Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 8, Agustus 2025, Hal. 2183-2192
DOI: https://doi.org/10.52436/1.jpti.934
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis Sentimen Publik Terhadap Program Lapor Mas Wapres Periode 2024 di Media Sosial X Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM)

Sylvi Anadas[*1], Ryan Randy Suryono[2]

[1]Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
[2]Program Studi Magister Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1]sylvi_anadas@teknokrat.ac.id, [2]ryan@teknokrat.ac.id

Abstrak

Program Lapor Mas Wapres merupakan platform digital yang difasilitasi oleh pemerintah guna menyalurkan aspirasi, keluhan, dan masukan masyarakat diajukan langsung kepada Wakil Presiden Republik Indonesia. Penelitian ini bertujuan untuk mengevaluasi opini masyarakat terhadap program tersebut melalui platform media sosial X dengan menerapkan algoritma klasifikasi Naïve Bayes dan Support Vector Machine (SVM). Dari total 8.108 tweet yang berhasil dikumpulkan, sebanyak 6.967 data digunakan setelah melalui tahap pembersihan dan pra-pemrosesan. Untuk mengatasi ketidakseimbangan kelas sentimen, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), dengan pembagian data 70% untuk pelatihan dan 30% untuk pengujian. Berdasarkan hasil analisis, algoritma Naïve Bayes menunjukkan akurasi awal sebesar 63% sebelum penerapan SMOTE, yang kemudian meningkat menjadi 78% setelah penerapan SMOTE. Sementara itu, algoritma SVM menunjukkan kinerja lebih unggul dengan akurasi sebesar 80% sebelum SMOTE dan meningkat menjadi 91% setelah SMOTE. Sentimen positif publik umumnya dipengaruhi oleh persepsi terhadap keterbukaan informasi dan kemudahan akses program, sedangkan sentimen negatif muncul akibat respons pemerintah yang dianggap lambat atau tidak memadai. Penerapan SMOTE terbukti efektif dalam meningkatkan kinerja model klasifikasi, terutama pada SVM. Temuan ini memberikan wawasan bagi pemerintah untuk meningkatkan efektifitas layanan pengaduan digital dan menyusun strategi komunikasi publik yang lebih responsif.

Kata kunci: *Analisis Sentimen, Media Sosial X, Naive Bayes, Program Lapor Mas Wapres, SMOTE, SVM.*

Public Sentiment Analysis of the Vice President's Lapor Mas Program for the 2024 Period on Social Media X Using the Naive Bayes Method and Support Vector Machine (SVM)

Abstract

The Lapor Mas Wapres program is a digital platform facilitated by the government to channel public aspirations, complaints, and input submitted directly to the Vice President of the Republic of Indonesia. This study aims to evaluate public opinion on the program through the social media platform X by applying the Naïve Bayes and Support Vector Machine (SVM) classification algorithms. Of the total 8,108 tweets that were successfully collected, 6,967 data were used after going through the cleaning and pre-processing stages. To overcome the imbalance of sentiment classes, the Synthetic Minority Over-sampling Technique (SMOTE) technique was applied, with 70% data divided for training and 30% for testing. Based on the analysis results, the Naïve Bayes algorithm showed an initial accuracy of 63% before the application of SMOTE, which then increased to 78% after the application of SMOTE. Meanwhile, the SVM algorithm showed superior performance with an accuracy of 80% before SMOTE and increased to 91% after SMOTE. Positive public sentiment is generally influenced by perceptions of information transparency and ease of program access, while negative sentiment arises due to the government's response which is considered slow or inadequate. The implementation of SMOTE has proven effective in improving the performance of classification models, especially in SVM. This finding provides insight for the government to improve the effectiveness of digital complaint services and develop a more responsive public communication strategy.

Keywords: *Lapor Mas Wapres Program, Naive Bayes, Sentiment Analysis, SMOTE, Social Media X, SVM.*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan signifikan pada berbagai aspek kehidupan, termasuk dalam mengubah interaksi antara pihak pemerintah dan warga masyarakat, salah satunya melalui platform pengaduan berbasis teknologi. Di Indonesia, Program Lapor Mas Wapres adalah kanal resmi yang memungkinkan warga menyampaikan aspirasi, kritik, dan saran langsung kepada Wakil Presiden. Program ini bertujuan meningkatkan transparasi dan akuntabilitas pemerintah melalui komunikasi dua arah yang lebih terbuka dan responsif [1], [2].

Di era digital, Media sosial telah menjadi media utama bagi masyarakat dalam mengekspresikan opini terhadap kebijakan dan layanan publik. Platform seperti X (sebelumnya Twitter) memfasilitasi interaksi publik yang dinamis, dan karenanya menjadi sumber data yang potensial untuk mengukur persepsi publik. Analisis sentimen merupakan metode dalam pemrosesan bahasa alami yang bertujuan untuk mengenali dan mengklasifikasikan opini atau perasaan dalam teks, serta menentukan apakah sentimen tersebut positif, negatif, maupun netral [3], [4].

Dalam konteks pengolahan data teks, metode klasifikasi seperti Naïve Bayes dan Support Vector Machine (SVM) telah banyak digunakan. Naïve Bayes dikenal karena kesederhanaan dan efisiensinya dalam memproses data besar [5], sementara SVM unggul dalam klasifikasi dengan margin optimal dan sering menghasilkan akurasi tinggi [6], [7]. Namun, tantangan seperti ketidakseimbangan kelas sentimen dalam data sering mengurangi kinerja model. Untuk mengatasinya, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) telah digunakan secara luas dan terbukti dalam meningkatkan representasi kelas minoritas [8], [9], [10].

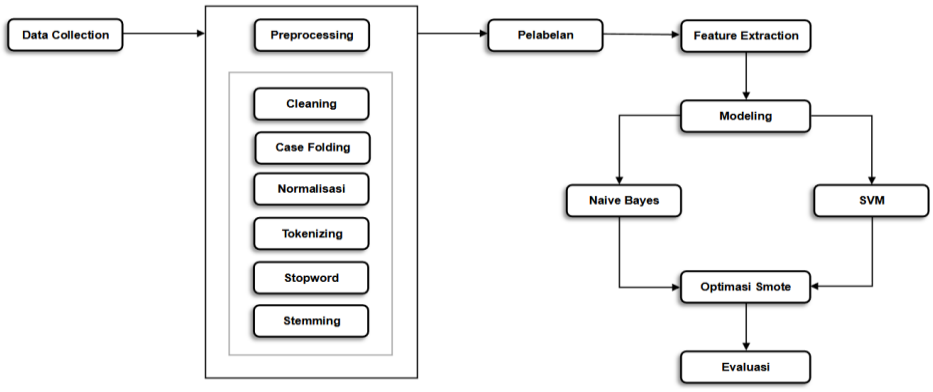
Meskipun algoritma Naïve Bayes dan SVM telah banyak digunakan dalam penelitian analisis sentimen, sebagian besar fokus pada konteks umum seperti opini terhadap produk komersial, isu politik dan aplikasi digital [4]. Belum banyak penelitian yang secara khusus mengkaji sentimen publik terhadap program pengaduan digital pemerintah seperti Lapor Mas Wapres, apalagi dengan mengombinasikan kedua algoritma tersebut bersama SMOTE secara simultan. Selain itu, studi sebelumnya cenderung menitikberatkan pada evaluasi performa model tanpa mengaitkan langsung hasil klasifikasi dengan implikasi kebijakan publik [6], [9]. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan pendekatan analisis sentimen berbasis machine learning yang tidak hanya mengevaluasi kinerja algoritma, tetapi juga memberikan rekomendasi strategis bagi pengambil kebijakan berdasarkan persepsi publik yang terukur.

Secara praktis, hasil penelitian ini diharapkan dapat memberikan masukan bagi pemerintah dalam meningkatkan efektivitas layanan pengaduan digital, serta menjadi dasar dalam menyusun strategi komunikasi publik yang lebih responsif dan berbasis data. Selain itu, pendekatan berbasis machine learning yang digunakan juga dapat menjadi acuan teknis dalam pengembangan sistem monitoring opini publik secara otomatis.

Tujuan dari penelitian ini adalah untuk menganalisis sentimen publik terhadap Program Lapor Mas Wapres di media sosial X menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM), serta mengevaluasi efektivitas penerapan teknik SMOTE dalam meningkatkan kinerja klasifikasi.

2. METODE PENELITIAN

Tujuan dari penelitian ini untuk mengevaluasi sentimen publik terkait Program Lapor Mas Wapres melalui media sosial X (sebelumnya Twitter) dengan menggunakan dua algoritma klasifikasi teks, yaitu Naïve Bayes dan Support Vector Machine (SVM). Pendekatan kuantitatif diterapkan dalam penelitian ini dengan beberapa tahapan sistematis, mulai dari pengumpulan data hingga evaluasi kinerja model. Adapun diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.1 Pengumpulan Data

Data diperoleh dengan metode web scraping dari media sosial X memanfaatkan kata kunci yang relevan dengan "Lapor Mas Wapres". Proses scraping dilakukan untuk mengumpulkan tweet yang mengandung opini atau tanggapan publik terhadap program tersebut. Data dikumpulkan dalam rentan waktu 11 November 2024 hingga 10 Februari 2025. Berdasarkan hasil pengumpulan data, diperoleh sebanyak 8.108 data yang kemudian disimpan dalam format *Comma-Separated Values* (CSV).

2.2 Preprocessing

Tahapan *preprocessing* data teks merupakan tahap penting dalam mempersiapkan data mentah menjadi format yang lebih terstruktur dan siap untuk dianalisis lebih lanjut [11]. Proses ini ditunjukan untuk memperbaiki kualitas data dengan menghapus elemen-elemen yang tidak relevan atau mengganggu, agar analisis model dapat berjalan dengan lebih optimal dan efisiensi.

1. Cleaning

Pembersihan data menghapus karakter-karakter khusus, angka, tanda baca, dan symbol yang tidak diperlukan dalam analisis. Pada tahap ini, karakter atau elemen yang tidak relevan untuk dianalisis dihapus, seperti "@" dan "?". Karakter non-tekstual dan simbol umumnya tidak memiliki makna sentimen dan dapat menyebabkan noise pada proses klasifikasi. Tabel 1 hasil contoh dari cleaning:

Tabel 1. *Cleaning*

Tweet	Cleaning
@FayaAtika @gibran_tweet lapor mas wapres klo ada apa-apa itu bilang nya kmn ???	lapor mas wapres klo ada apaapa itu bilang nya kmn

2. Case Folding

Proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menjamin konsistensi terhadap pengolahan data, sehingga kata-kata seperti "Lapor" dan "lapor" dianggap sama oleh sistem. Tabel 2 hasil contoh dari case folding:

Tabel 2. *Case Folding*

Cleaning	Case Folding
Lapor Mas Wapres keknya sdh alm Skrg blng Nanti kalau ada apaapa Emangnya kek lapor ke Kelurahan Duh gini amat	lapor mas wapres keknya sdh alm skrg blng nanti kalau ada apaapa emangnya kek lapor ke kelurahan duh gini amat

3. Normalisasi

Pada tahap ini, kata-kata yang memiliki bentuk tidak baku atau singkatan diubah menjadi bentuk standar. Misalnya, kata "gimana" diubah menjadi "bagaimana" [12]. Normalisasi penting agar kata-kata dengan makna sama tapi bentuk berbeda tidak dianggap entitas yang berbeda oleh algoritma klasifikasi. Tabel 3 hasil contoh dari normalisasi:

Tabel 3. *Normalisasi*

Case Folding	Normalisasi
lapor mas wapres gimana kabarnya tuh	lapor mas wapres bagaimana kabarnya tuh

4. Tokenizing

Proses memecah kalimat menjadi kata-kata terpisah untuk setiap kata agar lebih mudah dianalisis oleh model *machine learning* [6]. Misalnya, kalimat "lapor mas wapres" akan diubah menjadi kata: ['lapor', 'mas', 'wapres']. Tabel 4 hasil contoh dari tokenizing:

Tabel 4. *Tokenizing*

Normalisasi	Tokenizing
sih ibu harus ya tanya ke mas ngabarin ya kemana mau kabarin ke lapor mas wapres sudah tak aktif	['sih', 'ibu', 'harus', 'ya', 'tanya', 'ke', 'mas', 'ngabarin', 'ya', 'kemana', 'mau', 'kabarin', 'ke', 'lapor', 'mas', 'wapres', 'sudah', 'tak', 'aktif']

p-ISSN: 2775-4227
e-ISSN: 2775-4219

5. *Stopword*

Pada tahap ini, kata-kata umum yang sering muncul dalam teks namun tidak memberikan kontribusi signifikan terhadap analisis, seperti kata penghubung, preposisi, atau kata-kata umum lainnya. Misalnya "kalau", " sama", dan "apa". Menghapus *stopword* membantu model fokus pada kata-kata bermakna yang berkontribusi terhadap pengklasifikasian sentimen. Tabel 5 hasil contoh dari stopword:

Tabel 5. *Stopword*

Tokenizing	Stopword
['iya', 'jawabnya', 'sama', 'pas', 'lapor', 'mas', 'wapres', 'kalau', 'ada', 'apa', 'bilang']	['iya', 'pas', 'lapor', 'mas', 'wapres', 'bilang']

6. Stemming

Mengubah kata ke bentuk dasarnya dengan menghapus imbuhan. Misalnya, "didengarkan" menjadi "dengar". Langkah ini menyatukan variasi kata agar model mengenali mereka sebagai satu entitas, meningkatkan efisiensi analisis. Tabel 6 hasil contoh dari stemming:

Tabel 6. *Stemming*

Stopword	Stemming
['didengarkan', 'lapor', 'mas', 'wapres', 'kagak', 'ngada', 'mengadakan', 'program', 'biar', 'mengasih', 'pekerjaan', 'timsesnya', 'palingan']	dengar lapor mas wapres kagak ngada ada program biar asih kerja timsesnya paling

2.3 Pelabelan Data

Setelah tahap pra-pemrosesan, jumlah tweet yang semula sebanyak 8.108 berkurang menjadi 6.967. Proses pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* dengan memanfaatkan kamus sentimen InSet (Indonesia Sentiment Lexicon) [13]. Setiap kata dalam tweet dibandingkan dengan entri dalam kamus InSet yang sudah dikategorikan sebagai positif, negatif, atau netral. Jika jumlah kata positif lebih banyak daripada kata negatif, tweet diklasifikasikan sebagai sentimen positif. Sebaliknya, jika kata negatif lebih dominan, tweet diberi label negatif. Apabila jumlah kata positif dan negatif seimbang atau tidak ada kata yang menunjukkan sentimen, tweet dikategorikan sebagai netral. Tabel 7 hasil contoh pelabelan data:

Tabel 7. Pelabelan Data

Tweet	Sentimen
netizen minimal pakai otak mas wapres sibuk capek rutinitas susu merespon lapor lapor mas wapres tunggu cari kerja mas gibran juta lapang kerja	Positif
irma chaniago mahluk bingung buka bilang partai korup cari nafkah kader partai korup nasdem mati bela program gibran lapor mas wapres bukti juntrung membabibu	Negatif
suruh lbp reshuffle tinggal lapor mas wapres	Netral

2.4 Feature Extraction

Penelitian ini menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk mengubah teks menjadi representasi numerik. TF-IDF menghitung bobot kata berdasarkan frekuensinya dalam dokumen dan kelangkaannya di seluruh korpus. Alasan pemilihan TF-IDF adalah kemampuannya dalam menyeimbangkan kata yang sering muncul dengan kata penting namun jarang, sehingga cocok untuk klasifikasi sentimen.

2.5 Modeling

Pada tahap pemodelan penelitian ini, digunakan dua algoritma machine learning yang umum untuk analisis sentimen, yaitu Naïve Bayes dan Support Vector Machine (SVM). Kedua algoritma ini telah banyak digunakan dalam klasifikasi teks dan memiliki ciri-ciri yang berbeda dalam pendekatannya. Model dilatih menggunakan 70% data pelatihan dan 30% data uji. Data pelatihan digunakan untuk memberikan pelatihan terhadap model dalam tahap klasifikasi, sedangkan data uji diterapkan dalam melakukan evaluasi model dalam mengklasifikasikan data baru. Berikut adalah penjelasan mengenai algoritma yang diterapkan dalam penelitian ini:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

1. Naïve Bayes

Algoritma klasifikasi probabilistik yang mengandung independensi antar fitur. Dalam konteks analisis sentimen, algoritma ini menghitung probabilitas suatu dokumen termasuk dalam kelas tertentu berdasarkan distribusi kata dalam dataset. Keunggulan utama Naïve Bayes terletak pada kemudahan dan efisiensinya dalam memproses data dalam jumlah besar [8].

2. Support Vector Machine (SVM)

Algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan data dari kelas yang berbeda dengan margin maksimum. Dalam analisis sentimen, SVM bekerja dengan mengubah teks data menjadi fitur vektor, kemudian menentukan batas terbaik antara kelas-kelas sentimen. SVM dikenal memiliki performa yang baik untuk menangani data berdimensi tinggi serta dapat memberikan akurasi yang tinggi dalam klasifikasi teks [9].

2.6 SMOTE

Dataset pada penelitian ini menunjukkan ketidakseimbangan jumlah antar kelas sentimen, di mana kelas positif jauh lebih dominan dibandingkan kelas negatif dan netral. Ketidakseimbangan ini berpotensi menurunkan akurasi model klasifikasi terhadap kelas minoritas. Untuk mengatasi hal tersebut, penelitian ini menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE bekerja dengan cara menghasilkan data sintetis baru pada kelas minoritas melalui interpolasi terhadap sejumlah tetangga terdekat. Dalam penelitian ini, SMOTE diterapkan menggunakan parameter default, yaitu (k_neighbors = 5) dan (sampling_strategy = 'auto'), yang secara otomatis menyamakan jumlah data pada setiap kelas. Pemilihan SMOTE didasarkan pada keefektifannya dalam meningkatkan representasi data minoritas tanpa harus menghapus data mayoritas. Teknik ini juga sesuai untuk data numerik hasil ekstraksi TF-IDF dan telah terbukti meningkatkan performa model, khususnya dalam metrik recall dan F1-score pada kelas negatif dan netral.

2.7 Evaluasi

Pada tahap evaluasi model analisis sentimen, digunakan beberapa metrik untuk menilai kinerja algoritma klasifikasi Naive Bayes dan Support Vector Machine (SVM). Metrik-metrik ini membantu untuk memahami seberapa baik model dalam mengklasifikasikan data teks ke dalam kategori sentimen yang benar. Berikut adalah matrik evaluasi yang umum digunakan:

1. Akurasi untuk mengukur proporsi prediksi umum yang benar terhadap total prediksi.
2. Presisi untuk mengukur keakuratan model dalam mengklasifikasikan sentimen positif, yakni berapa banyak prediksi positif yang benar.
3. Recall untuk mengukur kemampuan model dalam mendeteksi semua instance positif yang sebenarnya ada.
4. F1-Score merupakan rata-rata harmonis dari presisi dan recall. F1-Score memberikan keseimbangan antara presisi dan recall, terutama penting ketika distribusi kelas tidak seimbang [14].

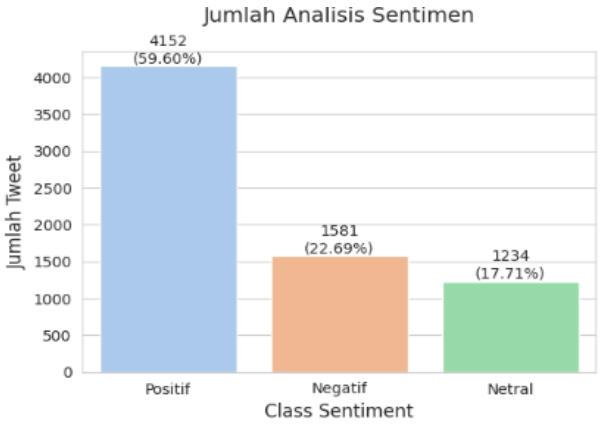
3. HASIL DAN PEMBAHASAN

3.1 Hasil

Penelitian ini menggunakan sebanyak 6.967 tweet hasil pra-pemrosesan dari media sosial X yang berkaitan dengan Program Lapor Mas Wapres. Data tersebut dikategorikan ke dalam tiga jenis sentiment, yaitu positif, negatif, dan netral, dengan menggunakan pendekatan *lexicon-based*. Proses pelabelan dilakukan dengan mencocokkan kata-kata dalam tweet dengan daftar kata berkonotasi positif dan negatif yang terdapat pada kamus sentimen InSet [15]. Distribusi hasil pelabelan ditampilkan pada Gambar 2.

Dari tahapan tersebut diperoleh 4.152 tweet dengan sentimen positif, 1.581 tweet dengan sentimen negatif, dan 1.234 tweet dengan sentimen netral. Setelah proses pelabelan, data digunakan untuk pelatihan model klasifikasi menggunakan dua algoritma, yaitu Naïve Bayes dan Support Vector Machine (SVM). Mengingat adanya ketidakseimbangan jumlah data pada setiap kelas sentimen, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) sebelum proses pelatihan model untuk menyeimbangkan distribusi data.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 2. Jumlah Analisis Sentimen

3.1.1 Hasil Klasifikasi Menggunakan Naïve Bayes

Model Naïve Bayes memberikan akurasi sebesar 63% sebelum penerapan SMOTE. Setelah dilakukan penyeimbangan data menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), akurasi meningkat menjadi 78%. Meskipun terjadi peningkatan signifikan, model ini menunjukkan kelemahan pada kelas sentimen netral, yang terlihat dari rendahnya nilai recall. Hal ini menunjukkan bahwa Naïve Bayes cenderung bias terhadap kelas mayoritas dan kurang akurat dalam mengenali variasi sentimen yang lebih halus seperti netral.

3.1.2 Hasil Klasifikasi Menggunakan Support Vector Machine

Model SVM menghasilkan akurasi awal yang lebih tinggi dibandingkan Naïve Bayes, yaitu 80%. Setelah SMOTE diterapkan, akurasi meningkat menjadi 91%. Selain peningkatan pada akurasi, nilai F1-score untuk kelas negatif dan netral juga menunjukkan perbaikan signifikan. Hal ini mengindikasikan bahwa SVM lebih stabil dan mampu memisahkan kelas-kelas sentimen dengan lebih baik, bahkan ketika data awalnya tidak seimbang.

3.1.3 Evaluasi Kinerja Model

Setelah proses persiapan data, dilakukan evaluasi dengan membandingkan algoritma Naïve Bayes dan Support Vector Machine (SVM). Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Untuk mengatasi ketidakseimbangan kelas, diterapkan teknik SMOTE yang menghasilkan sampel sintetis untuk kelas minoritas. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Tabel 8. Merupakan hasil pengujian sebelum dan setelah penerapan SMOTE.

Tabel 8. Perbandingan Hasil Klasifikasi Sentimen Sebelum dan Setelah SMOTE

Matriks	NB	NB + SMOTE	SVM	SVM + SMOTE
Accuracy	0.63	0.78	0.80	0.91
Sentimen Positif				
Precision	0.62	0.84	0.86	0.95
Recall	1.00	0.73	0.95	0.86
F1-Score	0.76	0.78	0.90	0.90
Sentimen Negatif				
Precision	0.90	0.72	0.73	0.95
Recall	0.12	0.95	0.77	0.95
F1-Score	0.21	0.82	0.75	0.95
Sentimen Netral				
Precision	0.75	0.81	0.59	0.84
Recall	0.01	0.66	0.33	0.92
F1-Score	0.02	0.73	0.43	0.88

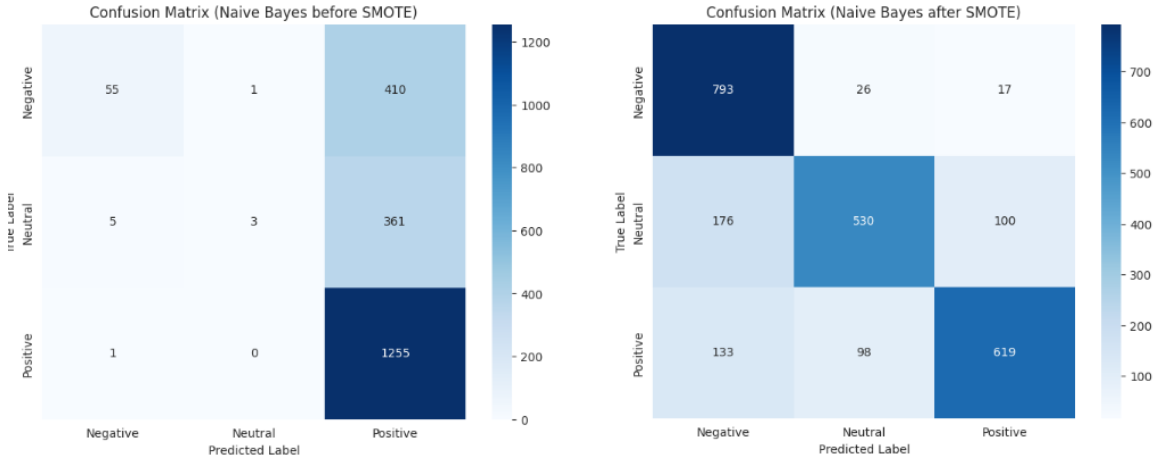
Berdasarkan hasil evaluasi, penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) terbukti meningkatkan akurasi pada kedua algoritma yang digunakan, yaitu Naïve Bayes dan Support Vector Machine (SVM). Peningkatan yang paling signifikan terlihat pada algoritma SVM, yang menunjukkan akurasi tertinggi sebesar 0.91 penerapan SMOTE. Hal ini mengidentifikasikan bahwa SMOTE efektif dalam mengatasi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

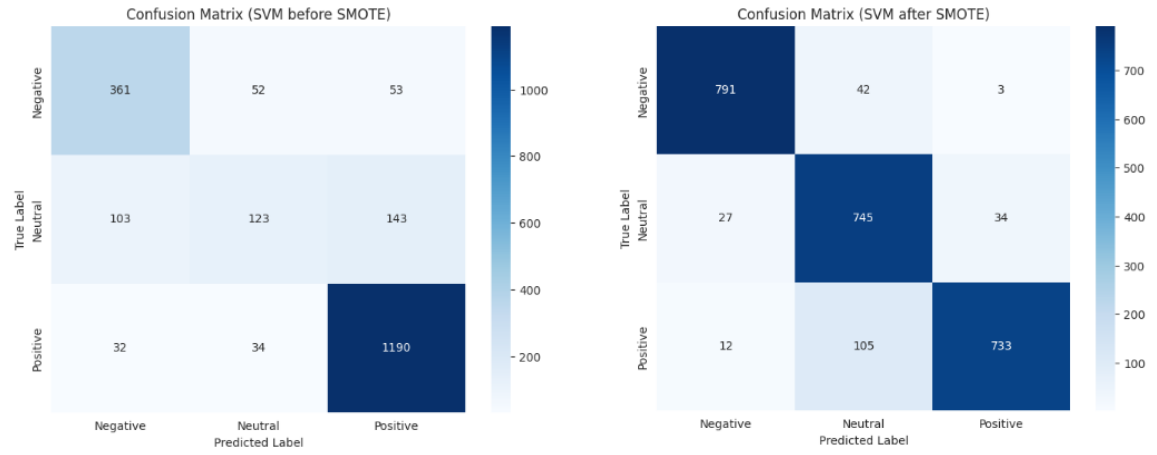
ketidakseimbangan kelas pada data, sehingga memperkuat kamampuan model dalam mengklasifikasikan sentimen data yang tidak seimbang. Secara umum, penerapan SMOTE juga meningkatkan kinerja model dalam mengklasifikasikan sentimen negatif dan netral, terutama pada metrik recall dan F1-score. Sementara itu, presisi pada sentimen positif dan negatif memperlihatkan peningkatan yang signifikan pada algoritma SVM setelah penerapan SMOTE.

Dalam rangka evaluasi kinerja model klasifikasi, penelitian ini melakukan analisis perbandingan dengan menggunakan *Confusion Matrix*. *Confusion matriks* digunakan untuk menilai sejauh mana kedua algoritma yang diterapkan dapat mengklasifikasikan data dengan tepat serta mengidentifikasi jenis kesalahan klasifikasi yang terjadi [16].

Confusion Matrix merupakan alat evaluasi yang sangat berguna dalam analisis klasifikasi model. Ini membantu memahami bagaimana model mengklasifikasikan data ke dalam tiga kelas sentimen: positif, negatif, dan netral. Dengan membandingkan model prediksi dengan nilai aktual dari data uji, kita dapat mengetahui seberapa akurat model dalam mengklasifikasikan sentimen. *Confusion Matrix* diterapkan untuk memberikan pemahaman yang komprehensif terhadap performa model klasifikasi, khususnya dalam konteks penanganan ketidakseimbangan data. Visualisasi hasil *Confusion Matrix* dapat dilihat pada gambar dibawah ini.



Gambar 3. *Confusion Matrix Naive Bayes* Sebelum Dan Sesudah Smote

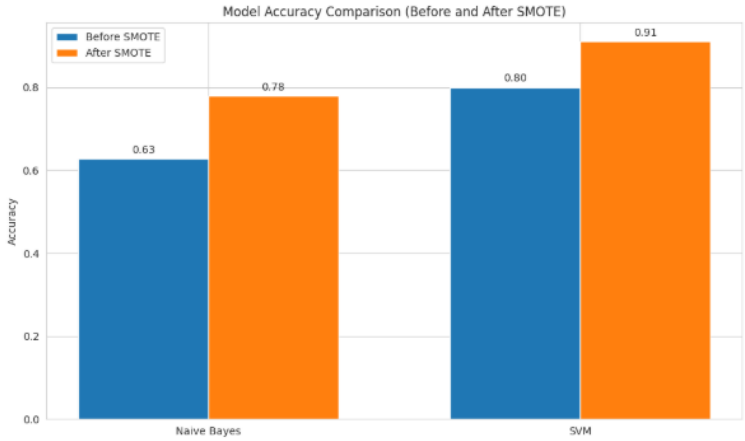


Gambar 4. *Confusion Matrix Support Vector Machine* Sebelum Dan Sesudah Smote

Evaluasi performa model klasifikasi dilakukan dengan menggunakan *confusion matrix* untuk memahami dampak penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) terhadap distribusi prediksi pada masing-masing kelas sentimen [17]. Pada algoritma Naïve Bayes, terjadi penurunan jumlah prediksi benar pada kelas positif dari 1.255 menjadi 619 setelah penerapan SMOTE. Sebaliknya, prediksi benar pada kelas negatif meningkat signifikan dari 410 menjadi 793, dan pada kelas netral meningkat dari 316 menjadi 530. Perubahan ini menunjukkan bahwa SMOTE membantu Naïve Bayes untuk meningkatkan kapasitas klasifikasi pada kelas negatif dan netral, meskipun terjadi penurunan pada kelas positif. Sementara itu, pada algoritma Support Vector Machine (SVM), jumlah prediksi benar pada kelas positif menurun dari 1.190 menjadi 733 setelah penerapan SMOTE.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Namun, prediksi benar pada kelas negatif meningkat dari 361 menjadi 791, dan pada kelas netral meningkat signifikan dari 143 menjadi 745. Hasil ini menunjukkan bahwa penerapan SMOTE pada SVM secara substansial mengoptimalkan performa model dalam mengklasifikasikan kelas netral dan negatif, meskipun terdapat penurunan pada kelas positif. Secara keseluruhan, penerapan SMOTE efektif dalam mengatasi ketidakseimbangan kelas pada data, terutama dalam meningkatkan performa klasifikasi pada kelas minoritas. Namun, perlu diperhatikan bahwa peningkatan pada kelas tertentu dapat disertai dengan penurunan pada kelas lainnya, sehingga pemilihan dan penerapan teknik penyeimbangan data perlu disesuaikan dengan kebutuhan khusus dari penelitian yang dilakukan. Selanjutnya yaitu visualisasi perbandingan akurasi model sebelum dan sesudah SMOTE dapat dilihat pada Gambar 5.



Gambar 5. Visualisasi Perbandingan Akurasi Model

Dari grafik tersebut, bisa dilihat bahwa penerapan SMOTE mempengaruhi positif pada peningkatan akurasi kedua algoritma. Peningkatan paling mencolok terjadi pada model SVM, dengan akurasi yang meningkat dari 80% menjadi 91%. Sementara itu, Naive Bayes juga menunjukkan perbaikan performa yang signifikan, dengan kenaikan akurasi dari 63% menjadi 78%. Hasil ini menunjukkan bahwa SMOTE mampu memperbaiki distribusi kelas yang tidak seimbang dan secara keseluruhan meningkatkan kinerja model dalam mengklasifikasikan sentimen secara lebih akurat.

3.1.4 Visualisasi WordCloud

Visualisasi WordCloud merupakan representasi visual dari data teks yang menampilkan kata-kata dengan ukuran yang bervariasi sesuai dengan frekuensi kemunculannya dalam korpus data. Semakin sering sebuah kata muncul, semakin besar ukuran font-nya dalam visualisasi tersebut [18]. Metode ini efektif untuk mengidentifikasi kata-kata yang paling dominan dalam kumpulan teks data, yang memungkinkan untuk mempermudah peneliti dalam menangkap inti pembahasan dan pola tematik dari data yang kompleks. Pada penelitian ini, Wordcloud digunakan untuk memvisualisasikan kata-kata yang paling sering muncul dalam tweet terkait program "Lapor Mas Wapres" setelah melalui proses pra-pemrosesan dan pelabelan sentimen. Visualisasi ini bertujuan untuk memberikan pemahaman menyeluru mengenai topik-topik yang paling sering dibicarakan oleh masyarakat sehubung dengan program lapor mas wapres. Hasil visualisasi wordcloud dapat dilihat pada Gambar 6.

Gambar 6. Hasil Visualisasi WordCloud

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.2 Diskusi

Hasil penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa klasifikasi sentimen yang lebih unggul dibandingkan Naïve Bayes. Peningkatan akurasi SVM dari 80% menjadi 91% setelah penerapan SMOTE membuktikan efektivitasnya dalam menangani data tidak seimbang, terutama untuk sentimen minoritas seperti negatif dan netral. Sementara itu, Naïve Bayes mengalami peningkatan dari 63% menjadi 78%, tetapi masih memiliki kelemahan dalam mendeteksi kelas netral secara akurat.

Temuan ini sejalan dengan penelitian dalam [7] yang menyimpulkan bahwa SVM lebih stabil dibandingkan Naïve Bayes dalam klasifikasi opini pada ulasan game, terutama ketika data memiliki distribusi kelas yang tidak seimbang. Selain itu, penelitian dalam [8] juga menemukan bahwa penerapan SMOTE pada data sentimen aplikasi digital dapat meningkatkan performa SVM secara signifikan.

Secara teknis, keunggulan SVM terletak pada kemampuannya memisahkan kelas dengan margin optimal, sementara Naïve Bayes cenderung lemah karena mengasumsikan independensi antar fitur. Implikasi praktisnya, model SVM dapat dimanfaatkan untuk meningkatkan efektivitas pemantauan opini publik secara digital. Secara teoretis, temuan ini mendukung bahwa kombinasi preprocessing yang tepat, TF-IDF, dan SMOTE merupakan pendekatan yang kuat dalam klasifikasi teks.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa klasifikasi sentimen yang lebih unggul dibandingkan Naïve Bayes, khususnya setelah penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Kombinasi antara preprocessing, ekstraksi fitur menggunakan TF-IDF, serta penyeimbangan data melalui SMOTE terbukti efektif dalam meningkatkan akurasi klasifikasi sentimen publik terhadap Program Lapor Mas Wapres. Model SVM menunjukkan kemampuan yang lebih baik dalam mengenali sentimen minoritas negatif dan netral, dengan peningkatan akurasi dari 80% menjadi 91% setelah SMOTE diterapkan.

Temuan ini memberikan kontribusi praktis dalam perancangan sistem pemantauan opini publik berbasis media sosial, yang dapat dimanfaatkan oleh pemerintah untuk memperkuat strategi komunikasi digital yang responsif dan berbasis data. Secara teoritis, hasil ini mendukung pendekatan machine learning dalam analisis sentimen sebagai alat evaluasi kebijakan publik. Untuk penelitian selanjutnya, disarankan mengeksplorasi pendekatan *ensemble learning* atau model *deep learning* seperti LSTM atau BERT, serta memperluas sumber data dari berbagai platform digital untuk memperoleh gambaran opini publik yang lebih komprehensif dan representatif.

DAFTAR PUSTAKA

[1] R. Alma Dwi Mayasari and E. Widowati, “Kelayakan Aplikasi Pengaduan Berbasis Android dalam Pencegahan Bahaya Psikososial Bagian Jurnalistik Perusahaan Media X,” *J. Heal. Sains*, vol. 2, no. 7, pp. 932–941, 2021, doi: 10.46799/jhs.v2i7.225.

[2] M. Massa and T. Keheningan, “Kata Kunci: Lapor Mas Wapres ; Media Massa; Teori Keheningan,” vol. 2, no. 1, pp. 120–127, 2025.

[3] A. Zein, S. Farizy, and E. Suharyanto, “Sentimen Analisis Pada Komentar Pendek Evaluasi Dosen Oleh Mahasiswa (Edom) Program Studi Sistem Informasi Universitas Pamulung,” *J. Ilmu Komput.*, vol. 5, no. 01, pp. 17–23, 2022, [Online]. Available: https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/113%0Ahttps://jurnal.pranataindonesia.ac.id/index.php/jik/article/download/113/66

[4] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.

[5] Diana Dwi Rahayu, Muhammad Fatchan, and Alfonsus Ligouri, “Analisis Sentimen Twitter Terpilihnya Prabowo - Gibran Menggunakan Metode Neural Network,” *Tematik*, vol. 11, no. 1, pp. 85–91, 2024, doi: 10.38204/tematik.v11i1.1943.

[6] D. N. Agustia, R. R. Suryono, U. T. Indonesia, L. Ratu, and K. B. Lampung, “COMPARISON OF NAÏVE BAYES , RANDOM FOREST , AND LOGISTIC REGRESSION ALGORITHMS FOR SENTIMENT ANALYSIS ONLINE GAMBLING KOMPARASI ALGORITMA NAÏVE BAYES , RANDOM FOREST , DAN LOGISTIC REGRESION UNTUK ANALISIS,” vol. 10, no. 1, pp. 284–295, 2025.

[7] M. Safrudin, M. Martanto, and U. Hayati, “Perbandingan Kinerja Naïve Bayes Dan Support Vector

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Machine Untuk Klasifikasi Sentimen Ulasan Game Genshin Impact," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3182–3188, 2024, doi: 10.36040/jati.v8i3.8415.

[8] "Eskiyaturrofikoh" and R. R. 'Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm)," *JIPI(Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1408–1419, 2024, [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/5392

[9] A. Kharisma and I. Ernawati, "Sentimen Analisis Opini Masyarakat Jakarta Pada Kinerja Pemerintah Jakarta Terhadap Isu Tenggelamnya Jakarta Menggunakan Algoritma Support Vector Machine," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 488–497, 2023.

[10] M. Mustaqim, B. Warsito, and B. Surarso, "Combination of synthetic minority oversampling technique (Smote) and backpropagation neural network to handle imbalanced class in predicting the use of contraceptive implants," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 2, pp. 116–127, 2019, doi: 10.26594/register.v5i2.1705.

[11] M. Riswan, A. Primajaya, A. Susilo, and Y. Irawan, "ANALISIS SENTIMEN TERHADAP PEMBERITAAN HASIL REKAPITULASI PEMILU PRESIDEN 2024 PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN NAIVE," vol. 13, no. 1, 2025.

[12] D. Sebastian and K. A. Nugraha, "Sistem Perbaikan Kata Tidak Baku Bahasa Indonesia Menggunakan Metode Crowdsourcing," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 3, pp. 386–396, 2020, doi: 10.28932/jutisi.v5i3.1983.

[13] M. Al Khadafi, Kurnia Paranitha Kartika, and Filda Febrinita, "Penerapan Metode Naïve Bayes Classifier Dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada Bpjs," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 725–733, 2022, doi: 10.36040/jati.v6i2.5633.

[14] M. Akrom, "KOMPARASI SVM KLASIK DAN KUANTUM DALAM KLASIFIKASI BINER BIJI GANDUM (SEEDS) COMPARISON OF CLASSICAL AND QUANTUM SVM IN BINARY," vol. 9, no. 1, pp. 49–58, 2024.

[15] S. A. Nugraha, "PENERAPAN LEXICON BASED UNTUK ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP DANANTARA," vol. 9, no. 3, pp. 4949–4957, 2025.

[16] F. Budiman and Y. M. Awaludin, "Optimasi Analisis Kesuburan Tanah dengan Pendekatan Soft Voting Ensemble," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 261–276, 2023, doi: 10.24176/simet.v14i2.11285.

[17] F. N. Salsabilla *et al.*, "PRESIDEN JOKOWI PADA MEDIA SOSIAL X," vol. 8, pp. 106–115, 2025.

[18] Novirianto, "Analisis Sentimen Berbasis Aspek Pada UlasanAplikasi Mybluebird Dengan Implementasi N-GramDan Algoritma Logistic Regression," 2023, [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/76656/1/IQBAL FARIZ NOVIRIANTO-FST.pdf