

Klasifikasi Sentimen Ulasan Pengguna Aplikasi Layanan Publik Google Play Store Menggunakan NLP dan ML

Dwi Shafira Akbar Rizki^{1,*}, Muhammad Syaiful Khabib², Novita Rahmayuna³, Victor Gayuh Utomo⁴

^{1*,2,4} Fakultas Teknologi Informasi dan Komunikasi, Sistem Informasi, Universitas Semarang, Kota Semarang, Indonesia

³ School of Information Systems, Sistem Informasi, Universitas Bina Nusantara, Kota Semarang, Indonesia

Email: ^{1,*}shafiraakbarizkii@gmail.com, ²khabibsyiaiful0@gmail.com, ³novita.rahmayuna@binus.ac.id,

⁴victor@usm.ac.id

^{*}) Email Penulis Utama

Abstrak— Perkembangan teknologi digital telah mendorong transformasi signifikan dalam penyediaan layanan publik di Indonesia. Peningkatan pemanfaatan aplikasi digital menuntut adanya evaluasi berbasis data guna memahami tingkat kepuasan masyarakat sekaligus menilai persepsi mereka terhadap kualitas layanan. Evaluasi ini penting tidak hanya dari sisi teknis, tetapi juga untuk memastikan bahwa manfaat layanan benar-benar dirasakan oleh para pengguna. Salah satu pendekatan yang relevan untuk melakukan penilaian tersebut adalah analisis sentimen dengan memanfaatkan metode Natural Language Processing (NLP) melalui ulasan pengguna yang tersedia pada platform distribusi aplikasi seperti Google Play Store. Penelitian ini bertujuan menerapkan serta membandingkan kinerja tiga algoritma klasifikasi teks, yakni Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, dalam menganalisis sentimen terhadap aplikasi layanan publik MyICON+, yang dikembangkan oleh PT Indonesia Comnets Plus (ICON+), anak perusahaan dari PT PLN (Persero). Metode penelitian menggunakan pendekatan kuantitatif eksperimental dengan desain komparatif. Data penelitian berupa 2.000 ulasan berbahasa Indonesia yang dikumpulkan menggunakan teknik web scraping. Data kemudian diproses melalui tahapan prapemrosesan teks yang meliputi *case folding*, *tokenisasi*, penghapusan *stopword*, dan *stemming* menggunakan pustaka Sastrawi. Representasi fitur diperoleh melalui metode Term Frequency–Inverse Document Frequency (TF-IDF), dengan data yang diberi label sentimen positif dan negatif. Dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi dilakukan dengan metrik akurasi, precision, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi 97% dan macro F1-score 0,87. Algoritma Random Forest berada pada posisi kedua dengan akurasi 95% dan macro F1-score 0,80, sedangkan Naïve Bayes memperoleh akurasi 92% dan macro F1-score 0,51. Selain unggul dalam akurasi, SVM juga lebih stabil dalam mendeteksi sentimen minoritas seperti ulasan positif. Secara keseluruhan, penelitian ini menegaskan bahwa SVM merupakan metode yang paling efektif dalam klasifikasi opini berbasis teks pada konteks layanan publik digital. Hasil studi diharapkan berkontribusi dalam pengembangan sistem analitik opini publik otomatis dan mendukung pengambilan keputusan yang lebih adaptif untuk peningkatan layanan digital pemerintah.

Kata Kunci: Analisis Sentimen, Support Vector Machine, Naïve Bayes, Random Forest, Natural Language Processing

Abstract— The rapid advancement of digital technology has significantly transformed the provision of public services in Indonesia. The growing utilization of digital applications necessitates data-driven evaluations to better understand public satisfaction and perceptions regarding service quality. Such evaluations are not only intended to measure technical performance but also to ensure that the benefits of digital services are genuinely experienced by users. One promising approach to achieve this goal is sentiment analysis using Natural Language Processing (NLP), particularly by analyzing user reviews available on application distribution platforms such as the Google Play Store. This study aims to implement and compare the performance of three text classification algorithms, namely Naïve Bayes, Support Vector Machine (SVM), and Random Forest, in examining public sentiment toward the MyICON+ application. MyICON+ is an official platform developed by PT Indonesia Comnets Plus (ICON+), a subsidiary of PT PLN (Persero), as part of the digital transformation in public service delivery. The research employed a quantitative experimental approach with a comparative design. A total of 2,000 Indonesian-language user reviews were collected automatically using a web scraping technique. The reviews then underwent a preprocessing stage consisting of case folding, tokenization, stopword removal, and stemming using the Sastrawi library. Feature representation was conducted using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and the dataset was labeled with positive and negative sentiments. The data was split into 80% for training and 20% for testing purposes. The models were evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The results indicate that SVM achieved the highest performance with 97% accuracy and a macro F1-score of 0.87. Random Forest ranked second with 95% accuracy and a macro F1-score of 0.80, while Naïve Bayes obtained 92% accuracy and a macro F1-score of 0.51. Furthermore, SVM demonstrated greater stability in detecting minority sentiments, particularly positive reviews. In conclusion, the findings confirm that SVM is the most effective method for text-based opinion classification in the context of digital public services. The study is expected to contribute to the development of automated public opinion analytics systems and support more adaptive decision-making for improving government digital services.

Keywords: Sentiment Analysis, Support Vector Machine, Naïve Bayes, Random Forest, Natural Language Processing

1. PENDAHULUAN

Teknologi informasi yang terus berkembang pesat telah menghadirkan perubahan mendasar pada berbagai sektor kehidupan, terutama dalam cara masyarakat mengakses dan menggunakan layanan publik. Aplikasi layanan publik kini hadir sebagai jembatan antara pemerintah atau badan usaha milik negara dengan masyarakat, memfasilitasi berbagai kebutuhan seperti pelaporan keluhan, pemantauan layanan, pembayaran tagihan, dan akses informasi berbasis *daring*. Namun, keberhasilan digitalisasi layanan publik tidak semata ditentukan oleh aspek teknis dan ketersediaan fitur, melainkan juga oleh bagaimana pengalaman serta kepuasan masyarakat tercermin melalui ulasan yang mereka berikan pada saluran penyebaran aplikasi, misalnya Google Play Store. Ulasan tersebut berperan sebagai indikator signifikan yang mencerminkan mutu layanan, kecepatan respons, serta kenyamanan dalam penggunaan aplikasi, yang pada akhirnya memengaruhi tingkat kepercayaan dan penerimaan publik terhadap layanan digital pemerintah [1], [2]. Dengan demikian, penerapan analisis serta klasifikasi sentimen berdasarkan ulasan masyarakat memiliki peran krusial untuk menggali persepsi publik secara lebih mendalam, menemukan masalah yang kerap terjadi, sekaligus menyediakan dasar bagi peningkatan mutu layanan digital [3], [4].

Dalam era keterbukaan informasi dan partisipasi digital, ulasan ini menjadi aset penting untuk mengukur keefektifan layanan secara *real-time*. Dalam lingkungan digital yang kompetitif dan dinamis, kebutuhan untuk memahami sentimen masyarakat terhadap layanan publik meningkat secara signifikan. Salah satu pendekatan yang terbukti efektif adalah analisis sentimen berbasis *Natural Language Processing* (NLP). Melalui NLP, teks yang tidak terstruktur dapat ditransformasi menjadi data yang siap dianalisis secara sistematis, sehingga memungkinkan pemahaman yang lebih komprehensif mengenai opini dan persepsi masyarakat. Indriyani et al. [5] dalam studi terkait aplikasi TikTok, menemukan bahwa algoritma Support Vector Machine (SVM) menghasilkan akurasi 84%, melampaui Naïve Bayes yang mencatat 79%. Studi serupa oleh Adela et al. [6] yang menunjukkan keunggulan SVM dalam mengklasifikasikan ulasan pada aplikasi Seabank. Sementara itu, Nevrada dan Syaputra [7] menemukan bahwa ada ketidaksesuaian antara rating pengguna dan isi ulasan pada aplikasi Telegram, menegaskan perlunya analisis sentimen berbasis teks agar dapat menggambarkan persepsi yang lebih akurat.

Meski telah banyak dilakukan analisis terhadap aplikasi komersial, jumlah penelitian yang secara khusus membahas aplikasi layanan publik, terutama yang dikembangkan oleh BUMN, masih tergolong terbatas. Padahal, aplikasi semacam ini berperan penting dalam menyediakan layanan dasar masyarakat yang bersifat esensial, seperti listrik, air, transportasi, dan konektivitas internet. Karakteristik pengguna aplikasi layanan publik cenderung lebih heterogen dari segi latar belakang sosial dan teknologi, dan tingkat urgensinya lebih tinggi dibanding aplikasi hiburan. Studi oleh Hardiansyah et al. [8] terhadap aplikasi MyTelkomsel menunjukkan bahwa SVM memiliki performa klasifikasi yang sangat baik. Namun, karena perbedaan konteks dan jenis layanan, hasil tersebut belum dapat digeneralisasikan ke aplikasi layanan publik lainnya, sehingga perlu ada studi lanjutan yang lebih spesifik.

Sebagai alternatif tambahan dari SVM dan Naïve Bayes, Random Forest sering diterapkan dalam analisis klasifikasi teks berkat ketahanannya dalam mengelola data berukuran tinggi, fitur yang rumit, dan distribusi kelas yang tidak proporsional. Algoritma ini bekerja dengan prinsip *ensemble learning*, yakni menyusun sejumlah pohon keputusan kemudian menggabungkannya untuk memperoleh hasil prediksi yang lebih andal dan tepat. Dalam studi oleh Adrian et al. [9], algoritma ini berhasil mencapai akurasi sebesar 74% untuk analisis sentimen aplikasi SIREKAP, bahkan dalam kondisi data yang sangat tidak seimbang (lebih dari 4.000 ulasan negatif) [10]. mengindikasikan bahwa Random Forest dapat mencapai tingkat akurasi sebesar 88,33% dalam konteks aplikasi e-commerce seperti Shopee dan Tokopedia, memperlihatkan potensinya dalam domain aplikasi publik. Penelitian lain oleh Dan et al. [11] dalam konteks sentimen terhadap kebijakan PSBB juga menunjukkan performa yang kompetitif dari Random Forest dibandingkan metode lain, memperkuat alasan mengapa algoritma ini layak diuji dalam domain layanan publik digital.

Dalam konteks ini, aplikasi MyICON+ menjadi objek studi yang relevan. MyICON+ adalah aplikasi resmi yang dikembangkan oleh PT Indonesia Comnets Plus (ICON+), anak perusahaan dari PT PLN (Persero), untuk mendukung layanan konektivitas internet rumah tangga berbasis fiber optic. Aplikasi ini memberikan kemudahan kepada pelanggan dalam memantau status layanan, melaporkan gangguan, serta mengakses layanan pelanggan secara digital. Sebagai bagian dari ekosistem layanan publik digital di Indonesia, aplikasi ini memiliki potensi data ulasan yang kaya dan dapat digunakan untuk menganalisis kualitas layanan secara langsung dari sudut pandang pengguna. Oleh karena itu, penting untuk melakukan kajian mendalam terhadap ulasan aplikasi MyICON+ untuk mengidentifikasi persepsi pelanggan, serta mengevaluasi model klasifikasi sentimen yang terbukti paling efisien pada konteks penelitian ini.

Penelitian ini difokuskan pada evaluasi performa tiga algoritma terkenal Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam melakukan analisis sentimen terhadap ulasan yang diberikan pengguna aplikasi MyICON+. Penelitian ini dilaksanakan melalui pendekatan kuantitatif eksperimen dengan desain perbandingan. Data penelitian berupa 2.000 ulasan berbahasa Indonesia dikumpulkan dari platform Google Play Store menggunakan teknik penarikan data otomatis (*web scraping*). Selanjutnya, data tersebut melalui tahap

pra-pemrosesan, termasuk normalisasi teks (*case folding*), pemisahan kata (*tokenization*), penghapusan kata umum (*stopword removal*), dan stemming, dengan bantuan pustaka Sastrawi. Kemudian, Teks dikonversi ke bentuk numerik melalui metode *Term Frequency Inverse Document Frequency* (TF-IDF), sentimen dibagi menjadi dua kategori utama, yakni positif dan negatif. Selanjutnya, dataset dipisahkan menjadi 80% untuk pelatihan dan 20% untuk pengujian.

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta *confusion matrix* untuk mengamati performa masing-masing algoritma pada setiap kelas sentimen. Hasil analisis menunjukkan bahwa algoritma Support Vector Machine (SVM) mencatatkan performa terbaik dengan akurasi sebesar 97% dan nilai *macro F1-score* sebesar 0,87. Di posisi kedua, Random Forest mencatatkan akurasi sebesar 95% dan *macro F1-score* 0,80. Sementara itu, Naïve Bayes memperoleh akurasi 92% dan *macro F1-score* sebesar 0,51. SVM juga menunjukkan performa yang lebih stabil dalam mendeteksi ulasan dengan sentimen minoritas, khususnya ulasan positif, yang biasanya sulit teridentifikasi dalam dataset yang tidak seimbang. Hasil ini konsisten dengan temuan pada berbagai studi sebelumnya dan semakin mengukuhkan posisi SVM serta Random Forest sebagai algoritma unggulan dalam tugas klasifikasi opini berbasis teks, khususnya pada konteks layanan publik digital yang berbasis ulasan pengguna [5]-[7], [9]- [10].

Secara teoritis, penelitian ini berkontribusi dalam memperkaya literatur NLP dan *machine learning* di Indonesia, khususnya pada aplikasi layanan publik yang belum banyak diteliti. Penelitian ini juga memberikan pendekatan metodologis yang dapat direplikasi pada aplikasi digital lainnya dengan karakteristik sejenis. Dari sisi praktis, temuan ini memberikan panduan penting bagi pengembang dan pengelola aplikasi publik dalam menyusun strategi pengembangan fitur, merespon kritik pengguna, dan meningkatkan kepuasan layanan secara berkelanjutan. Dengan menggunakan algoritma klasifikasi yang terbukti efektif, instansi pemerintah maupun BUMN dapat lebih proaktif dalam memahami kebutuhan masyarakat melalui analisis opini yang sistematis dan berbasis data.

Dengan demikian, penerapan metode NLP dan pembelajaran mesin dalam klasifikasi sentimen pada ulasan aplikasi layanan publik seperti MyICON+ tidak hanya memberikan gambaran tentang efektivitas teknologi yang digunakan, tetapi juga menjadi langkah awal menuju sistem pelayanan publik yang lebih responsif, adaptif, dan berbasis pada suara pengguna. Penelitian ini berperan sebagai referensi strategis bagi akademisi, profesional, serta pengambil keputusan untuk meningkatkan kualitas transformasi digital sektor publik di Indonesia.

2. METODE PENELITIAN

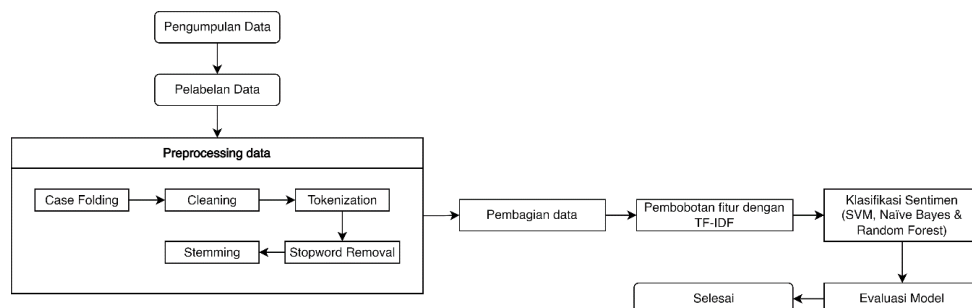
2.1 Jenis Penelitian dan Pendekatan yang Digunakan

Studi ini menerapkan metode kuantitatif komparatif dengan tujuan menilai serta membandingkan performa tiga algoritma klasifikasi populer, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, dalam analisis sentimen terhadap ulasan pengguna aplikasi MyICON+ di Google Play Store. Pemanfaatan Natural Language Processing (NLP) dan teknik machine learning dalam klasifikasi opini berbasis teks dianggap efektif karena memiliki kemampuan mengelola data dalam skala luas secara sistematis dan konsisten. Pendekatan ini juga didukung oleh penelitian sebelumnya [1] dan [7], yang mengimplementasikan metode serupa untuk aplikasi Telegram dan Tiktok dengan menghasilkan akurasi lebih dari 85%.

Kategori penelitian ini termasuk eksperimen kuantitatif dengan pendekatan komputasional dan statistik berbasis teks. Proses penelitian dilakukan secara terstruktur, mencakup akuisisi data, prapemrosesan teks, pelabelan sentimen, pelatihan model, evaluasi performa, dan analisis hasil. Fokus utamanya adalah mengukur efektivitas model dalam menangkap pola sentimen pengguna terhadap aplikasi layanan publik.

2.2 Tahapan Penelitian

Proses penelitian disusun secara sistematis untuk mencakup seluruh langkah yang dilaksanakan, mulai dari pengumpulan data hingga penilaian hasil klasifikasi. Setiap langkah dijelaskan secara detail pada Gambar 1.



Gambar 1. Diagram Tahapan Analisis Sentimen pada Aplikasi MyICON+

Dalam sebuah penelitian analisis sentimen berbasis teks, pendekatan yang sistematis diperlukan agar setiap tahap saling berkaitan dan mendukung keakuratan hasil. Alur tahapan seperti yang ditampilkan pada gambar memuat proses preprocessing, konversi fitur, pelatihan model, hingga evaluasi, yang merupakan praktik umum dalam berbagai studi NLP [12], [13]. Uraian dari setiap tahapan adalah sebagai berikut:

- a. Perolehan Data
Sebanyak 2.000 ulasan dari pengguna aplikasi MyICON+ dikumpulkan melalui Google Play Store secara publik. Proses pengambilan data dilakukan menggunakan teknik web scraping dengan memanfaatkan pustaka Google Play Scraper pada platform Python. Data difilter agar hanya mencakup ulasan berbahasa Indonesia dan memiliki panjang minimum lima kata untuk menjamin keberartian isi. Teknik serupa digunakan dalam studi [6] dan [5] yang menunjukkan efektivitas scraping sebagai metode akuisisi data ulasan pengguna dalam skala besar.
- b. Pelabelan Data
Setiap komentar dari pengguna Google Play Store diberi label sentimen manual dalam dua kelas, positif (1) dan negatif (0). Proses pelabelan mempertimbangkan polaritas opini, konteks bahasa, serta kata kunci yang menunjukkan muatan emosional. Proses anotasi ini dilakukan langsung oleh peneliti menggunakan antarmuka interaktif Google Colab, yang memungkinkan pembacaan dan pelabelan baris per baris dengan pencatatan data secara real-time. Misalnya, ulasan seperti “layanan sangat baik” akan diklasifikasikan sebagai positif, sedangkan ulasan seperti “tidak memuaskan” atau “buruk sekali” akan dilabeli negatif. Ulasan yang ambigu atau bernada netral seperti “biasa saja” dapat dikelompokkan sebagai netral atau diabaikan jika distribusinya sangat kecil, sebagaimana disarankan dalam beberapa penelitian sejenis [14], [15]. Strategi pelabelan seperti ini bertujuan menjaga konsistensi data latih untuk klasifikasi sentimen berbasis supervised learning.
- c. Preprocessing Data
Sebelum model dibangun, data teks melalui proses pre-pemrosesan untuk meningkatkan akurasi analisis. Proses ini merupakan standar dalam NLP untuk bahasa Indonesia dan sangat berpengaruh terhadap kualitas representasi data teks. Langkah-langkah meliputi:
 1. Case Folding:
Pada tahap case folding, seluruh karakter huruf dalam teks dikonversi ke bentuk huruf kecil agar data menjadi konsisten
 2. Cleaning
Proses pembersihan (cleaning) bertujuan menghilangkan elemen tidak penting misalnya angka, karakter khusus, link, serta emoji.
 3. Tokenization
Pada proses tokenization, teks diuraikan menjadi unit kata untuk memudahkan analisis lebih lanjut.
 4. Stopword Removal
Tahap stopword removal dilakukan untuk menyingkirkan kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis, seperti “yang”, “dan”, atau “adalah”.
 5. Stemming
Stemming merupakan tahap yang bertujuan mengubah kata-kata menjadi bentuk dasarnya (root word).
- d. Pembagian Data (Train-Test Split)
Setelah data melalui tahap prapemrosesan dan pelabelan, dataset kemudian dibagi menjadi dua bagian menggunakan teknik train-test split, dengan 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Pendekatan ini umum digunakan dalam evaluasi performa algoritma machine learning agar model diuji pada data yang belum pernah dikenali sebelumnya [16]. Pemisahan tersebut memungkinkan penilaian kemampuan model dalam melakukan generalisasi terhadap data baru secara lebih objektif.
- e. Pembobotan Fitur dengan TF-IDF
Teks kemudian diubah menjadi representasi angka melalui metode TF-IDF, yang menekankan kata-kata yang unik di setiap dokumen. Proses ini diterapkan dengan bantuan TfidfVectorizer pada library scikit-learn. TF-IDF terbukti efektif dalam meningkatkan representasi vektor dalam banyak penelitian klasifikasi opini berbasis teks, termasuk dalam domain aplikasi publik dan media sosial [12], [13].
- f. Analisis Sentimen
Analisis sentimen adalah metode dalam NLP yang bertujuan mengenali serta mengkategorikan ekspresi subjektif dalam teks, biasanya ke dalam kelas positif dan negatif. Penelitian sebelumnya [1] [3] menunjukkan bahwa teknik ini efektif sebagai alat untuk mengevaluasi kepuasan pengguna terhadap fitur dan kinerja suatu aplikasi. Ulasan dengan kata-kata apresiatif diklasifikasikan sebagai sentimen positif, sementara ulasan yang mengandung keluhan atau kritik termasuk ke dalam kategori negatif. Klasifikasi ini memungkinkan perusahaan memahami respons emosional pengguna secara lebih rinci, melampaui sekadar angka rating.
Untuk klasifikasi sentimen, penelitian ini menggunakan tiga algoritma machine learning:

1. Naïve Bayes (MultinomialNB)

Naïve Bayes merupakan metode klasifikasi berbasis probabilitas dan Teorema Bayes, engan asumsi independensi antar atribut. Kelebihannya terletak pada kesederhanaan implementasi, efisiensi komputasi, serta performa yang baik dalam tugas klasifikasi teks, khususnya analisis sentimen. Kelebihan utama Naïve Bayes adalah proses pelatihannya yang cepat, meskipun pada dataset berukuran besar, serta kemampuannya mengelola data teks yang jarang muncul pada representasi berbasis TF-IDF atau bag-of-words [17].

Sebagai metode yang relatif sederhana tetapi efektif untuk data diskrit, Naïve Bayes sering dijadikan acuan awal (baseline) dalam studi klasifikasi opini berbasis teks. Penelitian oleh Nurwanda et al. [1], memanfaatkan algoritma ini untuk menganalisis sentimen pengguna terhadap aplikasi Telegram yang tersedia di Google Play Store. Hasil penelitian tersebut menunjukkan bahwa Naïve Bayes memberikan performa yang baik, terutama setelah data melalui tahap tokenization dan stopword removal. Meski demikian, studi tersebut mencatat bahwa algoritma ini lebih rentan terhadap noise dan ketidakseimbangan kelas dibandingkan dengan metode lainnya

Penelitian serupa oleh Friska Aditia Indriyani et al. [5] pada aplikasi TikTok melaporkan bahwa Naïve Bayes mencapai tingkat akurasi 79%. Meskipun angka ini sedikit lebih rendah dibandingkan performa SVM yang diuji dalam penelitian tersebut, hasil ini menegaskan bahwa Naïve Bayes tetap relevan untuk klasifikasi opini, khususnya pada aplikasi hiburan dengan dataset yang besar dan dinamis.

Lebih lanjut, [3] menerapkan *Naïve Bayes* dalam konteks berbeda, yaitu untuk menganalisis sentimen pengguna platform pembelajaran daring Coursera. Dalam studi tersebut, *Naïve Bayes* digunakan sebagai bagian dari pendekatan ensemble classification yang juga melibatkan algoritma lain seperti *Decision Tree* dan KNN. Hasilnya menunjukkan bahwa *Naïve Bayes* memberikan kontribusi signifikan terhadap stabilitas model gabungan, yang menegaskan nilai strategisnya sebagai komponen klasifikasi multialgoritma.

Secara umum, *Naïve Bayes* sangat cocok diterapkan pada aplikasi klasifikasi teks yang memerlukan efisiensi dan kecepatan, terutama dalam konteks *real-time* atau volume data besar. Meskipun algoritma ini memiliki keterbatasan dalam menangani korelasi antar fitur, efektivitasnya tetap dapat dioptimalkan melalui strategi preprocessing dan pelabelan data yang tepat, sehingga menghasilkan kinerja yang kompeten dan andal dalam berbagai domain aplikasi klasifikasi opini publik [17]. Rumus dasar dalam metode Naïve Bayes, hal ini dapat dirumuskan sebagai berikut:

$$P(H|X) = \frac{P(H|X) \times P(H)}{P(X)} \quad (1)$$

Keterangan :

- X : data tanpa label
- H : prediksi kelas untuk data X
- P(H|X) : probabilitas posterior H diberikan data X
- P(H) : probabilitas apriori H
- P(H|X) : probabilitas data X diberikan H benar (likelihood)
- P(X) : probabilitas marginal data X.

2. Support Vector Machine (SVM dengan kernel linear)

SVM adalah metode klasifikasi margin-based yang mencari hyperplane paling optimal untuk memisahkan kelas data. Algoritma ini cocok untuk data berdimensi tinggi, termasuk teks, karena dapat mengelola fitur kompleks dari representasi TF-IDF atau *bag-of-words*. Keunggulan SVM juga terletak pada fleksibilitasnya dalam menangani klasifikasi linier maupun non-linier melalui pilihan kernel, di mana linear kernel sering digunakan dalam klasifikasi teks karena efisiensinya [12], [16]. SVM terbukti unggul dalam mengelola fitur-fitur yang saling terkait, yang umum dijumpai dalam representasi data teks. Penelitian oleh Nevrada dan Syaputra [7] menerapkan algoritma SVM untuk menganalisis sentimen ulasan aplikasi Telegram yang tersedia di Google Play Store. Studi tersebut melakukan prapemrosesan data melalui tahap normalisasi huruf, penghapusan stopword, dan stemming, serta memanfaatkan TF-IDF untuk merepresentasikan fitur. Hasil evaluasi menunjukkan bahwa SVM mampu mencapai rata-rata F1-score sebesar 81%, mengungguli model baseline lainnya yang diuji. Temuan ini menunjukkan keandalan SVM dalam menangani dataset besar dengan keragaman bahasa dan struktur kalimat.

Studi lain oleh Friska Aditia Indriyani et al. [5] pada aplikasi TikTok juga menunjukkan kinerja SVM yang kompetitif. Dengan menerapkan tahapan prapemrosesan dan teknik representasi fitur yang serupa, SVM berhasil mencapai tingkat akurasi 84%, melebihi performa Naïve Bayes yang dijadikan pembanding. Temuan ini menguatkan kesimpulan bahwa SVM sangat efektif untuk klasifikasi teks pada platform dengan gaya bahasa informal dan dinamis.

Selain itu, penelitian oleh Adela et al. [6] pada ulasan aplikasi Seabank sebuah aplikasi perbankan digital juga menunjukkan keunggulan SVM. Walaupun akurasi yang diperoleh hanya 63%, nilai ini masih lebih baik dibanding Naïve Bayes pada skenario yang sama. Studi ini menyoroti pentingnya penyesuaian parameter kernel dan regulasi untuk meningkatkan performa SVM dalam klasifikasi opini dengan struktur bahasa yang lebih kompleks dan bercampur.

Secara umum, dengan kemampuan menangani fitur-fitur yang saling bergantung, ketidakseimbangan kelas, dan data berdimensi tinggi, Algoritma SVM sering digunakan dalam berbagai penelitian terkait klasifikasi teks, terutama pada tugas analisis sentimen [12], [16]. Efektivitasnya dalam berbagai domain aplikasi, mulai dari hiburan hingga layanan perbankan dan publik, menjadikan SVM sebagai pilihan yang kuat dalam membangun sistem klasifikasi opini yang andal. Rumus dasar untuk metode SVM dapat dituliskan seperti di bawah ini:

$$(w \times xi) + b = 0 \quad (2)$$

Untuk setiap data x_i , yang tergolong dalam kelas -1, persamaannya dapat dituliskan sebagai berikut:

$$(w \times xi) + b \leq -1, yi = -1 \quad (3)$$

Sementara itu, untuk data x_i yang termasuk kelas +1, persamaannya dapat ditulis sebagai berikut:

$$Cw \times xi + b0 \geq 1, yi = -1 \quad (4)$$

3. Random Forest

Random Forest adalah metode ensemble learning yang mengintegrasikan prinsip pohon keputusan (decision tree) dengan teknik bagging (bootstrap aggregating). Algoritma ini menghasilkan banyak pohon keputusan secara acak dengan memilih subset data serta fitur berbeda pada setiap iterasi pelatihan, kemudian menggabungkan prediksi dari tiap pohon menggunakan mekanisme majority voting. Kelebihan Random Forest terletak pada kemampuannya mengurangi variansi model dan meminimalkan risiko overfitting, terutama ketika diterapkan pada dataset besar dan tidak terstruktur, seperti ulasan teks. Selain itu, metode ini juga tahan terhadap noise dan tetap mempertahankan performa meskipun menghadapi ketidakseimbangan kelas[9].

Penelitian oleh Adrian et al.[9] menerapkan algoritma Random Forest untuk menganalisis sentimen pada aplikasi SIREKAP yang digunakan dalam sistem pemilu. Walaupun dataset sangat tidak seimbang, dengan mayoritas ulasan bernada negatif, model ini mampu mencapai akurasi 74%, menegaskan kemampuan Random Forest dalam menangani ketidakseimbangan kelas. Selanjutnya, Tengke et al. [18] menggunakan Random Forest untuk klasifikasi sentimen pada platform e-commerce populer seperti Shopee dan Tokopedia, dengan akurasi mencapai 88,3%, menunjukkan efektivitas metode ini pada ulasan konsumen berbasis teks multi-kategori. Studi oleh Herjanto et al. [10] membandingkan Random Forest dengan SVM dalam analisis opini publik terkait kebijakan PSBB. Hasilnya memperlihatkan Random Forest mencatat akurasi 91%, lebih unggul dibanding SVM dengan memperoleh 87%, menegaskan keandalan algoritma ini dalam menangani opini publik dengan topik kontekstual dan fluktuatif.

Sementara itu, Dan et al. [11] menguji performa Random Forest dalam klasifikasi ulasan penumpang maskapai Indonesia. Hasil menunjukkan bahwa Random Forest mencatat nilai *macro F1-score* tertinggi dibandingkan algoritma lain seperti SVM dan Naïve Bayes, memperkuat reputasinya sebagai model yang stabil untuk pengolahan teks ulasan dengan struktur opini yang panjang dan kompleks. Penelitian oleh Sari dan Sentimen [19] terhadap aplikasi metaverse juga menegaskan keunggulan Random Forest, yang mencatatkan akurasi sebesar 91%, sedikit lebih tinggi dibandingkan SVM yang memperoleh 90%. Dalam domain baru dan emerging seperti metaverse, temuan ini menunjukkan bahwa Random Forest tetap mampu mempertahankan performa tinggi secara konsisten. Beberapa studi tersebut juga menggunakan metode Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi isu ketidakseimbangan kelas dalam data, terutama ketika jumlah ulasan dengan sentimen minoritas sangat terbatas. Dengan menggabungkan SMOTE dan Random Forest, model mampu meningkatkan sensitivitas terhadap kelas minoritas secara signifikan tanpa mengorbankan nilai *macro F1-score* secara keseluruhan. Pendekatan ini terbukti memperkuat ketahanan model dalam menghadapi distribusi kelas yang tidak merata dan meningkatkan akurasi klasifikasi dalam konteks nyata berbasis opini pengguna. [20].

Secara keseluruhan, Random Forest telah terbukti sebagai metode yang kompeten dan adaptif dalam berbagai domain aplikasi teks. Ketangguhannya dalam menghadapi kompleksitas bahasa alami, keanekaragaman ekspresi opini, serta distribusi kelas yang timpang membuat algoritma ini menjadi pilihan tepat untuk analisis sentimen pada aplikasi layanan publik seperti MyICON+.

g. Evaluasi Model

Penilaian kinerja diterapkan pada tiga algoritma klasifikasi, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, dengan memanfaatkan *confusion matrix* serta metrik evaluasi standar seperti accuracy, precision, recall, dan F1-score. Keempat metrik ini digunakan untuk menilai ketepatan, sensitivitas, dan keseimbangan hasil prediksi yang diberikan oleh masing-masing model terhadap dataset

ulasan pengguna yang telah diberi label secara manual. Evaluasi dilakukan secara menyeluruh dan konsisten untuk memastikan keandalan model, terutama pada data teks yang memiliki distribusi kelas tidak seimbang [13], [15].

Adapun rumus-rumus evaluasi tersebut adalah sebagai berikut:

$$akurasi = \frac{\text{jumlah prediksi benar}}{\text{total data}} \tag{5}$$

$$\text{presisi per kelas} = \frac{TP}{TP+FP} \tag{6}$$

$$\text{recall per kelas} = \frac{TP}{TP+FN} \tag{7}$$

$$f1\text{-score per kelas} = \frac{2 \times P \times R}{P + R} \tag{8}$$

Keterangan :

TP (True Positive) : jumlah prediksi positif yang benar

FP (False Positive) : jumlah prediksi positif yang salah

FN (False Negative) : jumlah prediksi negatif yang salah

TN (True Negative) : jumlah prediksi negatif yang benar

Evaluasi dilakukan untuk masing-masing model terhadap data uji guna mengetahui sejauh mana model dapat memberikan prediksi yang tepat

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset penelitian ini diperoleh dari ulasan publik pengguna aplikasi MyICON+ yang tersedia di Google Play Store. Pengumpulan data dilakukan melalui teknik web scraping, dengan memanfaatkan pustaka googleplay-scraper dalam lingkungan Python di Google Colab. Sebanyak 2.000 ulasan berhasil dikumpulkan dengan menyaring berdasarkan kategori “Most Relevant”.

Proses penyaringan ini dilakukan untuk menjamin bahwa data yang diproses tetap relevan dan memiliki konteks makna yang jelas. Teknik ini mengacu pada pendekatan serupa yang digunakan oleh [5], [6] dalam penelitian aplikasi digital berbasis pengguna.

Tabel 1. Data yang Dikumpulkan

no	user	score	at	content
1	Edw*****	1	23/05/2025 10:43	tiap ada gangguan pada jaringan wifi penanganan selalu lambat, kita komplain pun kayak percuma jawaban yang diberikan pasti slalu ambigu nggak pernah ad kepastian
2	Khu*****	5	23/05/2025 09:48	Pengajuan pemasangan cpt dan pelayanan sangat baik
3	Smi*****	5	23/05/2025 08:35	perbaiki lagi sistemnya
4	Da*****	5	23/05/2025 06:41	aplikasinya bagus dan bermanfaat banget
5	Moh*****	1	23/05/2025 06:06	setiap bulan ada saja gangguan ,bahkan perbaikannya lama bisa seharian lebih .. laporan dari aplikasi sudah ada teknisi kelapangan tapi realitanya seharian tidak ada, dan hari ini terjadi lagi, area klayan kab.cirebon
...
2000	Art*****	3	12/05/2023 02:35	Bagus bisa mengecek kapan jatuh tempo wifi sendiri, jeleknya tidak bisa langsung bayar melalui applikasi ini, sangat disayangkan

Hasil dari proses pengambilan data tersebut kemudian disajikan dalam bentuk tabel untuk memberikan gambaran awal mengenai persepsi pengguna. Tabel ini memuat informasi seperti nama pengguna, tanggal ulasan, rating, serta konten ulasan, yang selanjutnya digunakan sebagai acuan untuk proses analisis teks dan klasifikasi sentimen.

3.2 Pelabelan Data

Data ulasan yang terkumpul dari Google Play Store diberi label sentimen secara manual dan semi-otomatis, terbagi menjadi dua kelas utama, yakni positif (1) dan negatif (0). Penentuan label dilakukan secara langsung oleh peneliti pertama Dwi Shafira Akbar Rizki dan kedua Muhammad Syaiful Khabib, yang secara aktif terlibat dalam peninjauan, klasifikasi, serta validasi isi ulasan.

Pelabelan dilakukan dengan mempertimbangkan kata-kata kunci yang merepresentasikan polaritas opini (misalnya "bagus", "cepat", "puas" untuk sentimen positif; dan "lambat", "buruk", "tidak responsif" untuk sentimen negatif) serta konteks semantik dari kalimat secara keseluruhan. Penilaian tidak hanya bergantung pada kemunculan kata-kata tertentu, tetapi juga pada struktur dan intensitas makna yang disampaikan. Untuk menunjang efisiensi dan akurasi dalam anotasi data, digunakan *interface interaktif* berbasis Google Colaboratory (*Google Colab*). Dalam platform ini, penulis mengembangkan skrip berbasis Python yang secara dinamis menampilkan satu per satu ulasan dalam sel notebook, disertai tombol atau input untuk menentukan label yang sesuai. Setiap hasil anotasi secara otomatis tercatat dan disimpan ke dalam berkas CSV atau Google Sheet secara *real time*, sehingga proses pelabelan dapat dilanjutkan atau ditinjau ulang sewaktu-waktu dengan konsistensi yang tinggi.

Proses pelabelan semacam ini juga digunakan dalam berbagai studi sebelumnya, yang menunjukkan efektivitas metode semi otomatis dan keterlibatan langsung peneliti untuk menjamin validitas data pelatihan dalam supervised learning [14], [15]. Tabel 2 merepresentasikan prinsip dasar dalam anotasi manual yang digunakan untuk menyusun dataset pelatihan bagi model analisis sentimen berbasis teks.

Tabel 2. Proses Pemberian Label Sentimen pada Data

no	content	label sentimen
1	tiap ada gangguan pada jaringan wifi penanganan selalu lambat, kita komplain pun kayak percuma jawaban yang diberikan pasti slalu ambigu nggak pernah ad kepastian	0 (negatif)
2	Pengajuan pemasangan cpt dan pelayanan sangat baik	1 (positif)

3.3 Pre Processing Data

Tahap prapemrosesan data teks menjadi langkah penting dalam alur analisis sentimen, terutama ketika sumber data diperoleh dari ulasan pengguna di platform publik seperti Google Play Store. Proses ini bertujuan menormalkan, menyederhanakan, dan memfilter informasi relevan dari data mentah, sehingga bisa digunakan secara optimal dalam pembelajaran mesin. Tahap ini juga sangat memengaruhi akurasi dan efisiensi model klasifikasi yang diterapkan, seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, karena model sangat bergantung pada representasi fitur teks yang tersedia. Tabel 3 berikut menggambarkan secara sistematis hasil dari setiap tahapan pra-pemrosesan terhadap salah satu contoh kalimat ulasan pengguna aplikasi MyICON+. Tahapan-tahapan yang diterapkan meliputi:

Tabel 3. Pre-Processing Data

content	casefold	cleaned	tokenized	no_stopwords	stemmed
tiap ada	tiap ada	tiap ada	['tiap', 'ada',	['gangguan',	ganggu
gangguan pada	gangguan pada	gangguan	'gangguan', 'pada',	'jaringan',	jaring wifi
jaringan wifi	jaringan wifi	pada jaringan	'jaringan', 'wifi',	'wifi',	tangan
penanganan	penanganan	wifi	'penanganan',	'penanganan',	lambat
selalu lambat,	selalu lambat,	penanganan	'selalu', 'lambat',	'lambat',	komplain
kita komplain	kita komplain	selalu lambat	'kita', 'komplain',	'komplain',	kayak slalu
pun kayak	pun kayak	kita komplain	'pun', 'kayak',	'kayak', 'slalu',	ambigu
percuma	percuma	pun kayak	'percuma',	'ambigu',	nggak ad
jawaban yang	jawaban yang	percuma	'jawaban', 'yang',	'nggak', 'ad',	pasti
diberikan pasti	diberikan pasti	jawaban yang	'diberikan', 'pasti',	'kepastian']	
slalu ambigu	slalu ambigu	diberikan pasti	'slalu', 'ambigu',		
nggak pernah ad	nggak pernah ad	slalu ambigu	'nggak', 'pernah',		
kepastian	kepastian	nggak pernah	'ad', 'kepastian']		
		ad kepastian			

3.4 Pembagian Data

Setelah ulasan dikumpulkan, dibersihkan, dan dilabeli, langkah berikutnya adalah membagi dataset menjadi dua subset utama: data pelatihan (training set) dan data pengujian (testing set). Pembagian ini dilakukan menggunakan teknik train-test split dengan rasio 80:20, yang merupakan pendekatan umum dalam validasi model pembelajaran mesin untuk memastikan model tidak hanya mempelajari pola yang sudah ada, tetapi juga dapat diuji pada data baru yang sebelumnya belum pernah ditemui.

Sebanyak 1.600 data dialokasikan untuk proses pelatihan model, sedangkan 400 data lainnya digunakan untuk menguji kemampuan model pada data baru. Distribusi masing-masing label sentimen dalam kedua subset ditampilkan dalam Tabel 4. Hasilnya menunjukkan bahwa model akan dilatih dan diuji pada proporsi data yang relatif seimbang secara struktural, meskipun terjadi ketimpangan kuantitatif antar kelas sentimen.

Tabel 4. Pembagian Data

label	jumlah data (<i>train</i>)	jumlah data (<i>test</i>)
0.0 (negatif)	1392	367
1.0 (positif)	208	33
total	1600	400

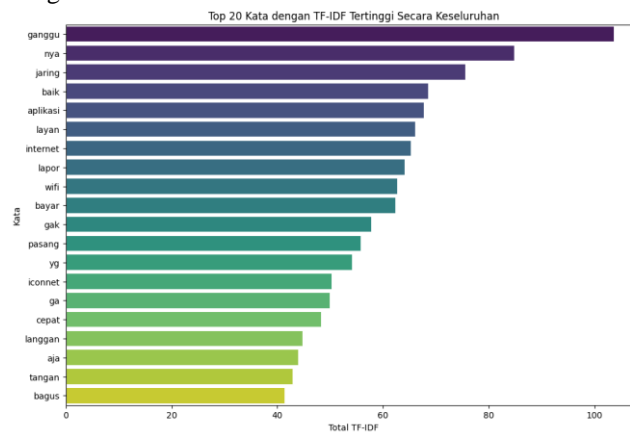
Terlihat bahwa kelas negatif mendominasi dataset, dengan proporsi lebih dari 85% dari total data, sedangkan kelas positif mencakup sisanya. Distribusi ini mencerminkan kondisi nyata dari ulasan pengguna, di mana pengguna cenderung lebih ekspresif dalam menyampaikan keluhan dibandingkan pujian. Ketidakseimbangan ini menjadi tantangan dalam pelatihan model klasifikasi, sebab model dapat lebih condong ke kelas mayoritas dan menurunkan akurasi untuk kelas yang lebih sedikit.

Untuk mengatasi tantangan tersebut, strategi seperti penggunaan metrik *macro-average* dan analisis *confusion matrix* per kelas menjadi penting, sehingga performa model tidak hanya dinilai dari akurasi keseluruhan, tetapi juga dari kemampuan model dalam mengklasifikasikan setiap kelas secara seimbang. Oleh karena itu, dalam penelitian ini, teknik *stratified split* diterapkan saat membagi dataset menjadi set pelatihan dan pengujian. Tujuan pendekatan ini adalah menjaga proporsi label sentimen positif dan negatif tetap seimbang di kedua subset data. Strategi ini sesuai dengan praktik validasi standar dalam pembelajaran mesin dan telah banyak diterapkan dalam penelitian klasifikasi sentimen, khususnya untuk teks berbahasa Indonesia.[16]. Dengan pembagian data yang representatif, kinerja ketiga algoritma klasifikasi, yakni Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, dapat dievaluasi secara akurat dapat dievaluasi secara objektif. Hal ini memungkinkan analisis yang lebih valid terhadap efektivitas model dalam mengatasi tantangan klasifikasi pada domain layanan publik digital yang nyata dan sering kali tidak seimbang secara distribusi data.

3.5 Pembobotan Fitur

Setelah seluruh tahapan prapemrosesan dilakukan, termasuk case folding, cleaning, tokenization, stopword removal, dan stemming, langkah berikutnya adalah mengubah data teks yang bersifat tidak terstruktur menjadi representasi numerik agar dapat digunakan oleh algoritma pembelajaran mesin. Konversi ini dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), yang diimplementasikan melalui *TfidfVectorizer* bawaan pustaka *Scikit-learn*. Metode TF-IDF menghitung bobot setiap kata dalam dokumen berdasarkan dua aspek utama: (1) frekuensi kata muncul dalam dokumen tertentu (Term Frequency) dan (2) seberapa jarang kata muncul di seluruh dokumen korpus (Inverse Document Frequency). Semakin sering suatu kata muncul dalam dokumen dan semakin jarang muncul dalam dokumen lain, maka bobot TF-IDF-nya akan

semakin tinggi. Prinsip ini memungkinkan algoritma mengenali kosakata yang berfungsi diskriminatif, yaitu kosakata dengan peran penting dalam membedakan sentimen antar ulasan.



Gambar 2. Bobot TF-IDF

Seperti terlihat pada Gambar 2, hasil dari proses TF-IDF menghasilkan daftar 20 kata dengan skor tertinggi yang secara signifikan mewakili isi korpus ulasan pengguna. Kata-kata seperti “ganggu”, “nya”, “jaring”, “baik”, dan “aplikasi” muncul sebagai kata paling informatif karena memiliki bobot TF-IDF tinggi. Hal ini menunjukkan bahwa kata-kata tersebut jarang muncul di seluruh dokumen, namun sangat sering muncul dalam dokumen tertentu, sehingga memiliki nilai diskriminatif tinggi. Meskipun representasi awal korpus berupa sparse matrix dengan mayoritas nilai nol karena tidak semua kata muncul dalam tiap ulasan, nilai TF-IDF yang bukan nol inilah yang penting. Bobot tersebut menunjukkan tingkat relevansi sebuah kata pada konteks ulasan spesifik serta dimanfaatkan oleh model klasifikasi seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk mengenali pola sentimen. Dengan demikian, pembobotan TF-IDF berperan penting dalam menekan pengaruh kata-kata umum dan menonjolkan kata yang benar-benar mewakili opini pengguna [12], [13]. Implementasi TF-IDF dalam penelitian ini berhasil membentuk vektor fitur dari 2.000 ulasan pengguna aplikasi MyICON+, yang kemudian dipakai dalam pelatihan dan pengujian model klasifikasi sentimen. TF-IDF terbukti sebagai salah satu teknik vektorisasi teks paling efektif untuk representasi data berdimensi tinggi, terutama pada bahasa alami seperti Bahasa Indonesia yang memiliki struktur morfologi kompleks.

3.6 Klasifikasi Sentimen & Evaluasi

Setelah seluruh tahapan prapemrosesan dan pembobotan fitur dengan metode TF-IDF selesai, langkah selanjutnya adalah melakukan pelatihan model klasifikasi serta evaluasi kinerjanya. Penelitian ini menggunakan tiga algoritma machine learning yang umum diterapkan dalam tugas klasifikasi teks, yaitu:

- a. Naïve Bayes (MultinomialNB)
Model ini diimplementasikan menggunakan pustaka *scikit-learn* dengan konfigurasi default dari MultinomialNB. Model ini dipilih karena telah umum diterapkan dan efektif pada data teks diskrit dalam bentuk representasi TF-IDF. Keunggulannya terletak pada kecepatan dan skalabilitas terhadap dataset besar [9], [14].
- b. Support Vector Machine (SVM – Linear Kernel)
SVM diterapkan memakai kernel linier (LinearSVC), tepat untuk data berdimensi tinggi. Parameter default digunakan, kecuali *random state* disetel untuk menjaga reproduisibilitas. SVM mampu menangani fitur saling bergantung dalam representasi TF-IDF serta efektif dalam mengoptimalkan margin pemisah antar kelas [12], [16].
- c. Random Forest
Random Forest diimplementasikan dengan 100 *estimators* dan pengaturan random state untuk menjaga reproduisibilitas. Model ini dikenal tangguh terhadap overfitting dan mampu menangani ketidakseimbangan data dengan lebih stabil [9]–[20].

Model menjalani proses pelatihan dengan dataset training dan dievaluasi menggunakan dataset testing yang telah disiapkan sebelumnya.

Evaluasi kinerja dilakukan dengan menggunakan metrik berikut:

- a. Akurasi: Persentase prediksi yang benar dari total data pengujian.
- b. Precision, Recall, F1-Score: Mengukur performa tiap kelas.
- c. Confusion Matrix: Distribusi prediksi untuk tiap kelas.

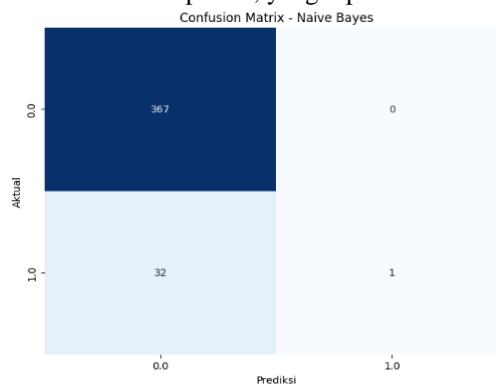
Hasil evaluasi metrik untuk kedua algoritma dirangkum pada Tabel 6, sementara visualisasi confusion matrix tiap model ditampilkan pada Gambar 3 (Naïve Bayes), Gambar 4 (SVM), dan Gambar 5 (Random Forest).

Untuk mendapatkan pemahaman komprehensif atas performa tiap algoritma, dihitung juga nilai macro average, yaitu rata-rata nilai *precision*, *recall*, dan *F1-score* tanpa memperhatikan proporsi distribusi data antar kelas. Ini penting dalam konteks data tidak seimbang seperti ulasan pengguna layanan publik, di mana sentimen negatif mendominasi.

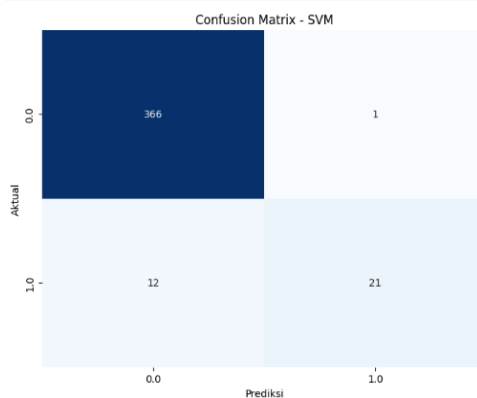
Tabel 6. Evaluasi

Metrik	Naïve Bayes	SVM	Random Forest
Akurasi	0.92	0.97	0.95
Precision (Macro)	0.96	0.96	0.98
Recall (Macro)	0.52	0.82	0.73
F1-Score (Macro)	0.51	0.87	0.80
F1-Score (Weighted)	0.88	0.96	0.95

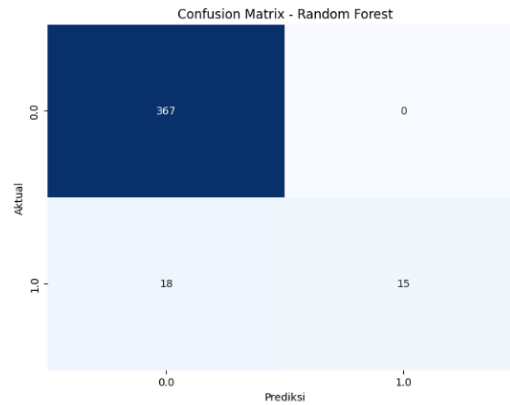
Hasil evaluasi menunjukkan bahwa model SVM secara jelas lebih unggul dibandingkan Naïve Bayes dan Random Forest dalam klasifikasi sentimen. SVM memperoleh akurasi senilai 97% serta macro F1-score dengan nilai 0,87, lebih tinggi daripada Random Forest dengan macro F1-score 0,80 dan Naïve Bayes dengan macro F1-score hanya 0,51. Hal ini menandakan bahwa SVM mampu menangani variabilitas dan sparsitas data teks dengan lebih efektif, serta tetap stabil menghadapi ketidakseimbangan kelas. Rendahnya macro F1-score pada Naïve Bayes terutama disebabkan ketidakseimbangan antara precision dan recall; meskipun precision cukup tinggi, recall yang rendah secara signifikan menurunkan nilai F1. Hal ini terjadi karena F1-score dihitung sebagai rata-rata harmonik dari precision dan recall, sehingga jika salah satunya sangat rendah, maka nilai F1 akan turun drastis. Recall yang rendah, khususnya pada kelas positif, mencerminkan banyaknya *false positive* (FP) atau kesalahan model dalam mengklasifikasikan data positif, yang dipicu oleh dominasi data negatif dalam dataset.



Gambar 3. Confusion Matrix hasil prediksi model Naïve Bayes



Gambar 4. Confusion Matrix hasil prediksi model Support Vector Machine



Gambar 5. Confusion Matrix hasil prediksi model Random Forest

Temuan ini konsisten dengan penelitian oleh Adrian et al. [9] dan Herjanto et al. [10], yang menyatakan bahwa Random Forest dapat bersaing atau bahkan mengungguli SVM dalam beberapa domain klasifikasi opini. Namun demikian, SVM tetap menjadi algoritma yang paling konsisten dalam mengenali pola sentimen dari teks, khususnya dalam pengolahan bahasa Indonesia yang tidak seimbang [5], [7], [19].

Keseluruhan tahapan penelitian, dari pengumpulan data, prapemrosesan, pelabelan, konversi fitur, hingga pemilihan dan evaluasi model, menunjukkan bahwa pemanfaatan NLP dikombinasikan dengan klasifikasi machine learning menghasilkan sistem analisis opini publik yang efektif dan dapat diandalkan. Oleh karena itu, hipotesis alternatif diterima bahwa SVM memiliki performa klasifikasi yang paling optimal, sementara Random Forest merupakan alternatif yang kompetitif, dan Naïve Bayes kurang efektif dalam menangani data tidak seimbang dalam konteks ulasan layanan publik digital.

4. KESIMPULAN

Penelitian ini dilakukan untuk menjawab permasalahan utama dalam pengembangan sistem klasifikasi opini publik berbasis teks, khususnya dalam konteks layanan publik digital di Indonesia. Studi ini memfokuskan pada evaluasi dan komparasi performa tiga algoritma klasifikasi teks, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam menganalisis sentimen pengguna terhadap aplikasi MyICON+ yang dikembangkan oleh PT Indonesia Comnets Plus (ICON+), anak perusahaan PT PLN (Persero). Melalui pendekatan kuantitatif eksperimental dan desain komparatif, penelitian ini mengolah sebanyak 2.000 ulasan berbahasa Indonesia dari Google Play Store, dimulai dari tahap prapemrosesan, transformasi fitur menggunakan TF-IDF, pelabelan sentimen, hingga evaluasi performa model.

Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) mencatat performa tertinggi dalam klasifikasi sentimen ulasan aplikasi layanan publik, dengan akurasi 97% dan macro F1-score sebesar 0,87. Random Forest berada di posisi kedua dengan akurasi 95% dan macro F1-score 0,80, sedangkan Naïve Bayes memperoleh akurasi 92% dan macro F1-score 0,51. Rendahnya performa Naïve Bayes terutama dipengaruhi oleh ketidakseimbangan data, di mana recall rendah pada kelas minoritas menyebabkan macro F1-score menurun secara signifikan. Untuk memastikan evaluasi merata di semua kelas, performa model diukur menggunakan akurasi, precision, recall, serta macro F1-score. Secara teoretis, penelitian ini menambah literatur NLP berbahasa Indonesia dalam konteks layanan publik digital, sedangkan secara praktis, pendekatan sistematis dalam prapemrosesan serta pemilihan algoritma yang tepat terbukti penting untuk membangun sistem analisis opini publik berbasis data yang efektif dan responsif.

Dari sisi praktis, temuan ini sangat bermanfaat bagi pengembang aplikasi layanan publik dan pemangku kebijakan digital dalam merancang sistem evaluasi layanan berbasis data. Implementasi model SVM dan Random Forest dapat digunakan sebagai bagian dari sistem pemantauan otomatis terhadap opini pengguna, yang memungkinkan deteksi cepat terhadap keluhan atau kepuasan pengguna. Hal ini dapat mendorong perbaikan layanan secara berkelanjutan, meningkatkan responsivitas pemerintah atau BUMN, serta memperkuat kepercayaan publik terhadap transformasi digital sektor layanan publik. Dari sisi kebijakan, hasil ini membuka peluang integrasi opinion mining sebagai indikator tambahan dalam proses audit layanan digital. Dengan menerapkan pendekatan data-driven, lembaga penyedia layanan publik dapat mengidentifikasi kebutuhan masyarakat secara lebih akurat, sehingga kebijakan yang diambil menjadi lebih adaptif dan berbasis pada bukti empirik.

Sebagai tindak lanjut, penelitian ini merekomendasikan eksplorasi terhadap model berbasis deep learning seperti IndoBERT atau fine-tuned transformer lainnya, yang berpotensi menghasilkan akurasi dan pemahaman semantik yang lebih dalam terhadap struktur bahasa Indonesia. Selain itu, penerapan teknik penyeimbangan data

seperti SMOTE atau ADASYN juga layak dipertimbangkan untuk mengatasi distribusi label yang tidak seimbang. Cakupan studi juga dapat diperluas dengan melibatkan lebih dari satu aplikasi layanan publik agar temuan dapat digeneralisasi ke konteks layanan digital lainnya, seperti transportasi, e-government, atau utilitas energi. Dengan cara ini, penelitian ini tidak hanya menyumbang secara metodologis dan empiris, tetapi juga membuka arah baru untuk pengembangan riset dan inovasi dalam klasifikasi sentimen berbasis teks di sektor publik Indonesia.

REFERENCES

- [1] N. Nurwanda, N. Suarna, and W. Prihartono, "Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1841–1846, 2024, doi: 10.36040/jati.v8i2.8469.
- [2] R. Damanhuri and V. A. Husein, "Analisis Sentimen pada Ulasan Aplikasi Access by KAI Berbahasa Indonesia Menggunakan Word-Embedding dan Classical Machine Learning," vol. 15, no. November, 2024, doi: 10.14710/jmasif.15.2.62383.
- [3] J. P. Hidayat and I. Nurhaida, "Analisis Sentimen pada Ulasan LMS Pembelajaran Menggunakan Metode Natural Language Processing," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 565–575, Jan. 2025, doi: 10.29100/jupi.v10i1.5899.
- [4] J. Friadi and D. Ely, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam : Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine," vol. 04, 2024, doi: 10.21456/vol14iss4pp403-407.
- [5] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [6] C. N. Adela, S. Karnila, S. Sutedi, and M. Agarina, "Analisis Ulasan Pengguna Aplikasi Seabank Dengan Support Vector Machine Dan Naïve Bayes," *J. Tekno Kompak*, vol. 18, no. 2, p. 441, 2024, doi: 10.33365/jtk.v18i2.4156.
- [7] N. A. Nevrada and M. A. Syaputra, "Sentiment Analysis of Telegram App Reviews on Google Play Store Using the Support Vector Machine (SVM) Algorithm," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 96–105, Jan. 2025, doi: 10.30871/jaic.v9i1.8851.
- [8] D. Hardiansyah, R. A. Aziz, and M. S. Hasibuan, "The Classification Method is Used for Sentiment Analysis in My Telkomsel," *Int. J. Artif. Intell. Res.*, vol. 8, no. 2, p. 169, Dec. 2024, doi: 10.29099/ijair.v8i2.1229.
- [9] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.
- [10] M. F. Y. Herjanto and C. Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [11] L. Dan and B. Menggunakan, "Analisis sentimen aplikasi shopee, tokopedia, lazada dan blibli menggunakan leksikon dan random forest," vol. 12, no. 3, 2024.
- [12] A. S. Iedwan, N. Mauliza, Y. Prityanto, A. D. Hartanto, and A. N. Rohman, "Comparative Performance of SVM and Multinomial Naïve Bayes in Sentiment Analysis of the Film ' Dirty Vote ',," vol. 11, no. 3, pp. 839–848, 2024, doi: 10.15294/sji.v11i3.10290.
- [13] D. Prastyo, I. H. Mursyidin, and D. Irawan, "Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten," vol. 7, no. 2, 2024, doi: 10.32877/bt.v7i2.1833.
- [14] D. Shalikhah, "Improved Accuracy of Naïve Bayes Algorithm and Support Vector Machine Using Particle Swarm Optimization for Menstrual Cup Sentiment Analysis on Twitter," *J. Adv. Inf. Syst. Technol.*, vol. 4, no. 2, pp. 139–148, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/jaist>
- [15] K. Pengaduan, P. Fakultas, and T. Universitas, "Arus Jurnal Sains dan Teknologi (AJST) Klasifikasi Pengaduan Pelayanan Fakultas Teknik Universitas Muhammadiyah Makassar menggunakan Natural Language Processing," vol. 2, no. 2, 2024.
- [16] N. A. Laia and S. P. Barus, "E ISSN : 2809-4069 Analisis Sentimen YouTube : " Di Balik Ambisi Jokowi dalam IKN ",," vol. 5, no. 1, pp. 7–12, 2025.
- [17] O. N. Julianti, N. Suarna, and W. Prihartono, "Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online Di Media Sosial Twitter," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2936–2941, 2024, doi: 10.36040/jati.v8i3.9613.
- [18] R. Tangke *et al.*, "ANALISIS SENTIMEN APLIKASI TIKTOK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST," vol. XIII, no. 2, pp. 53–62, 2024.
- [19] A. Sentimen *et al.*, "The Indonesian Journal of Computer Science," vol. 13, no. 5, pp. 8437–8445, 2024.

- [20] R. R. S. Putri Kumala Sari, "Vol 7 No 1 , Februari 2024 KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM," vol. 7, no. 1, pp. 31–39, 2024.