SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 8, Nomor 2, Juli 2025, Halaman 257-268 E-ISSN: 2721-4788

Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Terhadap Penggunaan *True Wireless Stereo* (TWS)

Risca Lusiana Pratiwi[1], Zulia Imami Alfianti[2], Ahmad Fauzi[3], Ginabila[4*]
[1]Fakultas Teknologi dan Informasi, Sistem Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia
[2,3,4]Fakultas Teknik dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
E-mail: [1]risca.ral@nusamandiri.ac.id, [2]zulia.zim@bsi.ac.id, [3]ahmad.aau@bsi.ac.id, [4*]gina.gnb@bsi.ac.id
(*: corresponding author)

Abstrak

Penggunaan perangkat audio nirkabel *True Wireless Stereo* (TWS) semakin meningkat sebagai alternatif earphone berkabel di Indonesia. Untuk memahami persepsi publik terhadap TWS, penelitian ini melakukan analisis sentimen terhadap opini pengguna di media sosial X. Dua algoritma pembelajaran mesin, yaitu Naive Bayes dan Support Vector Machine (SVM), digunakan untuk klasifikasi sentimen. Masalah ketidakseimbangan data diatasi dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE), sementara performa model dioptimalkan menggunakan *Particle Swarm Optimization* (PSO). Hasil penelitian menunjukkan bahwa SVM memiliki kinerja lebih unggul dengan akurasi 80,46% dan nilai *Area Under Curve* (AUC) sebesar 0,854, serta distribusi *precision* dan *recall* yang seimbang. Sementara itu, Naive Bayes unggul dalam mendeteksi sentimen negatif, namun memiliki akurasi lebih rendah sebesar 78,00% dan AUC 0,780. Kontribusi penelitian ini mencakup pengembangan model analisis sentimen berbasis media sosial yang efisien, penerapan teknik penyeimbangan dan optimasi data dalam klasifikasi, serta memberikan wawasan kepada produsen dan konsumen mengenai persepsi publik terhadap TWS.

Kata kunci: *True Wireless Stereo* (TWS), Analisis Sentimen, Naive Bayes, Support Vector Machine, SMOTE.

Abstract

The use of True Wireless Stereo (TWS) audio devices is increasingly popular in Indonesia as an alternative to wired earphones. To understand public perception of TWS, this study conducts sentiment analysis on user opinions expressed on the X social media platform. Two machine learning algorithms, Naive Bayes and Support Vector Machine (SVM), are employed for sentiment classification. The issue of data imbalance is addressed using the Synthetic Minority Oversampling Technique (SMOTE), while model performance is optimized using Particle Swarm Optimization (PSO). The results show that SVM outperforms Naive Bayes with an accuracy of 80.46% and an Area Under Curve (AUC) score of 0.854, along with a balanced distribution of precision and recall across sentiment classes. In contrast, Naive Bayes performs better in detecting negative sentiment but yields a lower accuracy of 78.00% and AUC of 0.780. The contributions of this study include the development of an efficient social media-based sentiment analysis model, the application of data balancing and optimization techniques in classification, and the provision of insights for manufacturers and consumers regarding public perception of TWS devices.

***Keywords**: True Wireless Stereo (TWS), Sentiment Analysis, Naive Bayes, Support Vector Machine, SMOTE.*

1. PENDAHULUAN

Dalam berbagai aktivitas sehari-hari, mulai dari perjalanan, pekerjaan, hingga hiburan, banyak masyarakat mengandalkan perangkat audio pribadi sebagai sarana utama untuk mengakses informasi dan memenuhi kebutuhan di tengah kegiatan yang sedang berlangsung. Sebelumnya, perangkat audio seperti earphone atau headset umumnya masih menggunakan kabel sebagai media koneksi. Meskipun fungsional, keterbatasan ini seringkali menimbulkan kendala, terutama saat pengguna sedang aktif bergerak atau membutuhkan fleksibilitas tinggi.

Seiring dengan kekurangan yang ada pada perangkat earphone berkabel, perkembangan teknologi telah memberikan solusi untuk mengatasi keterbatasan yang ada perangkat sebelumnya. Teknologi *True Wireless Stereo* (TWS) merupakan inovasi dari teknologi perangkat audio yang mengatasi masalah keterbatasan fungsi. TWS memungkinkan pengguna untuk mendengarkan audio tanpa hambatan kabel, memberikan kebebasan bergerak yang maksimal untuk pengguna

dan meningkatkan produktivitas serta kenyamanan dalam berbagai kegiatan sehari-hari.

True Wireless Stereo (TWS) adalah jenis perangkat audio tanpa kabel yang memanfaatkan teknologi *bluetooth* untuk mentransmisikan suara secara nirkabel. Kehadiran TWS beriringan dengan perkembangan *Bluetooth* 5.0, yang memungkinkan pengiriman sinyal audio dari *smartphone* atau perangkat lain yang mendukung koneksi *bluetooth* [1]. Seiring dengan kemajuan tersebut, penggunaan TWS kini telah menjadi tren di kalangan masyarakat untuk mendampingi berbagai kegiatan sehari-hari. Desain TWS yang ringkas, tanpa kabel, dan kemudahan koneksi perangkat, menjadikan TWS pilihan utama bagi banyak orang untuk mendengarkan musik saat berolahraga, mengikuti meeting online saat bekerja, atau hanya mengalihkan suara bising yang terdengar.

Setiap kemunculan inovasi atau hal baru dalam kehidupan sering kali memicu beragam tanggapan dan pandangan dari berbagai lapisan masyarakat [2]. Melihat semakin luasnya tren penggunaan TWS di masyarakat, persepsi dan pengalaman pengguna terhadap perangkat ini menjadi sangat penting. Oleh karena itu, analisis sentimen juga menjadi sangat penting untuk mengidentifikasi opini, preferensi, dan potensi masalah yang mungkin muncul dari penggunaan TWS. Dengan menganalisis sentimen dari berbagai pengguna tws pada media sosial peneliti dapat memperoleh informasi mengenai kepuasan pengguna, fitur yang paling menarik, serta area yang memerlukan perbaikan. Informasi akan bermanfaat bagi produsen untuk pengembangan produk di masa mendatang dan dapat menjadi pertimbangan konsumen untuk membuat keputusan pembelian yang lebih efisien di tengah pilihan TWS yang semakin beragam di pasaran.

Sentiment analysis merupakan salah satu bidang dalam penelitian *text mining* yang berfokus pada pengklasifikasian dokumen berbasis teks. Melalui teknik pemrosesan bahasa alami (*Natural Language Processing/NLP*), analisis ini mampu mengekstraksi opini, emosi, serta evaluasi tertulis seseorang terhadap suatu topik tertentu [3]. Kajian terhadap analisis sentimen menjadi salah satu topik yang cukup banyak diminati dalam dunia penelitian, karena dinilai mampu memberikan kontribusi positif di berbagai bidang penerapan. [4].

Penelitian terdahulu terkait analisis sentimen telah dilakukan oleh Karimah, Lestari dan Mulyawan pada tahun 2024 dengan judul "Analisis Sentimen Komentar Video Mobil Listrik Di *Platform* Youtube Dengan Metode Naive Bayes". Studi ini mengevaluasi sentimen publik Indonesia terhadap kendaraan listrik dengan menganalisis komentar pada video YouTube, menggunakan algoritma Naive Bayes. Hasil evaluasi model menunjukkan tingkat akurasi sebesar 70,69% setelah dilakukan peningkatan data menggunakan teknik SMOTE *Upsampling* [5].

Penelitian selanjutnya dilakukan pada tahun 2023 oleh Alfaris dan Kusnawi dengan judul "Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee" yang membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes dalam menganalisis sentimen dari ulasan pengguna aplikasi Shopee di Google Playstore. Hasil analisis menunjukkan bahwa Naïve Bayes menghasilkan tingkat akurasi yang lebih tinggi, yaitu 71%, dibandingkan dengan KNN yang mencapai akurasi sebesar 70%. [6].

Penelitian ketiga dilakukan oleh Sepriadi dkk pada Tahun 2023 dengan judul "Analisis Sentimen *Review* Aplikasi MyPertamina menggunakan *Word Embedding Fasttext* dan Algoritma K-Nearest Neighbor". Penelitian ini menganalisis sentimen aplikasi MyPertamina menggunakan pendekatan *Lexicon Based* (kamus Vader) untuk pelabelan dan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi. Akurasi tertinggi sebesar 73% dicapai pada dataset dengan porsi 80/20 dan nilai K=9 tanpa *undersampling [7]*.

Penelitian keempat dilakukan oleh Fadillah Dkk pada tahun 2025 dengan judul "Analisis Sentimen Masyarakat Terhadap Pembatasan BBM Pertalite Menggunakan Random Forest dan K-Nearest Neighbor". Penelitian ini menganalisis opini publik terhadap kebijakan pembatasan penggunaan BBM Pertalite dari komentar Instagram. Dengan algoritma K-Nearest Neighbor (KNN) dan Random Forest, komentar diklasifikasikan menjadi sentimen positif, negatif, atau netral. Metode Random Forest menghasilkan akurasi tertinggi sebesar 73% pada rasio pembagian data 80:20 [8].

Penelitian kelima dilakukan oleh Ningsih dan rekan-rekannya pada tahun 2024 dengan judul "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada

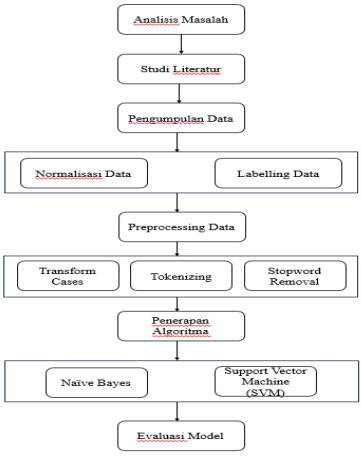
Penggunaan Mobil Listrik di Indonesia." Studi ini mengevaluasi sentimen masyarakat terhadap penggunaan mobil listrik di Indonesia dengan memanfaatkan data Twitter berbahasa Indonesia. Dua algoritma yang digunakan dalam proses klasifikasi adalah Naïve Bayes dan Support Vector Machine (SVM). Data dikumpulkan dari situs Kaggle.com menggunakan kata kunci "kendaraan listrik", kemudian diproses melalui tahap pelabelan manual, praproses data, dan pemodelan. Hasil analisis menunjukkan bahwa algoritma SVM memberikan akurasi lebih tinggi, yakni sebesar 70,82%, sedangkan Naïve Bayes mencatatkan akurasi sebesar 63,02%[9].

Mengacu pada latar belakang serta sejumlah studi sebelumnya yang telah dibahas, penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat Indonesia terhadap penggunaan *True Wireless Stereo* (TWS), dengan menganalisis opini para pengguna yang diungkapkan melalui platform media sosial "X". Dalam penelitian ini, digunakan dua algoritma klasifikasi, yaitu Naive Bayes dan Support Vector Machine (SVM), guna membandingkan tingkat akurasi yang dihasilkan masing-masing metode. Untuk meningkatkan kinerja kedua algoritma tersebut, peneliti menerapkan teknik optimasi *Particle Swarm Optimization* (PSO). Selain itu, metode SMOTE digunakan untuk mengatasi ketidakseimbangan data antara kelas positif dan negatif.

Particle Swarm Optimization (PSO) adalah salah satu teknik optimasi yang dikembangkan berdasarkan pola pergerakan kolektif kawanan burung, dan metode ini dapat digunakan untuk menyelesaikan berbagai permasalahan optimasi di berbagai bidang. Setiap partikel dalam PSO bergerak dalam ruang pencarian dengan menyesuaikan posisinya berdasarkan pengalaman terbaiknya sendiri dan partikel lainnya. Proses ini bertujuan untuk mencapai keseimbangan optimal dalam menemukan solusi terbaik [10]. Sedangkan, SMOTE (*Synthetic Minority Oversampling Technique*) merupakan pendekatan yang dirancang untuk menangani ketidakseimbangan data dan mengurangi risiko *overfitting*. Teknik ini bekerja dengan mengambil sampel secara acak dari kelas minoritas, kemudian mengidentifikasi K tetangga terdekat (*K-nearest neighbors*) dari setiap sampel tersebut. Berdasarkan informasi dari tetangga-tetangga ini, SMOTE menghasilkan data sintetis baru yang merepresentasikan kelas minoritas tambahan. Proses ini tidak hanya menambah jumlah data pada kelas minoritas, tetapi juga meningkatkan keragaman data, sehingga model dapat mempelajari pola dari kedua kelas secara lebih seimbang. Teknik ini digunakan karena telah terbukti efektif dalam meningkatkan performa klasifikasi pada data yang imbalanced tanpa menimbulkan overfitting akibat duplikasi data [11].

2. METODE PENELITIAN

Penelitian ini dirancang dengan tahapan yang terorganisir dan berurutan guna memastikan kelancaran proses serta memperoleh hasil yang optimal. Penelitian ini mengikuti beberapa tahapan mulai dari analisis masalah hingga evaluasi model, dengan harapan tujuan penelitian dapat tercapai dengan tepat. Langkah-langkah yang ditempuh dalam pelaksanaan penelitian ini disusun sebagai berikut:



Gambar 1. Kerangka Penelitian

2.1. Analisis Masalah

Menentukan rumusan masalah merupakan langkah awal yang harus dipastikan [12]. Pada tahap pertama penelitian, perumusan masalah dilakukan untuk menentukan permasalahan Utama yang akan diteliti. X Pada tahap ini peneliti menganalisis bagaimana persepsi masyarakat terhadap penggunaan TWS (*True Wireless Stereo*) dan menganalisis metode klasifikasi apa yang lebih efektif untuk menyelesaikan masalah.

2.2. Studi Literatur

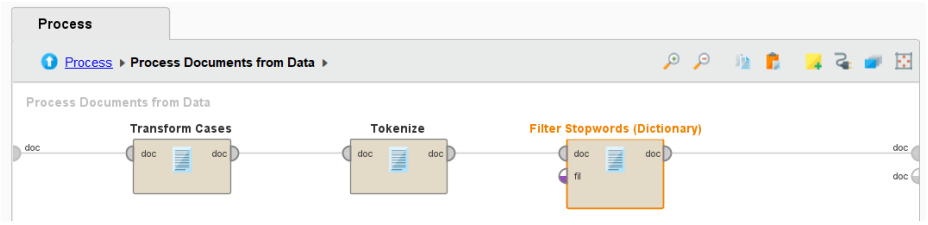
Studi kepustakaan atau yang juga dikenal sebagai *literature review* merupakan bentuk penelitian yang dilakukan dengan mengkaji secara kritis berbagai pengetahuan, pemikiran, dan temuan akademik yang relevan, serta menyusun kontribusi teoritis dan metodologis terhadap topik tertentu. Secara umum, studi kepustakaan dapat diartikan sebagai metode pengumpulan data melalui penelaahan berbagai sumber referensi seperti buku, jurnal, dan artikel ilmiah yang memiliki keterkaitan dengan permasalahan yang dikaji dalam suatu penelitian [13]. Studi literatur dijalani dalam penelitian mengkaji teori dan studi sebelumnya yang terkait dengan analisis sentimen. Peneliti melakukan studi pustaka dengan merujuk pada berbagai jurnal ilmiah sebelumnya guna memperdalam pemahaman mengenai konsep dasar analisis sentimen, tahapan pengolahan teks (*text preprocessing*), serta penggunaan metode klasifikasi seperti Naive Bayes dan Support Vector Machine (SVM). Kegiatan ini bertujuan untuk membangun dasar teoritis yang kokoh dan mengidentifikasi strategi yang telah terbukti efektif dalam studi-studi sejenis.

2.3. Pengumpulan Data

Platform X menjadi salah satu media penyebaran informasi dalam bentuk teks [14]. Pada tahap pengambilan data, peneliti melakukan proses *crawling* dari *platform* media sosial "X" dengan memanfaatkan bahasa pemrograman Python yang dijalankan melalui Google Colab. Google Colab merupakan tempat untuk menulis dan menjalankan program Python, di mana prosesnya dilakukan di server milik Google mempunyai komputer dengan kemampuan tinggi. Dari segi software, Google Colab sudah menyediakan banyak pustaka (*library*) yang sering dipakai sehingga pengguna bisa langsung memakainya tanpa perlu instalasi tambahan [15]. Data yang dikumpukan merupakan cuitan berbahasa Indonesia dari pengguna X yang telah memakai TWS atau persepsi terhadap lingkungannya yang menggunakan TWS. Pada saat *crawling*, *keyword* yang dicari adalah "Pake TWS". *Keyword* “Pake TWS” digunakan oleh peneliti karena mayoritas pengguna media sosial menggunakan bahasa yang santai digunakan sehari-hari pada saat mengunggah cuitan sehingga data yang didapatkan dari crawling berjumlah lebih banyak dibandingkan dengan menggunakan bahasa baku. Rentang waktu cuitan yang diambil adalah dari bulan April sampai dengan Mei 2025. Data yang diperoleh dari hasil *crawling* tersebut berjumlah 521 dengan perbandingan 302 merupakan sentimen positif dan 219 sentimen negatif. Setelah data terkumpul, data tersebut melalui proses normalisasi data yaitu proses perubahan kalimat yang tidak baku menjadi baku secara manual oleh peneliti. Normalisasi ini juga menyatukan ragam kosakata dari berbagai cuitan, sehingga proses analisis menjadi lebih akurat. Sebagai contoh, kata seperti “kamu”, “lo”, atau “lu” akan dikonversi menjadi “Anda” agar penggunaan kata lebih konsisten. Setelah proses normalisasi data selesai dilakukan, setiap data kemudian diklasifikasikan secara manual ke dalam kategori sentimen positif atau negatif.

2.4. *Preprocessing Data*

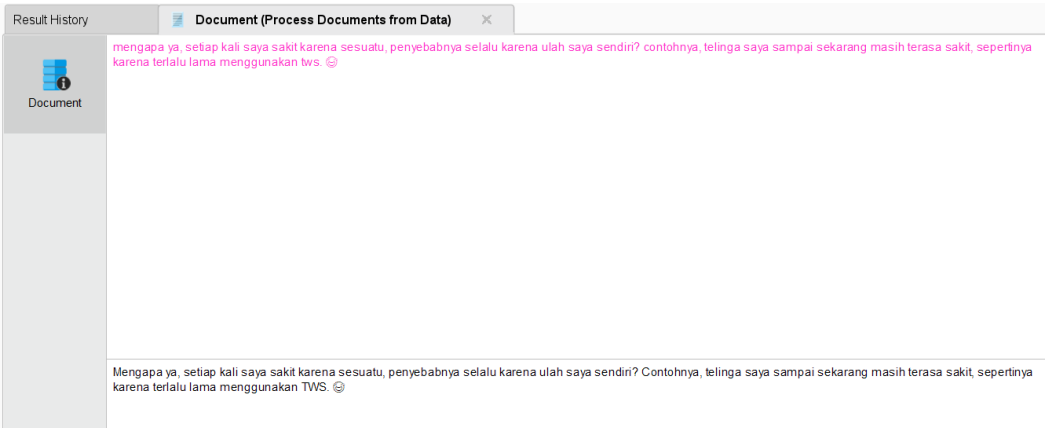
Teknik prapemrosesan data merupakan langkah yang sangat penting dalam proses data mining. Kualitas hasil analisis sangat bergantung pada kualitas data yang digunakan [16]. Data yang telah dinormalisasi dan diberikan label, tahap selanjutnya data tersebut akan melalui tahap *preprocess* pada aplikasi Altair Studio. Altair AI Studio, yang sebelumnya dikenal sebagai RapidMiner, merupakan *platform* perangkat lunak *open-source* yang dirancang untuk keperluan analisis data dan pembelajaran mesin. Platform ini menyediakan beragam operator dan fitur yang mendukung berbagai aktivitas analitik, seperti praproses data, klasifikasi, regresi, klastering, dan lainnya [17].Tahapan yang dilakukan meliputi *Transform Cases*, *Tokenizing* dan *Stopword Removal*. Berikut merupakan rancangan tahap *preprocess* pada aplikasi altair studio:



Gambar 2. Rancangan *Preprocess*

a. *Transform Cases*

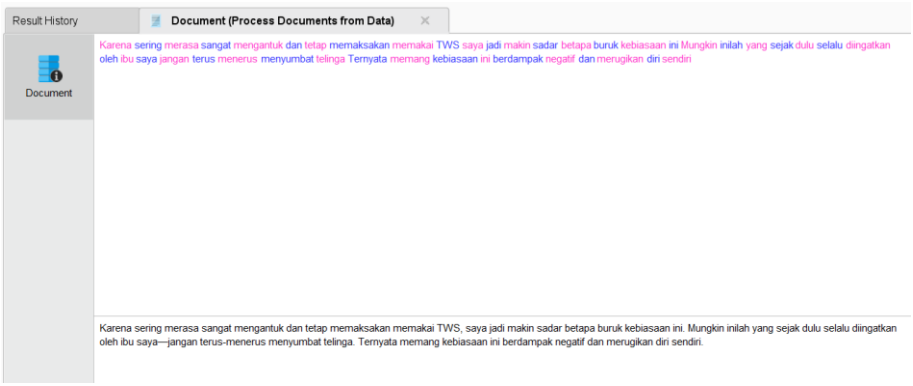
Dalam penelitian ini, proses transformasi huruf dilakukan untuk menyamakan format teks dengan mengonversi seluruh karakter menjadi huruf kecil, sehingga mempermudah tahapan analisis data. Berikut adalah proses *transform case* pada aplikasi altair studio:



Gambar 3. Proses *Transform Cases*

b. *Tokenizing*

Tokenisasi merupakan proses memecah teks menjadi unit-unit kecil seperti kata atau token, yang bertujuan untuk mempermudah identifikasi serta analisis setiap komponen teks secara terpisah dalam tahap praproses data. Berikut merupakan proses *tokenizing* dalam aplikasi Altair Studio:



Gambar 4. Proses *Tokenizing*

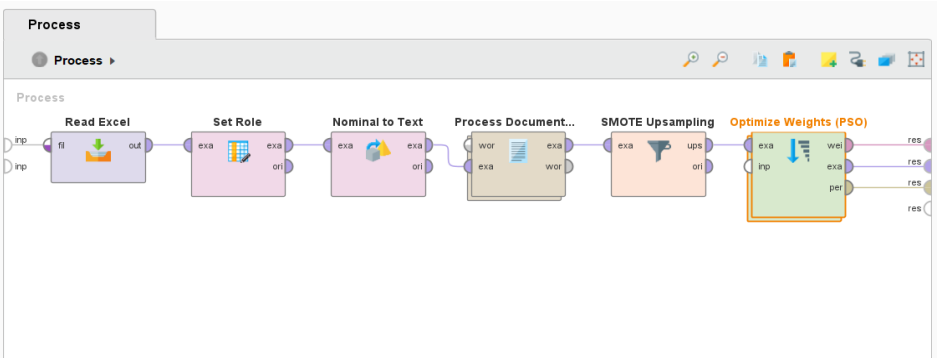
c. *Stopword Removal*

Stopword Removal berguna untuk proses penghapusan kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis, seperti “dan”, “yang”, atau “dengan”. Langkah ini dilakukan untuk meningkatkan fokus terhadap kata-kata yang lebih bermakna dalam proses pemrosesan dan klasifikasi teks. Berikut merupakan proses *stopword removal* pada teks dalam aplikasi altair studio:

Gambar 5. Proses *Stopword Removal*

2.5. Penerapan Algoritma

Setelah melalui tahap *pre-process* data, penerapan algoritma dilakukan. Penelitian ini menerapkan dua metode klasifikasi, yaitu Naive Bayes dan juga SVM (*Support Vector Machine*). Perbandingan antara keduanya diharapkan mampu memberikan pemahaman yang mendalam mengenai keunggulan dan keterbatasan masing-masing algoritma. Penelitian ini menghasilkan keluaran hasil algoritma mana untuk diketahui yang bekerja lebih baik dari segi tingkat akurasinya [18]. Algoritma Naive Bayes sering digunakan dalam pengolahan teks karena mampu bekerja dengan baik pada data yang berbentuk frekuensi kata, seperti seberapa sering kata tertentu muncul dalam sebuah kalimat atau dokumen. Sementara itu, algoritma SVM akan bekerja untuk mencari garis atau batas pemisah terbaik (*hyperplane*) untuk memisahkan data ke dalam beberapa kelas. SVM berusaha membuat jarak antara data dari dua kelas tersebut sejauh mungkin agar hasil klasifikasinya lebih akurat. Pada saat penerapann algoritma Teknik SMOTE digunakan untuk menangani ketidakseimbangan jumlah data antar kelas dengan cara membuat data sintetis pada kelas minoritas, sehingga model dapat belajar secara lebih seimbang dan menghasilkan performa klasifikasi yang lebih akurat. Metode *Particle Swarm Optimization* (PSO) ikut diterapkan dalam penelitian ini untuk menemukan parameter optimal selama proses pelatihan model, dengan tujuan meningkatkan akurasi serta performa algoritma klasifikasi yang digunakan. Berikut adalah *workflow* dari penerapan algoritma pada aplikasi altair studio:



Gambar 6. *Workflow*

Gambar tersebut menunjukkan *workflow* dalam Altair Studio yang menggambarkan rangkaian proses *machine learning* berbasis teks. Proses dimulai dengan membaca data dari file Excel, menetapkan peran atribut (fitur dan label), serta mengubah data nominal menjadi teks. Selanjutnya, dilakukan praproses teks seperti *transform case*, *tokenizing* dan *stopword removal* melalui operator *Process Documents*. Setelah itu, diterapkan SMOTE untuk menyeimbangkan distribusi kelas, dan diakhiri dengan optimasi bobot menggunakan algoritma PSO guna meningkatkan performa model klasifikasi.

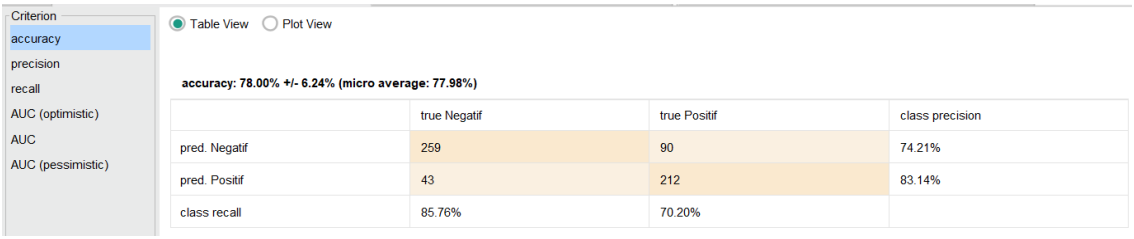
2.6. Evaluasi Model

Setelah proses penerapan kedua algoritma, Langkah selanjutnya yang dilakukan adalah menguji performa kedua metode dengan data. Uji performa kedua metode dalam tahap evaluasi model akan mengasilkan metrik evaluasi. Metrik evaluasi tersebut akan menampilak nilai dari akurasi, *precision, recall* dan grafik AUC dalam tahap ini juga akan ditampilkan. Akurasi akan mengukur persentase prediksi yang benar dari seluruh data, *Precision* mengindikasikan sejauh mana model mampu memberikan prediksi yang tepat terhadap kelas positif, sedangkan *recall* merefleksikan kemampuan model dalam mengidentifikasi seluruh *instance* yang benar-benar termasuk dalam kelas positif. Di sisi lain, nilai AUC yang diperoleh dari kurva ROC digunakan untuk mengevaluasi sejauh mana model dapat membedakan antara kelas positif dan negatif. Semakin dekat nilai AUC dengan angka satu, maka semakin baik performa model.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Output Algoritma Naïve Bayes

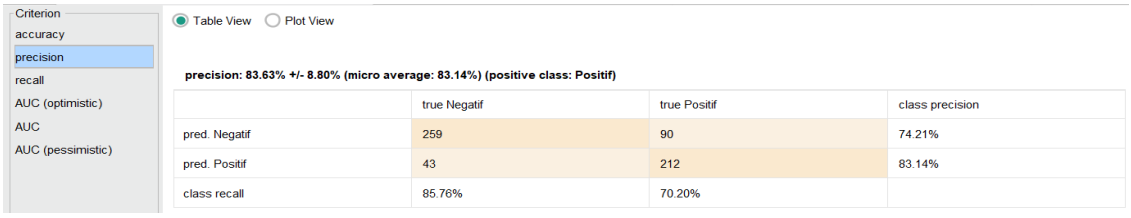
Confusion matrix yang dihasilkan oleh algoritma Naive Bayes mengindikasikan bahwa model memiliki akurasi sebesar 78,00% dengan deviasi ±6,24%, yang berarti sekitar 78% prediksi model sesuai dengan data sebenarnya. Dari 302 data negatif, sebanyak 259 diprediksi dengan benar (*true negative*), sementara 43 salah diklasifikasikan sebagai positif (*false positive*). Sementara itu, dari 302 data positif, sebanyak 212 diklasifikasikan dengan benar (*true positive*), namun 90 sisanya salah diprediksi sebagai negatif (*false negative*).



	true Negatif	true Positif	class precision
pred. Negatif	259	90	74.21%
pred. Positif	43	212	83.14%
class recall	85.76%	70.20%	

Gambar 7. *Confusion Matriks* Akurasi Algoritma Naive Bayes

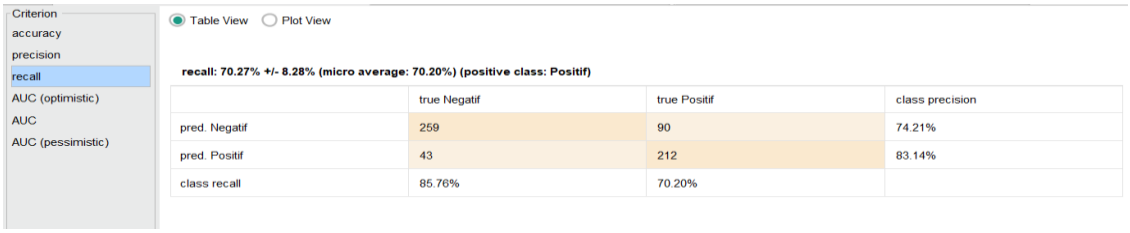
Precision untuk kelas positif mencapai 83,14%, menunjukkan bahwa sebagian besar prediksi positif memang benar, sedangkan *precision* untuk kelas negatif sebesar 74,21%.



	true Negatif	true Positif	class precision
pred. Negatif	259	90	74.21%
pred. Positif	43	212	83.14%
class recall	85.76%	70.20%	

Gambar 8. *Precision Confusion Matriks* Algoritma Naive Bayes

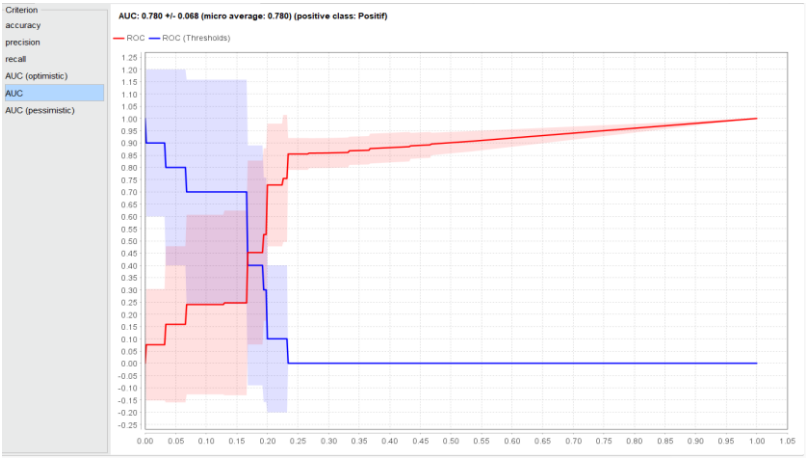
Recall kelas negatif lebih tinggi yaitu 85,76%, dibandingkan *recall* kelas positif yang hanya 70,20%, Temuan ini menunjukkan bahwa model memiliki kemampuan yang lebih tinggi dalam mengidentifikasi kelas negatif dibandingkan kelas positif. Secara umum, performa model dapat dikategorikan cukup baik.



	true Negatif	true Positif	class precision
pred. Negatif	259	90	74.21%
pred. Positif	43	212	83.14%
class recall	85.76%	70.20%	

Gambar 9. *Recall Confusion Matriks* Algoritma Naive Bayes

Nilai AUC dari yang didapatkan dari algoritma Naive Bayes sebesar 0.780 ± 0.068 dan *micro average* 0.780 untuk kelas positif. Nilai AUC yang diperoleh mencerminkan bahwa model mampu membedakan kelas positif dan negatif dengan cukup baik. Sementara itu, kurva ROC menggambarkan keterkaitan antara *true positive rate* dan *false positive rate* pada berbagai nilai *threshold* klasifikasi. Area berwarna di sekitar kurva menunjukkan variasi atau deviasi hasil dari proses validasi silang. Dengan nilai AUC tersebut, penerapan pendekatan Naïve Bayes pada penelitian ini tergolong dalam "*Fair Classification*".



Gambar 10. *AUC Curve* Algoritma Naive Bayes

Algoritma Naive Bayes cukup baik dalam mengenali data negatif, dengan hasil nilai *recall* dan *precision* yang cukup tinggi. Secara keseluruhan, model memiliki akurasi yang cukup baik (78%) dan performa yang cukup seimbang antar kelas meskipun *recall* kelas positif masih lebih rendah.

3.1. Analisis Output Algoritma Support Vector Machine (SVM)

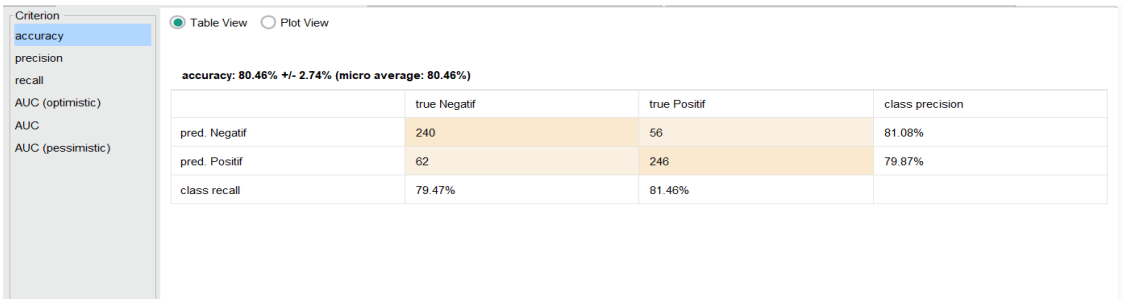
Dalam penelitian ini, algoritma SVM telah diuji dengan berbagai nilai parameter epsilon mulai dari 0,0 hingga 1,0 untuk mengamati pengaruhnya terhadap performa model. Parameter epsilon berperan dalam menentukan margin toleransi kesalahan pada model. Penggunaan nilai epsilon yang tepat dapat mempengaruhi kemampuan model dalam membedakan data antar kelas, menjaga margin yang optimal, serta menghindari *overfitting* atau *underfitting*. Dengan demikian, pemilihan epsilon yang sesuai berkontribusi pada kestabilan dan keseimbangan performa model secara keseluruhan. Berikut adalah tabel hasil akurasi dengan berbagai nilai Epsilon:

Tabel 1. Perbandingan Hasil Akurasi terhadap Berbagai Nilai Epsilon

Epsilon 0,0	78.14%
Epsilon 0,1	76.50%
Epsilon 0,2	76.84%
Epsilon 0,3	77.16%
Epsilon 0,4	77.32%
Epsilon 0,5	75.99%
Epsilon 0,6	76.84%
Epsilon 0,7	76.66%
Epsilon 0,8	77.49%
Epsilon 0,9	80.46%
Epsilon 1,0	49.51%

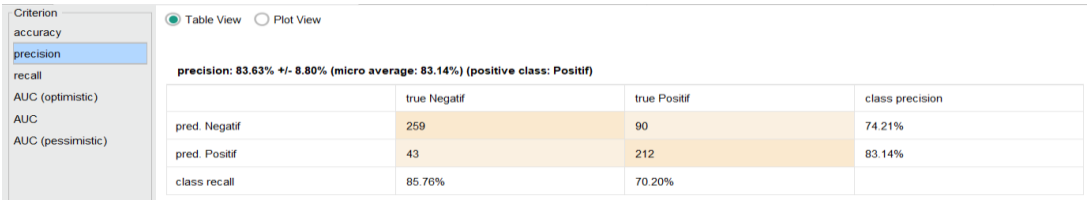
Berdasarkan hasil pengujian terhadap berbagai nilai epsilon, terlihat bahwa performa model SVM bervariasi pada setiap nilai yang digunakan. Nilai akurasi tertinggi dicapai saat

epsilon bernilai 0,9, yaitu sebesar 80,46%, sementara nilai epsilon lainnya menunjukkan akurasi yang lebih rendah, dan bahkan menurun drastis pada epsilon 1,0. Oleh karena itu, dalam penelitian ini digunakan nilai epsilon = 0,9 karena menghasilkan performa terbaik dalam hal akurasi. Pemilihan ini menunjukkan bahwa epsilon 0,9 memberikan keseimbangan yang optimal antara ketelitian model dan toleransi terhadap margin kesalahan. Tabel berikut menyajikan hasil confusion matrix dari model klasifikasi SVM:



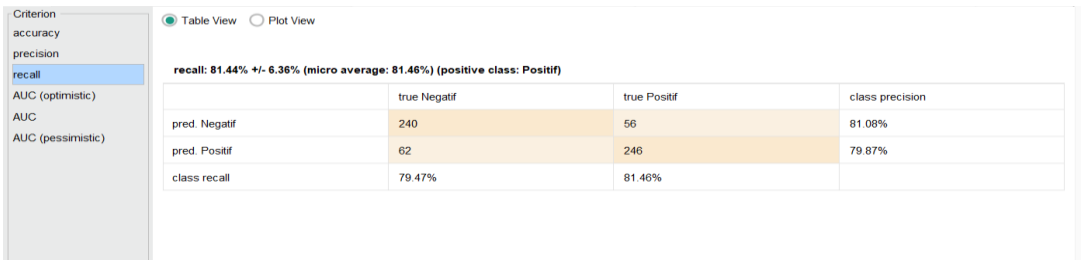
Gambar 11. *Confusion Matriks* Akurasi Algoritma SVM

Hasil *confusion matrix* dari algoritma SVM diatas menunjukkan bahwa model menghasilkan akurasi sebesar 80,46% ± 2,74%, yang berarti bahwa model mampu mengklasifikasikan data dengan benar sekitar 80% dari keseluruhan data, dengan tingkat variasi antar *fold* yang kecil. Pada data negatif, sebanyak 240 data berhasil diklasifikasikan dengan benar (*true negative*), sementara 62 data negatif salah diklasifikasikan sebagai positif (*false positive*). Untuk data positif, 246 data berhasil diprediksi dengan benar (*true positive*), sedangkan 56 data positif salah diklasifikasikan sebagai negatif (*false negative*).



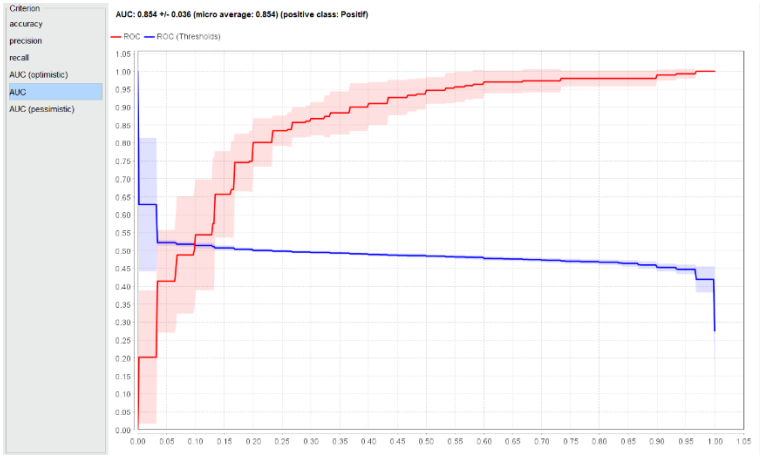
Gambar 12. *Precision Confusion Matriks* Algoritma SVM

Precision untuk kelas negatif sebesar 81,08% mengindikasikan bahwa dari seluruh data yang diprediksi sebagai negatif oleh model, sebanyak 81,08% benar-benar merupakan data negatif. Artinya, model memiliki ketepatan yang cukup tinggi dalam mengidentifikasi data yang termasuk ke dalam kelas negatif, dengan jumlah kesalahan (*false negative*) yang relatif rendah. Sementara itu, precision untuk kelas positif sebesar 79,87% menunjukkan bahwa dari seluruh prediksi yang dikategorikan sebagai positif, sebanyak 79,87% merupakan data positif diklasifikasikan benar. Nilai tersebut menggambarkan model cukup andal dalam mendeteksi data positif, meskipun ada sebagian kecil prediksi positif yang sebenarnya merupakan data negatif (*false positive*). Secara keseluruhan, nilai precision di kedua kelas berada pada tingkat yang cukup tinggi dan seimbang, yang mencerminkan model mampu melakukan klasifikasi dengan tingkat ketepatan yang baik terhadap kedua kelas.



Gambar 13. *Recall Confusion Matriks* Algoritma SVM

Recall pada kelas negatif mencapai 79,47%, yang berarti model mampu mengenali sebagian besar data negatif dengan benar. Meskipun demikian, masih terdapat sejumlah kecil data negatif yang salah diklasifikasikan sebagai positif (*false positive*). Di sisi lain, *recall* untuk kelas positif sebesar 81,46% menunjukkan bahwa model berhasil mendeteksi mayoritas data yang benar-benar termasuk dalam kelas positif. Nilai *recall* yang seimbang pada kedua kelas ini mengindikasikan bahwa model memiliki kinerja yang stabil dalam membedakan antara kelas positif dan negatif.



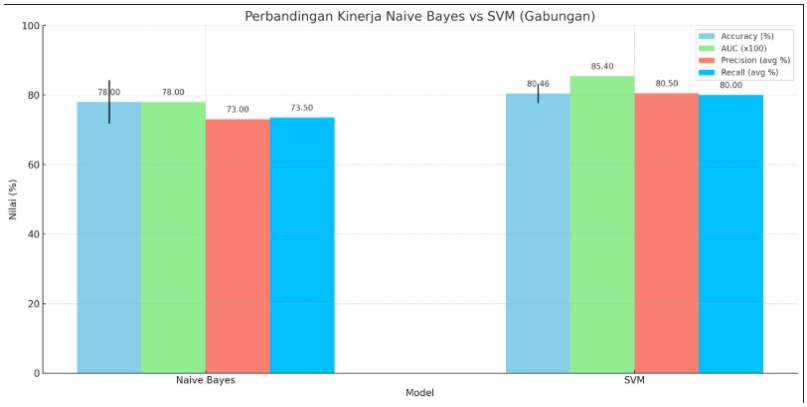
Gambar 14. *AUC Curve* Algoritma SVM

Model Support Vector Machine (SVM) menghasilkan nilai AUC (*Area Under Curve*) sebesar 0,854 ± 0,036, dengan nilai *micro average* yang sama untuk kelas positif. Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas positif dan negatif, sehingga dapat dikategorikan sebagai "*Good Classification*". Kurva ROC yang ditampilkan dalam warna merah cenderung mendekati sudut kiri atas grafik, yang menandakan tingginya *true positive rate* dan rendahnya *false positive rate*. Area yang diarsir pada grafik mencerminkan deviasi hasil dari proses *cross-validation*. Nilai AUC yang tinggi ini mengindikasikan bahwa model SVM mampu melakukan klasifikasi dengan tingkat akurasi dan keandalan yang tinggi terhadap kedua kelas.

Secara keseluruhan, model SVM menunjukkan performa yang lebih seimbang, hal tersebut ditunjukkan oleh nilai *precision* dan *recall* yang relatif seragam pada kelas positif maupun negatif. Model mampu mengidentifikasi data dari kedua kelas dengan akurasi yang tinggi dan deviasi yang rendah, yaitu sebesar 80,46% ± 2,74%, yang mencerminkan kestabilan kinerjanya selama proses validasi. Nilai *precision* dan *recall* yang saling mendekati antara kedua kelas menunjukkan bahwa model tidak cenderung berat sebelah, baik terhadap kelas positif maupun negatif. Jika dibandingkan dengan model sebelumnya, performa SVM dinilai lebih baik karena memberikan hasil klasifikasi yang lebih konsisten, seimbang, dan akurat, sehingga lebih dapat diandalkan dalam menangani permasalahan klasifikasi pada data yang digunakan.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi kedua algoritma yaitu Naive Bayes dan SVM mampu memberikan performa klasifikasi yang cukup baik, namun dengan karakteristik yang berbeda. Berikut merupakan ringkasan hasil dalam grafik:



Gambar 15. Perbandingan Hasil Kedua Algoritma

Model Naive Bayes menghasilkan akurasi sebesar 78,00% ± 6,24%, dengan nilai AUC 0.780, serta menunjukkan keunggulan dalam mengenali kelas negatif. Namun, performanya pada kelas positif masih sedikit lebih rendah. Sebaliknya, model SVM memperlihatkan kinerja yang lebih seimbang dalam mengklasifikasikan kelas positif dan negatif, disertai dengan tingkat akurasi yang lebih unggul sebesar 80,46% ± 2,74% dan nilai AUC 0.854, serta deviasi yang lebih kecil, menandakan kestabilan model. Nilai precision dan recall pada kedua kelas juga lebih seragam, fakta ini memperkuat bahwa SVM bekerja lebih konsisten dalam membedakan kedua kelas. Dengan demikian, dapat disimpulkan bahwa model SVM memberikan performa yang lebih unggul dan seimbang dibandingkan dengan Naive Bayes dalam menangani data pada penelitian ini. Sebagai pengembangan ke depan, penelitian ini memiliki potensi untuk diperluas baik dari segi kompleksitas data maupun variasi pendekatan algoritma. Salah satu arah pengembangan yang dapat dipertimbangkan adalah menguji performa model pada jenis data yang lebih lebih beragam atau berasal dari domain yang berbeda untuk melihat sejauh mana model dapat melakukan generalisasi.

DAFTAR PUSTAKA

[1] I. M. Ozifai, K. Fikri, and H. T. Hinggo, "Pengaruh Kualitas Produk Dan Harga Terhadap Keputusan Pembelian TWS Merek Xiaomi(Studi Kasus Masyarakat di Kota Pekanbaru)," *Jurnal Ilmiah Mahasiswa Merdeka EMBA*, vol. 3 No.2, no. 2, pp. 820–837, 2024.

[2] Ginabila and A. Fauzi, "Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine," *Jurnal Ilmiah ILKOMINFO*, vol. 6 No.2, pp. 111–496222, 2023, doi: https://doi.org/10.47324/ilkominfo.v6i2.180.

[3] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no.1, pp. 32–35, 2023, doi: https://doi.org/10.37905/jjeee.v5i1.16830.

[4] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA*, vol. 10, No.2, pp. 71–76, 2020, doi: http://dx.doi.org/10.32664/smatika.v10i02.455.

[5] A. Karimah, G. Dwilestari, and Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik di Platform Youtube dengan Metode Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8 No.1, pp. 767–773, 2024, doi: https://doi.org/10.36040/jati.v8i1.8373.

[6] S. Alfaris and Kusnawi, "Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee," *Indonesian Journal of Computer Science*, vol. 12, pp. 2023–2766, 2023, doi: https://doi.org/10.33022/ijcs.v12i5.3304.

[7] N. Sepriadi, E. Budianita, M. Fikry, and Pizaini, "Analisis Sentimen Review Aplikasi MyPertamina menggunakan Word Embedding Fasttext dan Algoritma K-Nearest Neighbor," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 15, no. 1, pp. 91–109, 2023, doi: https://doi.org/10.37424/informasi.v15i1.222.

[8] F. M. Fadillah, Y. Cahyana, and A. Fauzi, "Analisis Sentimen Masyarakat Terhadap Pembatasan BBM Pertalite Menggunakan Random Forest dan K-Nearest Neighbor," *Bulletin Of Computer Science Research*, vol. 5, pp. 340–352, 2025, doi: https://doi.org/10.47065/bulletincsr.v5i4.547.

[9] 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, pp. 556–562, 2024, doi: https://doi.org/10.57152/malcom.v4i2.1253.

[10] S. Bumbungan, K. Kusrini, and Kusnawi, "Penerapan Particle Swarm Optimization (PSO) dalam Pemilihan Parameter Secara Otomatis pada Support Vector Machine (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika," *Jurnal Teknik AMATA*, vol. 04, no.1, 2023, doi: https://doi.org/10.55334/jtam.v4i1.77.

[11] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model," *IEEE Access*, vol. 9, pp. 78621–78634, 2021, doi: 10.1109/ACCESS.2021.3083638.

[12] Asian Law Students' Association, *ALSA Legal Writing Guidebook: Advanced Guide to Level Up Legal Writing*. Jakarta, 2024.

[13] A. N. Ramadani, K. C. Kirana, U. Astuti, and A. Marini, "Pengaruh Penggunaan Media Pembelajaran Terhadap Dunia Pendidikan (Studi Literatur)," *JPDSH: Jurnal Pendidikan Dasar Dan Sosial Humaniora*, vol. 2 No.6, no. 6, pp. 749–756, 2023, https://doi.org/10.53625/jpdsh.v2i6.5432.

[14] A. Pratama and M. Rosyda, "Analisis Sentimen dalam Aplikasi X terhadap Pengungsi Rohingya dengan LSTM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 1, pp. 95–105, 2025, doi: https://doi.org/10.36080/skanika.v8i1.3329.

[15] R. Gelar Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 1, pp. 55–60, 2023, doi: https://doi.org/10.47233/jteksis.v5i1.750.

[16] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Oriental Journal of Computer Science and Technology*, vol. 13, pp. 78–81, 2021, doi: http://dx.doi.org/10.13005/ojcst13.0203.03.

[17] L. B. A. Prasetya, "Clustering Dalam Menentukan Tindak Lanjut HasilAnnual Check Mental Health Dengan Algoritma K-Means," *Computer Based Information System Journal*, vol. 13, no. 1, pp. 55-61, 2025, doi: https://doi.org/10.33884/cbis.v13i1.9564.

[18] N. Khoirunnisaa, *et al*., "Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naive Bayes dan SVM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7 No.1, pp. 64–73, 2024, https://doi.org/10.36080/skanika.v7i1.3138.