

Analisis Sentimen Publik Terhadap Kenaikan Pajak PPN di Indonesia Tahun 2024 Menggunakan Algoritma *Machine Learning*

Mhd Arief Hasan*¹, Novia Putri Bimby²

^{1,2}Program Studi Teknik Informatika, fakultas Ilmu Komputer, Universitas Lancang Kuning
m.arif@unilak.ac.id, 2255201031@filkom.unilak.ac.id

Abstract

The Indonesian government's policy to increase the Value-Added Tax (VAT) rate to 12% has sparked diverse public reactions, particularly on social media platforms like Twitter. The core issue addressed in this study is the need to automatically and accurately understand public sentiment toward this policy, which poses a challenge in processing unstructured textual data. This research aims to evaluate and compare the performance of three popular machine learning classification algorithms—Logistic Regression, Naive Bayes, and Support Vector Machine (SVM)—in analyzing public sentiment extracted from tweets related to the VAT increase. The analysis involves text preprocessing, feature transformation using TF-IDF, and performance evaluation using accuracy and F1-score metrics. The results show that the SVM algorithm achieved the highest accuracy at 82.24%, followed by Logistic Regression at 81.15% and Naive Bayes at 74.63%. These findings indicate that SVM is the most effective algorithm for distinguishing between positive, negative, and neutral sentiments in social media texts, making it a recommended choice for similar sentiment analysis tasks in the future.

Keywords: sentiment analysis, tweet, logistic regression, naive bayes, support vector machine.

Abstrak

Kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12% di Indonesia telah memicu beragam reaksi masyarakat, terutama di media sosial seperti Twitter. Permasalahan dalam penelitian ini terletak pada pentingnya memahami persepsi publik terhadap kebijakan tersebut secara otomatis dan akurat, yang menjadi tantangan tersendiri dalam pengolahan data teks. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja tiga algoritma klasifikasi machine learning populer, yaitu Logistic Regression, Naive Bayes, dan Support Vector Machine (SVM), dalam menganalisis sentimen publik dari data tweet terkait kebijakan kenaikan PPN. Proses analisis melibatkan tahapan preprocessing, transformasi teks menggunakan TF-IDF, serta evaluasi menggunakan metrik akurasi dan f1-score. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi sebesar 82,24%, diikuti oleh Logistic Regression (81,15%) dan Naive Bayes (74,63%). Temuan ini menunjukkan bahwa SVM lebih efektif dalam membedakan sentimen positif, negatif, dan netral pada teks media sosial, serta dapat dijadikan algoritma yang direkomendasikan untuk tugas analisis sentimen serupa di masa mendatang.

Kata kunci: analisis sentimen, tweet, logistic regression, naive bayes, support vector machine.

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Salah satu metode dalam pengolahan bahasa alami (Pengolahan Bahasa Alami) adalah analisis sentimen [1]. Ini digunakan untuk menemukan dan mengkategorikan pendapat atau perasaan yang terkandung dalam teks [2]. Platform media sosial seperti Twitter, yang sering digunakan untuk menyuarakan pendapat tentang masalah terkini, seperti kebijakan pemerintah, adalah salah satu aplikasi analisis sentimen yang semakin populer [3]. Salah satu topik yang banyak mendapatkan perhatian adalah kebijakan kenaikan PPN 12% yang baru-baru ini diumumkan [4]. Masyarakat mengungkapkan reaksi beragam terhadap kebijakan ini, mulai dari sentimen positif hingga negatif, yang menciptakan tantangan dalam analisis sentimen berbasis teks [5]. Untuk

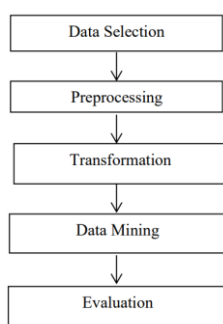
analisis sentimen, ada banyak algoritma klasifikasi, termasuk Logistic Regression, Naive Bayes, dan Support Vector Machine (SVM). Masing-masing algoritma memiliki fitur dan keunggulan tertentu dalam memproses dan mengklasifikasikan data teks. Logistic Regression sering dipilih karena kesederhanaannya dan kemampuan untuk menangani masalah klasifikasi biner, sementara Naive Bayes dikenal karena efisiensi [6].

Tujuan penelitian ini adalah untuk mengevaluasi kinerja ketiga algoritma klasifikasi tersebut dalam menganalisis sentimen tweet terkait kebijakan kenaikan PPN sebesar 12% [7]. Dengan membandingkan akurasi dan kemampuan masing-masing algoritma dalam mengklasifikasikan sentimen

positif, negatif, dan netral, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai algoritma mana yang paling efektif dalam menangani tugas ini [8].

2. Metode Penelitian

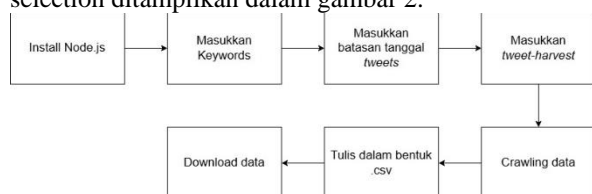
Metode yang digunakan dalam penelitian ini adalah klasifikasi machine learning. Ia dikenal karena kemampuannya untuk mengklasifikasikan data teks secara efektif berdasarkan pola dan fitur yang dikenali dalam data[8]. Pendekatan ini memungkinkan analisis sentimen dengan tingkat akurasi yang tinggi melalui penerapan algoritma klasifikasi yang telah teruji dalam berbagai penelitian sebelumnya[9]. Algoritma yang digunakan dalam penelitian ini meliputi Logistic Regression, Support Vector Machine (SVM), dan Naive Bayes, yang masing-masing memiliki keunggulan dalam menangani data teks berdimensi tinggi serta mengatasi tantangan dalam analisis sentimen[10].



Gambar 1. Metode Penelitian

2.1. Data Selection

Dalam penelitian ini, peneliti memilih untuk mengumpulkan data tweet yang menggunakan hashtag terkait PPN (Pajak Pertambahan Nilai) [11]. Pengambilan sampel ini bertujuan untuk mendapatkan wawasan yang mendalam mengenai pandangan masyarakat tentang kebijakan kenaikan PPN 12%. Sampel data yang digunakan adalah 1379 tweet yang diambil dari rentang waktu tertentu, yaitu mulai dari 29 Oktober 2021 hingga 31 Desember 2024. Karena ketentuan ini, crawling data memiliki keterbatasan. Anda harus membatasi jumlah data yang diambil dari tweets per hari. Alat crawling tweet-harvest digunakan untuk mengumpulkan data; alat ini memungkinkan peneliti untuk mengakses dan menyimpan tweet yang relevan dengan cepat dan efektif. Untuk memastikan bahwa tujuan dari penelitian ini tercapai, langkah-langkahnya disusun secara sistematis. Proses data selection ditampilkan dalam gambar 2.



Gambar 2. Alur Crawling Data Twitter

2.2. Preprocessing Data

Langkah preprocessing dilakukan setelah proses pemilihan data selesai. Pada langkah ini, data dibersihkan untuk menghilangkan gangguan, atau suara, dan disiapkan untuk analisis. Ini dilakukan agar hasil analisis menjadi lebih akurat dan ideal. Tahap pertama adalah pembersihan data. Hal ini termasuk menghapus karakter khusus, duplikat, dan tanda baca seperti tanda seru, titik, koma, dan tanda tanya. Tujuan dari proses ini adalah untuk memastikan bahwa data yang digunakan lebih bersih dan konsisten. Selanjutnya, kapitalisasi dilakukan. Ini adalah proses mengubah semua karakter dalam teks menjadi huruf kecil untuk memastikan konsistensi format. Langkah selanjutnya adalah menghilangkan kata-kata berhenti. Kata-kata yang dianggap tidak penting (misalnya “dalam”, “dan”, “apa”) dan kata-kata umum lainnya yang sering muncul tetapi tidak relevan dengan analisis. Tokenisasi, proses membagi teks menjadi bagian-bagian yang lebih kecil seperti kata atau frasa, digunakan untuk memproses data setelah stop word dihilangkan. Proses ini membuat lebih mudah untuk menganalisis setiap kata yang ada dalam data. Selain itu, stemming juga dilakukan. Stemming melibatkan penghapusan imbuhan dan akhiran dari kata-kata dan mengubahnya menjadi bentuk dasarnya. Lemmatisasi adalah proses mengubah suatu kata menjadi bentuk dasarnya (lemma) dan sebanding dengan stemming kata. Namun, lemmatisasi lebih kompleks karena mempertimbangkan konteks dan makna kata saat mengubahnya menjadi bentuk dasar. Lemmatisasi memastikan kata-kata memiliki struktur tata bahasa yang tepat, berbeda dengan stemming, yang hanya menghapus awalan atau akhiran. Proses terakhir di bagian ini adalah pelabelan data menggunakan alat bantu seperti TextBlob, yang dibuat untuk mendukung analisis bahasa alami (NLP). Data ulasan dapat dikategorikan ke dalam kategori sentimen positif, negatif, atau netral melalui pelabelan ini. Hasil analisis sentimen ini membuat pengguna lebih mudah memahami isi ulasan tanpa membacanya secara menyeluruh.

Data akan berada dalam kondisi yang lebih siap untuk dianalisis setelah seluruh proses preprocessing ini diselesaikan, yang akan menghasilkan informasi yang lebih relevan dan akurat [12].

2.3. Transformation

Tujuan dari fase transformasi, proses vektorisasi, adalah memberi bobot pada kata-kata dalam tweet sehingga polanya dapat diidentifikasi saat memasuki fase pembangunan data.

2.4. Data Mining

a) Logistic Regression

Logistic Regression adalah metode pembelajaran untuk memodelkan hubungan antara variabel independen dan dependen yang bersifat biner atau kategorikal.

Prosesnya menggunakan Maximum Likelihood Estimation (MLE) untuk mengestimasi parameter dan menerapkan fungsi sigmoid untuk memetakan probabilitas antara 0 dan 1 [13]. Data dibagi menjadi dua bagian: dua puluh persen data uji untuk melatih dan mengevaluasi model dan delapan puluh persen data latih. Logistic Regression unggul dalam menangani berbagai jenis data dan membuat interpretasi lebih mudah dipahami. Untuk klasifikasi multi-kelas, metode ini dapat diperluas melalui pendekatan One-vs-Rest (OvR) atau Softmax Regression [14].

b) Naive Bayes

Naive Bayes adalah metode berbasis probabilitas yang menggunakan Teorema Bayes dan mengasumsikan fitur saling independen (naivety) [15]. Model ini melatih data dengan menghitung probabilitas prior dan likelihood setiap kelas, lalu memprediksi kelas dengan probabilitas tertinggi [16]. Data dibagi 80% untuk latih dan 20% untuk uji. Algoritma ini cepat, efisien untuk data berdimensi tinggi, dan efektif dalam klasifikasi teks seperti analisis sentimen dan deteksi spam, meskipun asumsi independensi tidak selalu terpenuhi [17].

c) Support Vector Machine (SVM)

SVM adalah metode pembelajaran yang memanfaatkan hyperplane untuk memisahkan kelas data secara optimal [18]. Data dibagi menjadi 80% latih dan 20% uji untuk melatih dan menguji model. SVM bekerja dengan memaksimalkan margin antara kelas melalui hyperplane terbaik, baik secara linear maupun non-linear, menggunakan teknik kernel untuk menangani data non-linear. Algoritma ini efektif pada data berdimensi tinggi, efisien memori, dan memiliki performa kuat dalam klasifikasi [19].

3. Hasil dan Pembahasan

3.1. Data Selection

Data Selection dilakukan melalui crawling data dari tweet yang berisi pendapat dari aplikasi Twitter menggunakan keywords yang relevan dengan subjek penelitian. Kata kunci yang digunakan mencakup hashtag seperti #PPN12Persen, #PPN2024, #PPNaiK, #KebijakanPPN, #PPNLangsung, #PPNaiK2024, #KenaikanPPN, #PajakPPN, #PajakIndonesia, #PPNaiK12, dan #PajakNaik2024. Alat tweet-harvest yang dibangun menggunakan Python digunakan untuk Crawling data ini. Proses ini mengumpulkan 1379 tweet dari 29 Oktober 2021 hingga 31 Desember 2024. Informasi seperti created_at, full_text, akun, dan tweet_url termasuk dalam data yang dihasilkan. Data ini kemudian digunakan sebagai dataset utama dalam penelitian untuk analisis sentimen mengenai kebijakan kenaikan PPN 12%.

3.2. Preprocessing Data

a) Cleaning Data

Proses ini membersihkan teks pada kolom "full_text" dari URL, mention (@username), tag HTML, angka, dan simbol yang tidak relevan. Tujuannya untuk menyisakan informasi yang bermakna agar analisis sentimen lebih akurat.

b) Case Folding

Pada tahap ini, semua karakter dalam teks diubah menjadi huruf kecil. (huruf kecil) Menyeimbangkan format dan mengurangi kompleksitas analisis data.

c) Tokenization

Proses ini membagi teks yang telah dibersihkan menjadi kata-kata (token) berdasarkan spasi, sehingga setiap kata dapat dianalisis satu per satu.

d) Stopword Removal

Hilangkan kata-kata yang umum dan tidak penting seperti "dan", "dalam", dan "ke" untuk memfokuskan analisis Anda pada kata-kata yang lebih penting.

e) Lemmatization

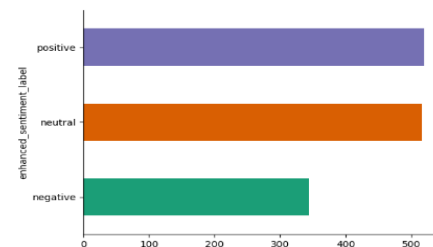
Ubah kata menjadi bentuk dasarnya untuk mengurangi variasi kata tanpa mengubah makna, menjadikan analisis lebih efisien dan akurat.

	full_text	original_text	after_remove1	after_lowerase	after_no_punctuation	tokens	after_stopwords	lemmatized_tokens
0	Direktor Eksekutif Parameter Pajak Indonesia...	Direktor Eksekutif Parameter Pajak Indonesia...	Direktor Eksekutif Parameter Pajak Indonesia...	direktor eksekutif parameter pajak indonesia...	direktor eksekutif parameter pajak indonesia...	{direktor, eksekutif, parameter, pajak, indo...}	{direktor, eksekutif, parameter, pajak, indo...}	{direktor, eksekutif, parameter, pajak, indo...}
1	Wakil Ketua Umum Majelis Ulama Indonesia (MUI)...	Wakil Ketua Umum Majelis Ulama Indonesia (MUI)...	Wakil Ketua Umum Majelis Ulama Indonesia (MUI)...	wakil ketua umum majelis ulama indonesia (mui)...	wakil ketua umum majelis ulama indonesia mui a...	{wakil, ketua, umum, majelis, ulama, indonesia...}	{wakil, ketua, umum, majelis, ulama, indonesia...}	{wakil, ketua, umum, majelis, ulama, indonesia...}
2	Para Para Pejabat Pemertajahan Yang Tidak Perlu...	Para Para Pejabat Pemertajahan Yang Tidak Perlu...	Para Para Pejabat Pemertajahan Yang Tidak Perlu...	para para pejabat pemertajahan yang tidak perlu...	para para pejabat pemertajahan yang tidak perlu...	{para, para, pejabat, pemertajahan, yang, tidak, perlu...}	{para, para, pejabat, pemertajahan, yang, tidak, perlu...}	{para, para, pejabat, pemertajahan, yang, tidak, perlu...}
3	Berobat ke Klinik atau Fasilitas Kesehatan PPN 12% MUI...	Berobat ke Klinik atau Fasilitas Kesehatan PPN 12% MUI...	Berobat ke Klinik atau Fasilitas Kesehatan PPN 12% MUI...	berobat ke klinik atau fasilitas kesehatan ppn 12% mui...	berobat ke klinik atau fasilitas kesehatan ppn 12% mui...	{berobat, ke, klinik, atau, fasilitas, kesehatan, ppn, mui...}	{berobat, ke, klinik, atau, fasilitas, kesehatan, ppn, mui...}	{berobat, ke, klinik, atau, fasilitas, kesehatan, ppn, mui...}
4	Konsep yang sama dengan #PPN12Persen yang salah...	Konsep yang sama dengan #PPN12Persen yang salah...	Konsep yang sama dengan yang salah ppsd1 ter...	konsep yang sama dengan yang salah ppsd1 ter...	konsep yang sama dengan yang salah ppsd1 ter...	{konsep, yang, sama, dengan, yang, salah, ppsd1, ter...}	{konsep, yang, sama, dengan, yang, salah, ppsd1, ter...}	{konsep, yang, sama, dengan, yang, salah, ppsd1, ter...}

Gambar 3. Hasil Preprocessing

f) Labeling Data

Memberikan label sentimen ('positif', 'negatif', atau 'netral') berdasarkan polaritas teks menggunakan metode analisis sentimen, untuk mempersiapkan data bagi model machine learning.

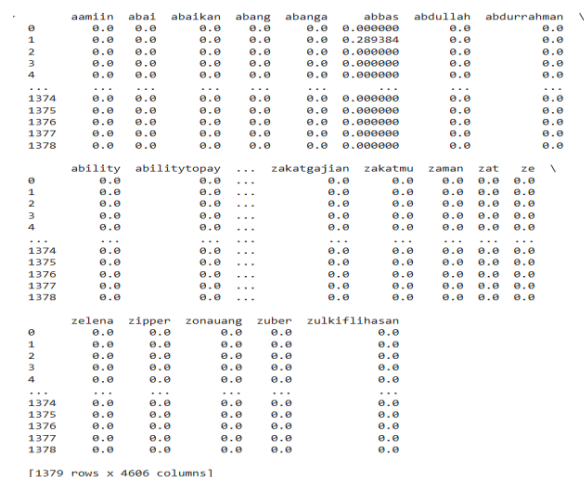


Gambar 4. Visualisasi hasil labelling data

3.3 Transformation

Prosesi transformasi melibatkan mengubah teks asli menjadi format numerik yang siap untuk dianalisis. Proses ini menilai kata-kata untuk menemukan pola saat klasifikasi. TF-IDF (Term Frekuensi – Inverse Document Frekuensi) merupakan salah satu teknik

yang digunakan. Hitung bobot sebuah kata berdasarkan seberapa sering kata tersebut muncul di dokumen dan berapa banyak kata yang lebih jarang muncul di dokumen lain. Ini membantu membedakan sentimen teks (positif, negatif, atau netral). Output transformasi ini adalah representasi numerik yang siap digunakan dalam algoritma klasifikasi [20].



Gambar 5. Representasi TF-IDF

3.4 Data Mining

Analisis sentimen menggunakan metode regresi logistik, naive bayes, dan support vector machine (SVM) dalam proses data mining. Prosesnya diawali dengan memberikan pelatihan model menggunakan data yang divektorisasi dengan metode TF-IDF. Vektorisasi ini memberikan bobot pada setiap kata sehingga model dapat menemukan pola yang sesuai. Logistic Regression bekerja dengan memprediksi probabilitas kelas sentimen (positif, negatif, netral) menggunakan fungsi logistik (sigmoid) untuk mengoptimalkan koefisien model agar memberikan hasil prediksi yang akurat. Naive Bayes menggunakan prinsip Teorema Bayes dengan asumsi independensi antar fitur (kata) untuk menghitung probabilitas setiap kelas. Metode ini sederhana, cepat, dan efisien meskipun data memiliki distribusi yang jarang muncul (sparse). Sementara itu, SVM menggunakan kernel linear untuk membangun hyperplane sebagai batas keputusan optimal yang memisahkan kelas sentimen. Metode ini dikenal efektif dalam menangani data kompleks dan noise. Ketiga algoritma ini memungkinkan analisis sentimen berjalan dengan akurasi tinggi melalui identifikasi pola dalam data teks yang dianalisis.

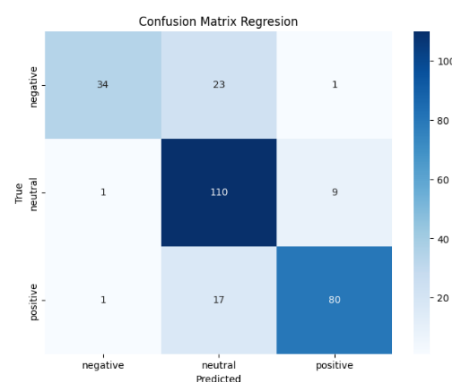
Table 1 Menunjukkan Akurasi Klasifikasi Untuk Setiap Model Yang Digunakan

Model	Akurasi
Logistic Regression	81,15%
Naive Bayes	74,63%
SVM	82,24%

Tabel 2 Menunjukkan metrik *precision*, *recall*, *f1-score*, dan *support* untuk setiap model yan digunakan

Klasifikasi	Precision	Recall	F1-Score	Support
Logistic Regression				
Negative	94%	59%	72%	58
Neutral	73%	92%	81%	120
Positive	89%	82%	85%	98
Naive Bayes				
Negative	74%	67%	70%	58
Neutral	70%	78%	74%	120
Positive	81%	76%	78%	98
SVM				
Negative	86%	66%	75%	58
Neutral	77%	89%	83%	120
Positive	88%	84%	86%	98

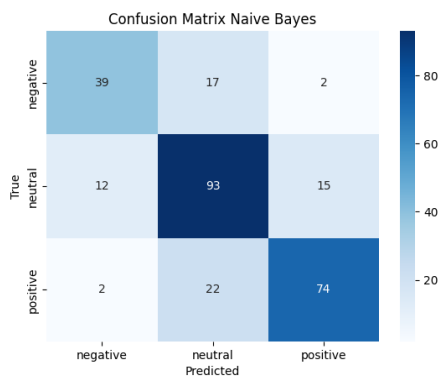
Berdasarkan Tabel 1 dan 2, Model Logistic Regression menunjukkan performa yang kompetitif dengan akurasi 81,15%. Model ini memiliki keunggulan dalam memprediksi probabilitas kelas dan sering digunakan karena kesederhanaan serta efektivitasnya dalam tugas klasifikasi teks. Di sisi lain, model Naive Bayes memiliki akurasi yang sedikit lebih rendah, yaitu 74,63%. Meskipun demikian, model ini tetap menjadi pilihan yang efisien untuk analisis teks karena kecepatan pemrosesan dan kemampuannya menangani data dengan fitur yang saling independen. Model SVM memberikan hasil terbaik secara keseluruhan. Pada kelas Netral, model ini memiliki F1-Score 82%, yang lebih tinggi dibandingkan model lainnya. Model SVM menunjukkan performa terbaik dengan akurasi tertinggi dan nilai F1-Score yang konsisten di semua kelas.



Gambar 6. Confusion Matric Regression

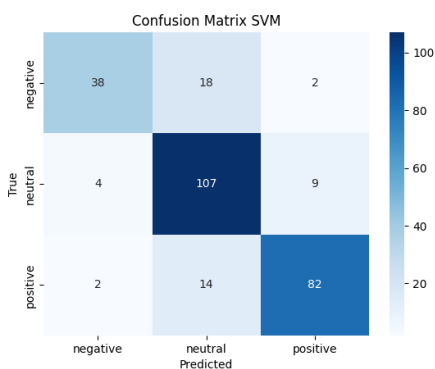
Gambar 6. memperlihatkan confusion matrix hasil klasifikasi sentimen menggunakan algoritma Logistic Regression. Dari hasil tersebut, terlihat bahwa model cukup efektif dalam mengidentifikasi sentimen netral dan positif, ditunjukkan oleh nilai recall masing-masing sebesar 92% dan 82%. Namun, performa model dalam mengenali kelas negatif masih rendah, dengan recall hanya mencapai 59%. Hal ini menunjukkan bahwa banyak data berlabel negatif yang

diklasifikasikan secara keliru sebagai netral atau positif. Kesalahan ini kemungkinan disebabkan oleh kemiripan pola bahasa dalam tweet netral dan negatif yang tidak dapat dipisahkan secara optimal oleh model. Meskipun demikian, model ini tetap memberikan performa yang kompetitif secara keseluruhan dengan akurasi mencapai 81,15%.



Gambar 7. Confusion Matrix Naive Bayes

Gambar 7. menunjukkan confusion matrix dari hasil klasifikasi menggunakan algoritma Naive Bayes. Model ini menampilkan performa yang cukup seimbang dalam mengidentifikasi ketiga kelas sentimen, meskipun dengan akurasi keseluruhan yang lebih rendah dibanding model lainnya. Recall tertinggi tercatat pada kelas netral (78%), diikuti oleh kelas positif (76%), dan negatif (67%). Hal ini menunjukkan bahwa model masih mampu menangkap pola umum dalam data, meskipun terjadi sejumlah kesalahan klasifikasi silang antar kelas.



Gambar 8. Confusion Matrix SVM

Gambar 8. memperlihatkan confusion matrix dari klasifikasi menggunakan algoritma Support Vector Machine (SVM). Hasil ini menegaskan bahwa SVM merupakan model dengan performa terbaik di antara ketiga algoritma yang diuji. Model menunjukkan nilai recall tinggi pada semua kelas, khususnya pada sentimen netral (89%) dan positif (84%), sementara kelas negatif juga teridentifikasi dengan baik (recall 66%). Ketepatan model dalam mengklasifikasikan tiap

kelas tercermin dari nilai F1-score yang konsisten dan tinggi di semua kategori.

Analisis confusion matrix menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi pada kelas Netral dan Positif; ini menunjukkan bahwa model masih mengalami kesulitan membedakan karakteristik antara kedua kelas tersebut. Model regresi logistik menemukan kelas Negatif dengan baik, tetapi tidak dapat menemukan kelas Positif sama sekali. Sementara itu, Naive Bayes dan SVM lebih baik dalam menemukan kelas Negatif.

Secara keseluruhan, hasil analisis menunjukkan bahwa model SVM mampu memberikan kinerja terbaik dalam klasifikasi sentimen teks dibandingkan model lainnya. Namun, masih diperlukan optimasi lebih lanjut untuk meningkatkan akurasi, terutama dalam membedakan kelas Positif dan Netral, yang cenderung menjadi tantangan dalam proses klasifikasi. Upaya perbaikan dapat difokuskan pada penanganan kesalahan klasifikasi dan peningkatan representasi data untuk kelas-kelas yang sulit dibedakan.

4. Kesimpulan

Studi ini mengevaluasi kinerja tiga algoritma klasifikasi, termasuk regresi logistik, Naive Bayes, dan support vector machine (SVM), dalam analisis sentimen tweet mengenai arahan kenaikan PPN 12%. Hasil eksperimen menunjukkan bahwa SVM memiliki akurasi tertinggi sebesar 82,24%, diikuti oleh Logistic Regression (81,15%) dan Naive Bayes (74,63%). Meskipun Logistic Regression kompetitif, SVM lebih stabil dalam membedakan kelas sentimen. Keunggulan SVM terletak pada kemampuannya menemukan hyperplane optimal untuk memisahkan data pada ruang berdimensi tinggi. Naive Bayes juga menunjukkan performa yang baik, meskipun sedikit lebih rendah. Temuan ini memberikan wawasan penting dalam memilih algoritma yang tepat untuk tugas analisis sentimen terhadap tweet, khususnya terkait kebijakan kenaikan PPN.

Daftar Rujukan

- [1] I. Syahrohim, S. D. Saputra, R. W. Saputra, V. H. Pranatawijaya, and R. Priskila, "Perbandingan Analisis Sentimen Setelah Pilpres 2024 Di Twitter Menggunakan Algoritma Machine Learning," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4249.
- [2] R. Bhaskaran *et al.*, "Intelligent Machine Learning with Metaheuristics Based Sentiment Analysis and Classification," *Computer Systems Science and Engineering*, vol. 44, no. 1, pp. 235–247, 2022, doi: 10.32604/csse.2023.024399.
- [3] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

- [4] R. Rita and P. Astuty, "Dampak Kenaikan Tarif Kenaikan Ppn Terhadap Inflasi," *Remittance: Jurnal Akuntansi Keuangan Dan Perbankan*, vol. 4, no. 1, pp. 38–43, 2023, doi: 10.56486/remittance.vol4no1.279.
- [5] S. Agustian *et al.*, "New Directions in Text Classification Research: Maximizing The Performance of Sentiment Classification from Limited Data Arah Baru Penelitian Klasifikasi Teks: Memaksimalkan Kinerja Klasifikasi Sentimen dari Data Terbatas," *Malcom*, pp. 1–10, 2024.
- [6] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Systems with Applications*, vol. 212, no. January 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.
- [7] Ferdian Maulana Akbar, Robby Hermansyah, Sofian Lusa, Dana Indra Sensuse, Nadya Safitri, and Damayanti Elisabeth, "Analisis Sentimen untuk Evaluasi Reputasi Merek Motor XYZ Berkaitan dengan Isu Rangka Motor di Twitter Menggunakan Pendekatan *Machine Learning*," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 3, pp. 647–654, 2024, doi: 10.25126/jtiik.938663.
- [8] W. Zhang, Y. Chang, Y. Ding, Y. Zhu, Y. Zhao, and R. Shi, "To Establish an Early Prediction Model for Acute Respiratory Distress Syndrome in Severe Acute Pancreatitis Using Machine Learning Algorithm," *Journal of Clinical Medicine*, vol. 12, no. 5, 2023, doi: 10.3390/jcm12051718.
- [9] L. Wu *et al.*, "Development of benchmark datasets for text mining and sentiment analysis to accelerate regulatory literature review," *Regulatory Toxicology and Pharmacology*, vol. 137, no. October 2022, p. 105287, 2023, doi: 10.1016/j.yrtph.2022.105287.
- [10] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," *Natural Language Processing Journal*, vol. 8, no. June, p. 100097, 2024, doi: 10.1016/j.nlp.2024.100097.
- [11] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model," *IEEE Access*, vol. 8, pp. 14630–14641, 2020, doi: 10.1109/ACCESS.2019.2963702.
- [12] S. Mansour, "Social media analysis of user's responses to terrorism using sentiment analysis and text mining," *Procedia Computer Science*, vol. 140, pp. 95–103, 2018, doi: 10.1016/j.procs.2018.10.297.
- [13] Y. Yang and N. Li, "Research on Residents' Travel Behavior Based on Multiple Logistic Regression Model," *IEEE Access*, vol. 11, no. July, pp. 74759–74767, 2023, doi: 10.1109/ACCESS.2023.3297497.
- [14] Imamah and F. H. Rachman, "Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regression," *Proceeding - 6th Information Technology International Seminar, ITIS 2020*, pp. 238–242, 2020, doi: 10.1109/ITIS50118.2020.9320958.
- [15] R. Sanjaya, E. Tohidi, E. Wahyudi, and K. Kaslani, "Analisis Sentimen Terhadap Berhentinya Tiktokshop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 507–514, 2024, doi: 10.36040/jati.v8i1.8443.
- [16] F. Meila, A. Sofyan, N. Sulistiyowati, and A. Voutama, "ANALISIS SENTIMEN TERHADAP RESPON PERUBAHAN NAMA TWITTER MENJADI ' X ' MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," vol. 8, no. 5, pp. 10987–10994, 2024.
- [17] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [18] L. Gunawan, M. S. Anggreainy, L. Wihan, Santy, G. Y. Lesmana, and S. Yusuf, "Support vector machine based emotional analysis of restaurant reviews," *Procedia Computer Science*, vol. 216, no. 2022, pp. 479–484, 2022, doi: 10.1016/j.procs.2022.12.160.
- [19] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. Alsobeh, K. Mansour, and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," *IEEE Access*, vol. 12, no. April, pp. 81793–81812, 2024, doi: 10.1109/ACCESS.2024.3382836.
- [20] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Computer Science*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.