muncul dalam analisis sentimen. Untuk memahami kata-kata yang paling banyak dibahas, kita dapat mengacu pada Gambar 7 yang menyajikan kata-kata dengan frekuensi tertinggi.



Gambar 7. Frekuensi kata tertinggi

Berdasarkan grafik frekuensi 10 kata dengan frekuensi tertinggi, dapat disimpulkan bahwa Kata "luhut" memiliki frekuensi jauh lebih tinggi dibandingkan kata lainnya yaitu 4568 kali. Ini menunjukkan bahwa diskusi atau analisis sentimen tersebut benar-benar terpusat dan sangat berfokus pada sosok Luhut Binsar Pandjaitan sebagai tokoh utama yang dibahas. Keberadaan namanya yang begitu dominan mengindikasikan bahwa berbagai isu atau pembahasan terkait erat dengan dirinya secara pribadi. Kata "indonesia", "jokowi", "binsar" muncul dengan frekuensi tinggi setelahnya. Ini mengaitkan pembahasan mengenai Luhut dengan konteks kebangsaan Indonesia serta pemerintahan di bawah kepemimpinan Presiden Jokowi. Kata "binsar" yang merupakan nama tengahnya juga sering disebutkan, menegaskan bahwa analisis ini secara khusus membahas dirinya sebagai individu. Kata "baik" dan "bagus" yang cukup sering muncul dapat mengindikasikan adanya penilaian atau sentimen positif terhadap kinerja dan tindakan Luhut. Namun, perlu dilihat konteks penggunaannya secara lebih rinci. Kata "opung" dan "orang" kemungkinan merujuk pada panggilan atau sebutan khusus untuk Luhut, mungkin berkaitan dengan latar belakang atau karakternya. Sementara kata "menko" dengan jelas menunjukkan jabatannya sebagai Menteri Koordinator. Meski frekuensinya paling rendah di antara 10 kata tersebut, kata "mereka" yang muncul 258 kali masih cukup signifikan. Hal ini dapat mengindikasikan adanya pembahasan mengenai pihak-pihak lain yang terkait atau terdampak dari kinerja dan kebijakan Luhut, seperti masyarakat, kelompok tertentu, atau pihak lain dalam pemerintahan.

Secara keseluruhan, frekuensi kata-kata ini menegaskan bahwa analisis sentimen tersebut berpusat pada sosok Luhut Binsar Pandjaitan, perannya sebagai Menteri Koordinator di pemerintahan Indonesia saat ini, serta penilaian terhadap kinerjanya dalam kapasitas tersebut, baik dari sisi positif maupun negatif. Konteks nasional Indonesia dan era kepemimpinan Presiden Jokowi juga menjadi latar belakang penting dalam pembahasan ini. Penjelasan yang lebih rinci ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai kesimpulan dari frekuensi kata-kata yang paling sering muncul dalam analisis sentimen terkait Luhut Binsar Pandjaitan

4. KESIMPULAN

Berdasarkan hasil dari penelitian ini telah dilakukan analisis sentimen terhadap opini masyarakat Indonesia mengenai tokoh publik luhut binsar pandjaitan di media sosial twitter. Data dalam penelitian ini menggunakan 4008 data mengenai topik luhut binsar pandjaitan. Dalam penelitian ini dilakukan perbandingan model algoritma Naïve Bayes dan support vector machine. Sebelum dilakukan perbandingan model algoritma naive bayes dan support vector machine, penelitian ini dilakukan optimasi SMOTE agar data minority dapat memiliki jumlah data yang sama dengan data majority. Hasil dari komparasi tersebut mendapatkan nilai algoritma Naïve Bayes menghasilkan akurasi 95% dan algoritma support vector machine menghasilkan akurasi 99%. Dan dapat disimpulkan algoritma dengan performa terbaik yaitu support vector machine dengan nilai akurasi 99%, precision 99%, recall 100%, dan F1-Score 99%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa dalam penelitian ini, algoritma Support Vector Machine (SVM) menunjukkan performa yang lebih baik dibandingkan dengan algoritma Naïve Bayes dalam melakukan analisis sentimen terhadap opini masyarakat Indonesia mengenai tokoh publik Luhut Binsar Pandjaitan di media sosial Twitter. Meskipun kedua algoritma memiliki akurasi yang cukup tinggi, SVM unggul dengan akurasi 99%, precision 99%, recall 100%, dan F1-Score 99%. Hasil ini menunjukkan bahwa SVM mampu mengklasifikasikan sentimen positif dan negatif dengan lebih akurat dibandingkan Naïve Bayes. Oleh karena itu, untuk analisis sentimen terhadap tokoh publik, algoritma SVM dapat menjadi pilihan yang lebih baik dibandingkan Naïve Bayes dalam kasus ini.

UCAPAN TERIMAKASIH

Kepada Perempuan yang pernah kebersamai penulis, terimakasih atas patah hati yang diberikan pada saat proses perkuliahan, karena dengan patah hati penulis dapat menyelesaikan penelitian ini dengan lancar yang sekarang bisa menjadi pengingat untuk penulis sehingga dapat membuktikan bahwa anda akan tetap menjadi alasan penulis untuk terus berproses menjadi pribadi yang lebih baik. walau tak dapat dipungkiri terkadang rasa sedih kerap kali datang dan menjadi kendala tersendiri. Terimakasih telah mengisi cerita di perjalanan penulis menuju cita-cita dan menjadi bagian menyenangkan serta menyakitkan dari proses pendewasaan penulis.

REFERENCES

- [1] S. W. Ritonga, . Y., M. Fikry, and E. P. Cynthia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3535.
- [2] T. T. Widowati and M. Sadikin, "ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE," *Jurnal SIMETRIS*, vol. 11, no. 2, 2020, Accessed: Jun. 19, 2024. [Online]. Available: <http://dx.doi.org/10.24176/simet.v11i2.4568>
- [3] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 556–562, Feb. 2024, doi: 10.57152/malcom.v4i2.1253.
- [4] R. Satria Buana, W. Gata, A. Zevana, P. Widodo, H. Setiawan, and K. Hilyati, "2023 1,2,3,5 Fakultas Teknologi Informasi," vol. 7, no. 2, 2023, doi: 10.35870/jti.

- [5] P. W. Ratiasasadara, S. Sudarno, and T. Tarno, "ANALISIS SENTIMEN PENERAPAN PPKM PADA TWITTER MENGGUNAKAN NAIVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI-SQUARE," *Jurnal Gaussian*, vol. 11, no. 4, pp. 580–590, Feb. 2023, doi: 10.14710/j.gauss.11.4.580-590.
- [6] D. F. Salsabillah, D. E. Ratnawati, and N. Y. Setiawan, "Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma Support Vector Machine dengan Naive bayes (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 107–116, Feb. 2024, doi: 10.25126/jtiik.20241117584.
- [7] T. Rosyida, H. P. Putro, and H. Wahyono, "ANALISIS SENTIMEN TERHADAP PILPRES 2024 BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN NAÏVE BAYES DAN SVM," *ANALISIS SENTIMEN TERHADAP PILPRES 2024 BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN NAÏVE BAYES DAN SVM*, vol. Vol.26, 2023, [Online]. Available: www.apjii.or.id
- [8] S. Riyadiiban and S. Riyadi, "Analisis Sentimen Opini Masyarakat Terhadap Stadion Jakarta Internasional Stadium (Jis) Pada Twitter ... Analisis Sentimen Opini Masyarakat Terhadap Stadion Jakarta Internasional Stadium (Jis) Pada Twitter Dengan Perbandingan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, p. 2024, doi: 10.55338/saintek.v5i3.2962.
- [9] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.
- [10] D. Alita and R. A. Shodiqin, "Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 1–12, Feb. 2023, doi: 10.58602/jaiti.v1i1.20.
- [11] Ade Dwi Dayani, Yuhandri, and G. Widi Nurcahyo, "Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *Jurnal KomtekInfo*, pp. 1–10, Mar. 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [12] R. I. Agustin, "KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN TWITTER KORUPSI BANSOS BERAS MASA PANDEMI," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4020.
- [13] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.
- [14] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [15] J. Homepage, N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, "The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, 2022, Accessed: Jun. 18, 2024. [Online]. Available: <https://doi.org/10.57152/malcom.v2i1.195>
- [16] M. Y. Siregar, A. Davy Wiranata, and R. A. Saputra, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Streaming Vidio Menggunakan Metode Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer Media Online*, vol. 4, no. 5, pp. 2419–2429, 2024, doi: 10.30865/klik.v4i5.1787.
- [17] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, 2024, doi: 10.30865/mib.v8i2.7517.
- [18] D. Oktavia and Y. R. Ramadahan, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Media Online*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [19] I. A. Hidayah, R. Kusumawati, Z. Abidin, and M. Imamuddin, "Analysis of Public Sentiment Towards The Tiktok Application Using The Naive Bayes Algorithm and Support Vector Machine," *Architecture and High Performance Computing*, vol. 6, no. 2, 2024, doi: 10.47709/cnahpc.v6i2.3990.
- [20] S. Hidayatulloh, W. Putra, and D. Febriawan, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store," *Media Online*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1600.
- [21] S. Berliani and S. Lestari, "Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 951–960, Apr. 2024, doi: 10.55338/saintek.v5i3.2746.
- [22] B. Ramadhani, R. R. Suryono, and K. Kunci, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, 2024, doi: 10.30865/mib.v8i2.7458.