

Model Rekomendasi Destinasi Wisata Kreatif di Indonesia Berdasarkan Data Cuaca dan Ulasan Wisatawan

Kharisma¹, Irmma Dwijayanti², Ulfi Saidata Aesy³, Alfirna Rizqi Lahitani⁴

^{1,2,3}Sistem Informasi, Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta

⁴Teknologi Informasi, Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta

¹kharisma.ano@gmail.com, ²irmmade9@gmail.com, ³ulfaesy@gmail.com, ⁴alfirnalahitani@gmail.com

Abstract

Indonesia holds vast potential for creative tourism through its rich cultural heritage, natural beauty, and local creativity. However, travelers still face challenges in planning optimal trips due to the lack of context-aware and real-time recommendation systems. In practice, tourists often rely on Google Maps reviews, which are unorganized thematically, and there is limited integration with weather conditions—an important factor that significantly impacts travel experiences, particularly for nature-based destinations. This study aims to develop a recommendation model for creative tourism destinations in Indonesia by integrating two key aspects: sentiment analysis of Google Maps reviews and real-time weather data. The research utilizes tourist reviews from Google Maps alongside up-to-date weather information from destinations across Indonesia. The reviews are analyzed using the Support Vector Machine (SVM) algorithm to classify sentiments as positive or negative. These sentiment results are then combined with real-time weather data to build a Content-Based Filtering (CBF) recommendation system capable of providing more relevant and adaptive suggestions. The study successfully produced a recommendation system model with a testing accuracy of 90%.

Keywords: recommendation model, support vector machine, content-based filtering, tourism, weather

Abstrak

Indonesia memiliki potensi pariwisata kreatif yang besar melalui kekayaan budaya, alam, dan kreativitas lokal. Namun, wisatawan masih menghadapi kendala dalam merencanakan perjalanan yang optimal karena keterbatasan sistem rekomendasi yang bersifat kontekstual dan real-time. Umumnya, wisatawan mencari referensi melalui ulasan Google Maps yang belum terorganisasi secara tematik, serta minimnya integrasi dengan kondisi cuaca yang sangat berpengaruh terhadap pengalaman wisata, terutama untuk destinasi alam. Penelitian ini bertujuan membangun model rekomendasi destinasi wisata kreatif di Indonesia dengan mengintegrasikan dua aspek penting, yaitu analisis sentimen ulasan wisatawan dari Google Maps dan data cuaca real-time. Penelitian ini akan menggunakan data ulasan wisatawan di Google Maps dan data cuaca terkini dari destinasi wisata di Indonesia. Ulasan dianalisis menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan sentimen positif atau negatif. Hasil sentimen tersebut dikombinasikan dengan data cuaca real-time untuk membentuk sistem rekomendasi berbasis Content-Based Filtering (CBF) yang mampu menyarankan destinasi secara lebih relevan dan adaptif. Penelitian ini menghasilkan model sistem rekomendasi dengan tingkat akurasi pengujian data sebesar 90%.

Kata kunci: model rekomendasi, support vector machine, content-based filtering, wisata, cuaca

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

1. Pendahuluan

Indonesia merupakan negara yang kaya akan keragaman budaya, bentang alam, dan kreativitas lokal, menjadikannya arena pariwisata dengan potensi besar. Dalam lima tahun terakhir, tercatat peningkatan signifikan minat wisatawan terhadap berbagai jenis wisata [1]. Destinasi seperti wisata alam, kuliner, sejarah, edukasi, dan agrikultur terus berkembang pesat [2]. Di antara kategori ini, wisata kreatif muncul sebagai fokus perhatian karena menawarkan pengalaman berbasis partisipasi, inovasi, dan interaksi mendalam dengan komunitas lokal. Secara akademik, wisata kreatif didefinisikan sebagai keterlibatan wisatawan dalam aktivitas berbasis kreativitas—seperti seni, kerajinan, atau kuliner kontemporer—yang berbeda dengan wisata berbasis lanskap atau warisan. Relevansi wisata kreatif dalam studi ini terletak pada sifatnya yang dinamis dan sangat sensitif terhadap

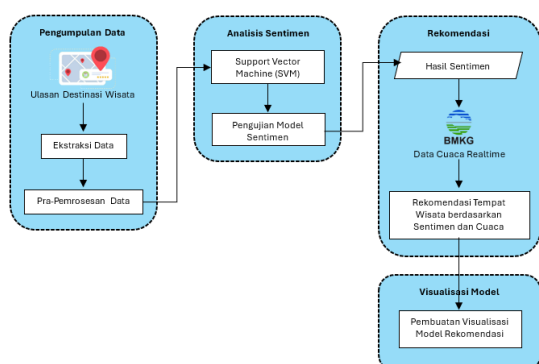
kualitas pengalaman wisatawan serta kondisi lingkungan di lokasi kunjungan. Variasi geografis Indonesia menjadikan cuaca sebagai faktor krusial, terutama karena destinasi kreatif sering kali merupakan ruang terbuka (pantai, pegunungan, desa wisata) yang rentan terhadap perubahan cuaca harian, yang pada akhirnya memengaruhi tingkat kepuasan dan aktivitas wisatawan.

Ulasan wisatawan di platform digital, khususnya Google Maps, merupakan sumber data yang kaya namun belum terintegrasi secara sistematis untuk mendukung pengambilan keputusan. Penelitian terdahulu telah memanfaatkan ulasan ini untuk analisis sentimen menggunakan berbagai metode, termasuk Naïve Bayes [3], [4], [5], [6], K-Nearest Neighbors (KNN) [7], [8], dan Support Vector Machine (SVM) [4], [9]. SVM sering dipilih karena menunjukkan kinerja superior, dengan akurasi klasifikasi teks

melebihi 80%, menjadikannya ideal untuk data berdimensi tinggi. Sementara itu, sistem rekomendasi wisata telah dieksplorasi melalui collaborative filtering [10], [11] dan case-based filtering [12], [13], [14], namun content-based filtering (CBF) [12], [13], [14], [15] dianggap paling relevan di sini karena kemampuannya merekomendasikan destinasi berdasarkan kecocokan atribut konten dengan preferensi pengguna [16], [17]. Meskipun demikian, analisis terhadap studi-studi ini menunjukkan kesenjangan penelitian (research gap) yang jelas.

Permasalahan inti yang belum terpecahkan adalah ketiadaan model rekomendasi destinasi wisata kreatif yang mampu mengintegrasikan sentimen wisatawan dan kondisi cuaca secara simultan. Pemanfaatan data ulasan Google Maps masih terbatas pada analisis sentimen dasar dan belum dioptimalkan untuk mendukung sistem rekomendasi yang adaptif, sehingga belum ada pendekatan yang efektif dalam memberikan rekomendasi secara kontekstual dan real-time sesuai dinamika lingkungan di destinasi. Berdasarkan kesenjangan ini, penelitian ini bertujuan utama untuk mengembangkan model rekomendasi destinasi wisata kreatif di Indonesia. Model yang diusulkan akan memanfaatkan sentimen positif dari ulasan yang diklasifikasikan menggunakan algoritma SVM, dan menghasilkan rekomendasi berdasarkan kecocokan preferensi melalui metode CBF. Melalui integrasi dengan data cuaca real-time, sistem ini diharapkan dapat memberikan kontribusi baru dengan menghasilkan rekomendasi yang lebih kontekstual, adaptif, dan relevan secara waktu nyata, mengatasi keterbatasan solusi yang ada.

2. Metode Penelitian



Gambar 1. Flowchart Tahapan Penelitian

2.1. Pengumpulan Data

Penelitian ini mengandalkan data ulasan wisatawan yang dikumpulkan dari Google Maps, mencakup 170 destinasi wisata yang tersebar di 17 provinsi di Indonesia. Penentuan 17 provinsi ini didasarkan pada kebutuhan akan representasi wilayah pariwisata yang tersebar di pulau-pulau besar seluruh Indonesia (seperti Jawa, Bali, Sumatera, dan Sulawesi, dll). Sementara itu, 170 destinasi dipilih dengan mempertimbangkan popularitas dan ketersediaan ulasan yang memadai,

dengan menetapkan batas minimal 100 ulasan untuk setiap destinasi.

Data ulasan ini diambil selama periode waktu satu bulan sampai satu tahun terakhir, tergantung jumlah ulasan tiap destinasi yang didapatkan. Proses pengambilan data (*scraping*) dilakukan dengan mengembangkan *tool* sendiri menggunakan kode python dan selenium, yang berfungsi mengekstrak informasi penting seperti teks ulasan, rating, dan waktu unggah. Perlu dicatat bahwa penggunaan *tool* ini memiliki keterbatasan, seperti potensi pembatasan jumlah permintaan (*query*) atau variasi format data yang mungkin memerlukan penyesuaian teknis lebih lanjut.

Setelah data mentah berhasil dikumpulkan, langkah selanjutnya adalah pra-pemrosesan data [18]. Tahap krusial ini bertujuan membersihkan dan mempersiapkan data sebelum dianalisis, mengingat ulasan sering kali menggunakan bahasa informal, singkatan, atau mengandung *noise*. Proses pembersihan ini meliputi: *case folding* (menyeragamkan teks menjadi huruf kecil), *filtering* (menghapus data yang *redundan*, ambigu, atau karakter non-alfanumerik), *tokenizing* (memecah kalimat menjadi unit kata), *stopwords removal* (menghilangkan kata-kata umum), dan *normalization* [19]. Melalui serangkaian tahapan ini, data yang tidak relevan dapat dieliminasi, sehingga kualitas ulasan terjamin dan representasi tekstual menjadi Bahasa Indonesia yang baku siap untuk diolah lebih lanjut.

2.2. Analisis Sentimen

Fase analisis berpusat pada klasifikasi sentimen dari ulasan teks yang telah dibersihkan. Sebelum model dilatih, data yang berjumlah total 6320 ulasan ini harus terlebih dahulu diberi label. Penelitian ini mengadopsi pelabelan yang mengacu pada Lexicon-based Labelling dengan memanfaatkan Indonesian Sentimen Lexicon (InSet Lexicon) [20]. Dalam skema ini, penilaian dengan rating 1, 2, dan 3 diklasifikasikan sebagai sentimen negatif, sementara rating 4 dan 5 diklasifikasikan sebagai sentimen positif [21]. Justifikasi akademik untuk skema ini didasarkan pada kecenderungan bahwa rating 3 sering mencerminkan pengalaman yang netral atau kurang memuaskan, sehingga secara fungsional lebih dekat ke kutub negatif, sejalan dengan praktik analisis ulasan yang luas. Setelah pelabelan, dilakukan pemeriksaan keseimbangan dataset; data terbagi menjadi 4350 ulasan positif dan 2020 ulasan negatif.

Langkah selanjutnya adalah ekstraksi fitur data menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Untuk mendapatkan representasi tekstual yang optimal, kami melakukan konfigurasi spesifik pada TF-IDF: menggunakan n-gram unigram dan bigram (range 1,2), membatasi fitur hingga 95% dari total dokumen (*max_features*), dan menetapkan *min_df* sebesar 2 untuk mengabaikan istilah yang terlalu jarang muncul. Hasil vektor TF-IDF ini

kemudian digunakan sebagai masukan untuk melatih model klasifikasi *Support Vector Machine* (SVM) [22].

Implementasi SVM menggunakan *scikit-learn* (`sklearn.svm.SVC`) dijalankan dengan konfigurasi standar yang meliputi kernel RBF, nilai $C = 1.0$ sebagai parameter regularisasi, dan $\gamma = \text{"scale"}$ untuk menyesuaikan pengaruh fitur berdasarkan variansi data. Parameter tambahan seperti $\text{degree} = 3$, $\text{shrinking} = \text{True}$, dan $\text{probability} = \text{False}$ mengikuti pengaturan bawaan yang memungkinkan model membangun fungsi pemisah secara optimal tanpa perlu penyesuaian lanjutan. Konfigurasi dasar ini menjadi dasar pemodelan awal sebelum dilakukan optimasi lebih lanjut.

Optimasi kinerja dilakukan melalui teknik K-Fold Cross-Validation dengan pembagian awal data sebesar 80% untuk pelatihan dan 20% untuk pengujian. Proses validasi ini menghasilkan berbagai skenario pembagian, dan akurasi tertinggi digunakan untuk menentukan proporsi terbaik sebelum diterapkan pada pelatihan akhir model. Setelah model selesai dilatih, evaluasi dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix untuk menilai kemampuan model dalam mengklasifikasikan data baru secara lebih komprehensif. Setelah analisis sentimen menggunakan SVM ini berhasil mengklasifikasikan ulasan wisatawan, hasil sentimen positif dan negatif tersebut akan diintegrasikan sebagai komponen utama dalam pengembangan model rekomendasi wisata, yang juga mempertimbangkan data cuaca dan preferensi wisatawan.

2.3. Rekomendasi

Model rekomendasi dalam penelitian ini menggunakan pendekatan Content-Based Filtering (CBF) yang diperluas dengan data sentimen dan cuaca sehingga hasil rekomendasi menjadi lebih relevan dengan kondisi nyata. Profil preferensi pengguna disusun dari beberapa aspek, seperti kategori wisata, tema aktivitas, lokasi tujuan, dan waktu kunjungan, lalu dibandingkan dengan fitur konten setiap destinasi, seperti jenis wisata, fasilitas, kepadatan pengunjung, dan karakteristik lingkungan. Nilai Skor Relevansi Terpadu (R_{score}) dihitung untuk setiap destinasi dengan menggabungkan kesesuaian konten, sentimen ulasan, dan kondisi cuaca. Rumus yang digunakan adalah:

$$R_{score} = w_c \cdot Sim_{content} + w_s \cdot Sentiment_{score} + w_w \cdot Weather_{score} \quad (1)$$

dengan setiap komponen diberi bobot sesuai perannya. Kesesuaian konten dihitung melalui cosine similarity, skor sentimen diperoleh dari analisis ulasan pengunjung, sedangkan skor cuaca dihitung berdasarkan kondisi cuaca terbaru yang diperoleh dari BMKG.

Data cuaca diambil melalui API BMKG dengan menggunakan nama kelurahan sebagai titik referensi.

Informasi yang diambil meliputi kondisi langit, intensitas hujan, dan parameter dasar cuaca lainnya. Kelayakan cuaca kemudian ditentukan menggunakan kriteria sederhana, misalnya cerah (1.0), berawan (0.7), hujan ringan (0.4), dan hujan sedang atau deras (0.0). Kriteria tersebut digunakan karena berkaitan langsung dengan kenyamanan dan keamanan aktivitas wisata. Setelah semua komponen dihitung, sistem menampilkan lima destinasi dengan R_{score} tertinggi yang sudah disesuaikan dengan waktu keberangkatan pengguna. Pendekatan ini membuat model rekomendasi lebih fleksibel, realistis, dan peka terhadap kondisi yang benar-benar dialami wisatawan.

2.4. Visualisasi Model

Tahap akhir metodologi ini menekankan bagaimana hasil rekomendasi divisualisasikan, sehingga informasi dari perhitungan R_{score} dapat dipahami dengan mudah oleh pengguna. Visualisasi ini berfungsi sebagai antarmuka yang merangkum cara kerja model CBF adaptif—mulai dari analisis konten, sentimen, hingga kondisi cuaca—dan menyajikannya dalam bentuk yang lebih terstruktur dan informatif. Tujuannya bukan menampilkan cerita detail dari output, tetapi menekankan bagaimana sistem mengubah proses komputasi menjadi tampilan yang jelas dan mudah diikuti.

Struktur visual sengaja dibagi menjadi dua bagian utama agar pengguna bisa menangkap informasi secara cepat. Bagian pertama memperlihatkan lokasi acuan, yaitu titik awal yang digunakan sistem untuk menyesuaikan konteks rekomendasi. Bagian kedua menampilkan daftar rekomendasi teratas yang sudah diperingkat berdasarkan nilai R_{score} . Seluruh hasil ditampilkan dalam bentuk tabel agar lebih mudah dibaca dan dibandingkan antar lokasi. Pendekatan ini memastikan bahwa visualisasi tidak hanya berfungsi sebagai keluaran akhir, tetapi juga sebagai cara untuk membantu pengguna memahami logika rekomendasi yang dihasilkan model.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data ulasan (*review*) merupakan komponen krusial dalam sistem rekomendasi ini, khususnya untuk mengukur faktor Sentimen sebagai proksi pengalaman pengguna kolektif. Proses akuisisi data dilakukan dengan strategi sampling representatif yang mencakup 17 provinsi di Indonesia, merepresentasikan penyebaran geografis dari Sumatera hingga Papua. Pendekatan ini memastikan keberagaman karakteristik destinasi dan ulasan pengguna. Dari setiap provinsi yang dipilih, diambil sampel sebanyak 10 lokasi wisata yang memiliki popularitas tinggi atau rating yang signifikan, sehingga total data mencakup 170 lokasi wisata unik.

Secara keseluruhan, data yang berhasil dihimpun dari *review* Google Maps mencapai volume masif, yaitu

sebanyak 179.039 data ulasan. Volume data ini memberikan basis yang kuat dan memadai untuk pelatihan model analisis sentimen serta komputasi skor relevansi.

[37]:	lokasi	provinsi	ulasan	rating	tanggal	ulasan_clean	pos_score	neg_score
0	Air Terjun Sipisopiso	Sumatera Utara	Tempatnya bersih dan indah trus disediakan te...	5	08/28/2025 00:52:18	tempat bersih indah trus disediakan tempat sam...	7	-26
1	Air Terjun Sipisopiso	Sumatera Utara	Gorgeous plunging waterfall! The views from th...	5	08/26/2025 01:52:49	gorgeous plunging waterfall the views from the...	5	-3
2	Air Terjun Sipisopiso	Sumatera Utara	Suasana asri banget	5	08/22/2025 07:26:55	suasana asri	8	-3
3	Air Terjun Sipisopiso	Sumatera Utara	Yes, another waterfall with a viewing platform...	4	08/20/2025 06:07:50	yes another waterfall with a viewing platform ...	15	-6
4	Air Terjun Sipisopiso	Sumatera Utara	Sangat di sayangkan, kurang perawatan dr pemer...	4	08/19/2025 12:11:16	sangat sayangkan perawatan dr pemerintah setem...	11	-20

Gambar 2. Hasil Pra-Pemrosesan Data

Data mentah yang telah dihimpun memerlukan serangkaian tahapan pra-pemrosesan (*pre-processing*) untuk mengatasi *noise* dan standarisasi data, memastikan kualitas dan keakuratan input bagi model. Langkah pra-pemrosesan yang diterapkan meliputi pembersihan data (*cleaning*), normalisasi kasus (*case folding*), tokenisasi, penghapusan *stop words* untuk mengekstrak inti kata.

Gambar 2 menunjukkan hasil dari tahap pra-pemrosesan data sebelum dilakukan analisis sentimen, diwakili oleh sampel dataframe dari lokasi "Air Terjun Sipisopiso" di Sumatera Utara. Dapat dilihat kolom ulasan menunjukkan teks ulasan asli, kolom ulasan_clean menunjukkan hasil dari pembersihan dan normalisasi teks, menghilangkan karakter yang tidak relevan dan menstandarisasi kata. Teks pada kolom ini kemudian menjadi input utama untuk model analisis sentimen. Kolom pos_score dan neg_score merupakan hasil sementara dari leksikon sentimen, menunjukkan jumlah kata positif dan negatif yang terdeteksi dalam ulasan tersebut, yang kemudian akan dikalkulasi untuk mendapatkan skor sentimen final ($S_{Sentimen}$) untuk setiap destinasi. Proses pra-pemrosesan berhasil mengubah data mentah menjadi format yang terstruktur dan siap untuk tahap pemodelan selanjutnya, khususnya pada penerapan Analisis Sentimen untuk mengkuantifikasi faktor pengalaman pengguna.

3.2. Analisis Sentimen

Tahap selanjutnya dari pemrosesan data adalah Analisis Sentimen terhadap teks ulasan yang telah dibersihkan (ulasan_clean). Analisis ini bertujuan untuk mengkuantifikasi sentimen (positif atau negatif) dari setiap ulasan, yang kemudian akan diaggregasi untuk mendapatkan skor sentimen destinasi ($S_{Sentimen}$). Metode klasifikasi yang diterapkan adalah *Support Vector Machine* (SVM), yang dikenal efektif dalam tugas klasifikasi teks berdimensi tinggi.

Sebelum model SVM dilatih, fitur tekstual diekstraksi menggunakan *Term Frequency-Inverse Document*

Frequency (TF-IDF). Teknik ini digunakan untuk mengubah kumpulan teks tidak terstruktur menjadi representasi vektor numerik, di mana nilai setiap fitur (kata) mencerminkan frekuensi kemunculannya dalam dokumen relatif terhadap kumpulan teks secara keseluruhan.

Untuk menguji dan evaluasi generalisasi dan stabilitas model, dilakukan proses validasi silang (*cross-validation*) menggunakan skema K-Fold. Konfigurasi K-Fold yang dipilih sebagai yang paling optimal adalah K=5, yang secara implisit berarti pembagian data latih sebesar 80% dan data uji sebesar 20% pada setiap fold ($1-1/K$).

Tujuan dari pengujian dan evaluasi ini adalah untuk mendapatkan gambaran kinerja model yang komprehensif, stabil, dan tidak bias terhadap pembagian data tertentu. Proses ini melibatkan penghitungan empat metrik kunci: Akurasi, Presisi, *Recall*, dan *F1-Score* pada lima subset data pelatihan dan pengujian yang berbeda (*fold*).

Tabel 1 merupakan hasil dari penghitungan metrik evaluasi kinerja (yaitu accuracy, precision, recall, dan f1-score) yang dilakukan pada sebuah model klasifikasi.

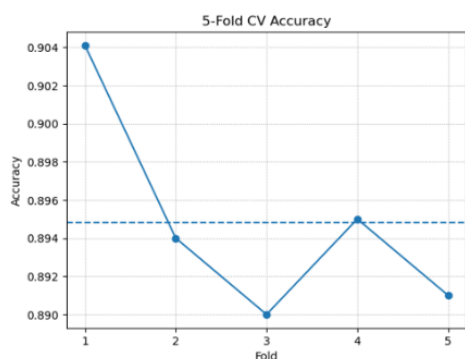
Penghitungan metrik ini dilakukan secara berulang untuk setiap fold (dari 1 hingga 5) sebagai bagian dari proses K-Fold Cross-Validation.

Tabel 1. Metrik per Fold

Fold	Accuracy	Precision	Recall	F1
1	90.41%	91.68%	94.39%	93.01%
2	89.40%	90.61%	94.09%	92.32%
3	89.00%	89.65%	94.68%	92.10%
4	89.50%	90.74%	94.09%	92.39%
5	89.10%	91.04%	93.06%	92.04%

Grafik yang ditampilkan selanjutnya merupakan representasi visual dari profil kinerja model, yang merangkum hasil metrik evaluasi utama: Akurasi, Presisi, Recall, dan F1-Score. Visualisasi ini sangat penting untuk menilai efektivitas prediktif model serta mengukur kualitas generalisasinya pada data yang belum terlihat.

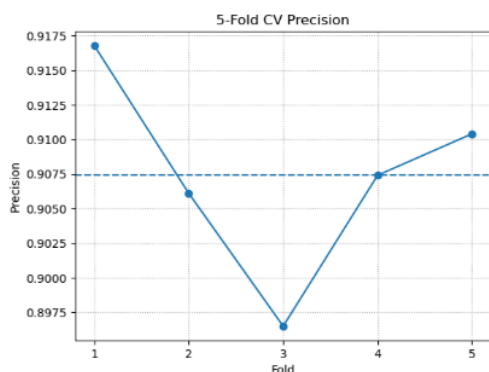
1. Akurasi



Gambar 3. Evaluasi Akurasi

Gambar 3 menyajikan akurasi model, yang merupakan metrik paling dasar yang menggambarkan proporsi prediksi yang benar secara keseluruhan. Akurasi tertinggi dicapai pada Fold 1 dengan nilai mencapai 0.904, menunjukkan kemampuan model yang unggul dalam memberikan rekomendasi yang benar pada pembagian data ini. Namun, terdapat fluktuasi yang signifikan, di mana nilai akurasi terendah tercatat pada Fold 3 yaitu tepat di angka 0.890. Akurasi rata-rata yang diwakili oleh garis putus-putus berada di kisaran 0.895. Variasi ini, meskipun relatif kecil (sekitar 1.4%), menyoroti bahwa kinerja model sedikit sensitif terhadap pembagian data pelatihan dan pengujian, sebuah fenomena yang biasa terjadi dalam *cross-validation*.

2. Presisi

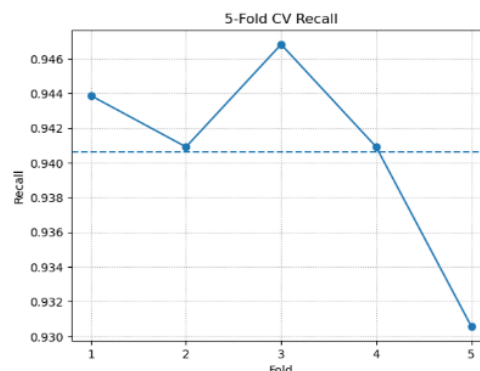


Gambar 4. Evaluasi Presisi

Gambar 4 mengukur tingkat ketepatan model dalam konteks rekomendasi; yaitu, seberapa sering rekomendasi yang diberikan model benar-benar relevan. Metrik ini sangat penting karena mencerminkan kemampuan model dalam meminimalkan *False Positive* (merekomendasikan sesuatu yang tidak disukai pengguna). Presisi tertinggi diamati pada Fold 1 dengan nilai mencapai 0.917, menunjukkan tingkat keandalan yang sangat tinggi pada fold ini. Konsisten dengan temuan akurasi, presisi terendah juga ditemukan pada Fold 3, dengan nilai di sekitar 0.897. Nilai presisi rata-rata berada di 0.9075. Hasil ini mengindikasikan bahwa pada Fold 3, model cenderung merekomendasikan lebih banyak item yang

ternyata tidak relevan (tingginya *false positive*), dibandingkan dengan Fold 1.

3. Recall

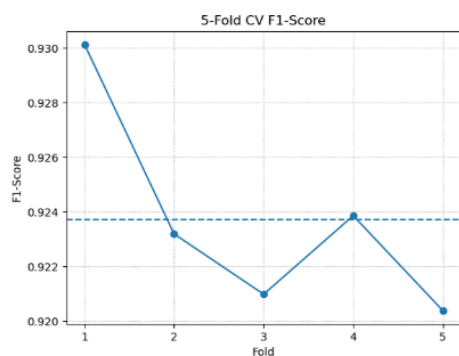


Gambar 5. Evaluasi Recall

Gambar 5 menampilkan hasil *Recall*, yang mengilustrasikan sensitivitas model, atau tingkat keberhasilan model dalam menangkap seluruh rekomendasi yang benar-benar relevan. Metrik ini mencerminkan kemampuan model dalam meminimalkan *False Negative* (gagal merekomendasikan item yang seharusnya disukai pengguna). Unikny, nilai *Recall* tertinggi justru dicapai pada Fold 3 dengan angka tertinggi mendekati 0.947. Hasil ini sangat kontras dengan metrik Presisi dan Akurasi di fold yang sama, yang berada pada titik terendah. *Recall* rata-rata adalah sekitar 0.940.

Kontradiksi antara Presisi rendah dan *Recall* tinggi pada Fold 3 menunjukkan adanya trade-off yang jelas. Penyebab utama anomali pada Fold 3 terletak pada ketidakseimbangan trade-off antara Precision dan Recall dibandingkan Fold lainnya. Pada Fold ini, model menghasilkan Recall paling tinggi (0.946824), menunjukkan sensitivitas superior dalam mengidentifikasi semua ulasan positif yang sebenarnya. Namun, tingginya Recall ini dicapai dengan mengorbankan Precision yang paling rendah (0.896503), yang berarti model menghasilkan banyak False Positives (ulasan negatif diklasifikasikan sebagai positif). Anomali ini mengindikasikan bahwa Fold 3 kemungkinan besar berisi proporsi kasus batas (ambigu) yang lebih tinggi, memaksa model menjadi sangat inklusif (recall tinggi) tetapi kurang tepat (precision rendah) dalam prediksinya.

4. F1-Score



Gambar 6. Evaluasi F1-Score

Gambar 6 memperlihatkan F1-Score, metrik gabungan yang sangat penting karena memberikan gambaran tunggal tentang kinerja model dengan menyeimbangkan Presisi dan *Recall*. Metrik ini sangat berguna untuk data yang tidak seimbang. Nilai *F1-Score* tertinggi, yang menunjukkan kinerja paling seimbang, konsisten dicapai pada Fold 1 dengan nilai sekitar 0.930. Sementara itu, nilai terendah tercatat pada Fold 5 (sekitar 0.920). Nilai rata-rata *F1-Score* berada di 0.924. *F1-Score* pada Fold 3 (sekitar 0.910) berada di bawah rata-rata, mengkonfirmasi bahwa meskipun nilai *Recall* sangat tinggi, Presisi yang rendah menyeret nilai *F1-Score* ke bawah.

```
[40]: (  fold accuracy precision recall  f1
0    1  0.994096  0.916786  0.943870  0.938131
1    2  0.894000  0.906117  0.940916  0.923188
2    3  0.890000  0.896503  0.946824  0.920977
3    4  0.895000  0.907407  0.940916  0.923858
4    5  0.891000  0.910405  0.930576  0.920380,
{'accuracy': 0.8948191808191808,
'precision': 0.9074436795665811,
'recall': 0.9406283840472674,
'f1': 0.9237068257954458})
```

Gambar 7. Cross Validation pada Data Latih

Gambar 7 menyajikan hasil evaluasi kinerja model klasifikasi SVM pada setiap fold dan rata-rata kinerja keseluruhannya. Hasil pengujian pada data uji (rata-rata 20% data pada setiap fold) menunjukkan kinerja model yang sangat baik. Rata-rata Akurasi (*Accuracy*) yang diperoleh adalah 0.8948 (sekitar 89.48%). Akurasi ini mengindikasikan bahwa model mampu memprediksi sentimen ulasan (positif atau negatif) dengan benar hampir 90% dari keseluruhan kasus. Presisi diperoleh rata-rata 0.9074, menunjukkan bahwa dari semua prediksi positif yang dibuat oleh model, sekitar 90.74% di antaranya benar-benar positif. *Recall* diperoleh rata-rata 0.9406, menunjukkan bahwa model mampu mengidentifikasi sekitar 94.06% dari semua ulasan positif yang sebenarnya ada dalam data uji. *F1-Score* diperoleh rata-rata 0.9237. Metrik *F1-Score* yang merupakan *harmonic mean* dari *Precision* dan *Recall*, menunjukkan keseimbangan kinerja model yang sangat baik, terutama relevan untuk dataset yang mungkin mengalami sedikit ketidakseimbangan kelas.

Kinerja yang stabil dan tinggi di seluruh 5 fold validasi silang mengonfirmasi bahwa model SVM, yang diekstraksi fitur dengan TF-IDF, efektif dan *reliable* untuk mengkuantifikasi sentimen pengguna,

sehingga layak digunakan sebagai salah satu faktor penentu R_{score} dalam sistem rekomendasi.

3.3. Rekomendasi

Setelah kuantifikasi sentimen ulasan berhasil divalidasi, langkah selanjutnya adalah implementasi proses Rekomendasi Adaptif menggunakan pendekatan *Content-Based Filtering* (CBF) yang disempurnakan dengan faktor kontekstual. Proses ini mengintegrasikan Skor Sentimen ($S_{Sentimen}$) dari ulasan pengguna dan data cuaca real-time dari BMKG, yang secara kolektif digunakan untuk mengkalkulasi Skor Relevansi Terpadu (R_{score}). Model rekomendasi beroperasi secara iteratif, dimulai dengan penetapan Lokasi Wisata Acuan sebagai titik query awal. Alur kerja ini mencakup Ekstraksi Data Acuan, Pengecekan Kualitas Konten, dan implementasi Pengecekan Adaptif Cuaca Real-time sebagai faktor guardrail. Jika kondisi cuaca pada lokasi acuan dinilai tidak optimal, rekomendasi lokasi tersebut akan dibatalkan demi menjamin keselamatan pengguna. Selanjutnya, R_{score} dihitung untuk semua destinasi potensial dengan mengintegrasikan kecocokan konten, $S_{Sentimen}$, dan data cuaca, sebelum Hasil Rekomendasi akhir disajikan kepada pengguna berdasarkan ketiga faktor tersebut.

Meskipun model analisis sentimen SVM yang digunakan terbukti stabil dan layak diintegrasikan (dengan *F1-Score* rata-rata 0.9237), penelitian ini menemukan adanya anomali *trade-off* yang signifikan pada Fold 3, menunjukkan perlunya optimasi *Precision*. Untuk memastikan $S_{Sentimen}$ yang digunakan dalam perhitungan R_{score} memiliki kredibilitas dan akurasi tinggi (meminimalkan False Positives), direkomendasikan untuk melakukan penyetelan *hyperparameter* SVM yang memprioritaskan *Precision*. Selain itu, analisis kualitatif mendalam pada sampel data anomali serta penyempurnaan pra-pemrosesan data seperti implementasi teknik *Negation Handling* disarankan. Tindakan korektif ini akan secara langsung meningkatkan keandalan output sentimen, yang pada gilirannya akan memperkuat validitas dan akurasi R_{score} serta kinerja keseluruhan sistem rekomendasi adaptif.

3.4. Visualisasi Model

Visualisasi hasil akhir model berfungsi sebagai antarmuka yang mengkomunikasikan keputusan sistem kepada pengguna, sekaligus menunjukkan transparansi mekanisme adaptif yang diimplementasikan. Gambar 8 menampilkan keluaran dari proses rekomendasi, menyajikan 5 lokasi terbaik yang dihasilkan setelah integrasi faktor sentimen dan cuaca.

```

=== Info Lokasi Acuan ===
Nama      : Gurun_Pasir_Bintan
Kelurahan : Busung
Rating    : 4.45
Ulasan    : 689
Sentimen+ : 308
Cuaca     : berawan
Status    : ❌ Tidak masuk rekomendasi karena cuaca berawan

=== Rekomendasi 5 Lokasi Terbaik ===
      lokasi      kelurahan  kecamatan \
11  Istana_Siak_Sri_Indrapura  Kampung Dalam  Siak
16  Masjid_Raya_Medan          Mesjid         Medan Kota
26  Taman_Nasional_Kerinci_Seblat  Lubung Gadang Selatan  Sangir
21  Pantai_Nongsa_Batam          Sambau         Nongsa
20  Pantai_Carocok_Painan        Painan         IV Jurai

      kabupaten  rata_rating  total_ulasan  sentimen_pos  cuaca
11  Siak          4.709925    4161          2635          cerah berawan
16  Medan         4.778571    1120          901           cerah berawan
26  Solok Selatan  4.559701    670           406           cerah
21  Batam         4.176360    533           378           cerah berawan
20  Pesisir Selatan  4.490769    650           352           cerah

```

✅ Hasil rekomendasi tersimpan di 'rekomendasi_lokasi.xlsx'

Gambar 8. Rekomendasi Destinasi Wisata Terbaik Berdasarkan Analisis Sentimen dan Cuaca

Model Rekomendasi dirancang untuk beroperasi berdasarkan konteks geografis yang spesifik, sehingga memerlukan masukan berupa Lokasi Acuan dan Kelurahan terkait. Data Kelurahan ini memiliki fungsi ganda: pertama, sebagai mekanisme awal untuk membatasi rekomendasi wilayah di sekitar lokasi acuan, dan kedua, sebagai parameter kunci untuk mendapatkan kondisi cuaca real-time melalui API BMKG. Namun, implementasi saat ini masih menggunakan nama Kelurahan sebagai parameter masukan utama untuk permintaan data cuaca dari API BMKG, dan belum memanfaatkan alamat koordinat yang lebih presisi dari Lokasi Acuan. Ketergantungan pada Kelurahan sebagai batas administratif yang relatif luas dapat menimbulkan isu granularity, di mana kondisi cuaca yang dilaporkan mungkin tidak sepenuhnya mencerminkan kondisi mikro pada titik koordinat spesifik destinasi. Untuk meningkatkan akurasi kontekstual dari faktor guardrail cuaca, langkah selanjutnya yang krusial adalah memigrasikan parameter masukan API BMKG dari nama Kelurahan menjadi koordinat lokasi acuan, sehingga data cuaca yang diperoleh benar-benar merepresentasikan kondisi di lokasi yang akan dikunjungi.

Bagian awal visualisasi menampilkan Informasi Lokasi Acuan (misalnya, "Gurun_Pasir_Bintan"). Komponen ini penting untuk validasi kontekstual. Selain menampilkan atribut statis seperti Nama, Kelurahan, Rating (4.45), dan jumlah Ulasan (689), sistem juga menyajikan metrik internal Sentimen Positif (308) dan, yang paling kritis, Kondisi Cuaca Real-time (Berawan). Dalam contoh ini, faktor cuaca memicu mekanisme guardrail sistem, yang ditunjukkan oleh Status: "❌ Tidak masuk rekomendasi karena cuaca berawan". Ini memvalidasi implementasi adaptif model: meskipun lokasi acuan mungkin memiliki rating dan sentimen yang baik, ketidaksesuaian kondisi lingkungan aktual (cuaca) secara otomatis mengecualikannya dari rekomendasi, memastikan keselamatan dan pengalaman pengguna yang optimal.

Bagian Rekomendasi 5 Lokasi Terbaik menyajikan keluaran dalam format tabular terstruktur. Format ini

dirancang untuk memungkinkan perbandingan cepat oleh pengguna, didukung oleh data terkuantifikasi. Gambar 8 mencakup :

1. Informasi Geografis: Menyajikan Lokasi, Kelurahan, Kecamatan, dan Kabupaten (misalnya, Istana Siak Sri Indrapura di Kabupaten Siak), yang memenuhi kriteria Content-Based Filtering (CBF) berdasarkan profil geografis dan preferensi pengguna.
2. Metrik Performa: Mencakup rata_rating, total_ulasan, dan sentimen_pos. Nilai-nilai ini, seperti Sentimen Positif yang tinggi (misalnya, 2635 untuk Istana Siak), menunjukkan kontribusi signifikan dari faktor pengalaman sosial (SSentimen) dalam menentukan Rscore lokasi tersebut.
3. Faktor Adaptif Cuaca: Kolom cuaca menjadi justifikasi kontekstual utama. Rekomendasi yang disajikan (misalnya, Pantai Carocok Painan dan Solok Selatan) sebagian besar menunjukkan kondisi "cerah" atau "cerah berawan".

Penyajian ini secara eksplisit menunjukkan bahwa rekomendasi yang diberikan oleh model telah melalui proses filtering dan penyesuaian yang ketat berdasarkan sentimen pengguna (pengalaman sosial) dan kondisi lingkungan (cuaca), yang secara efektif memvalidasi efektivitas pendekatan CBF adaptif yang diusulkan dalam penelitian ini. Hasil output ini membuktikan bahwa model tidak hanya bergantung pada kecocokan atribut statis, tetapi mampu memberikan rekomendasi yang adaptif dan sensitif konteks demi meningkatkan utilitas praktis sistem rekomendasi.

4. Kesimpulan

Analisis 5-Fold Cross-Validation (CV) membuktikan bahwa model klasifikasi sentimen berbasis SVM-TFIDF bekerja sangat stabil, dengan rata-rata metrik kinerja utama di atas 0.89. Kinerja tinggi ini memvalidasi skor sentimen sebagai penentu yang andal dalam perhitungan R_{score} model rekomendasi yang dikembangkan. Meskipun terdapat anomali kecil pada Fold 3 yang menunjukkan perlunya pemahaman lebih lanjut tentang data yang ambigu, keandalan keseluruhan model mendukung integrasi sentimen untuk meningkatkan akurasi prediktif sistem Content-Based Filtering (CBF) kontekstual. Model rekomendasi adaptif ini membawa implikasi strategis yang penting, antara lain: mendukung peningkatan mutu berbasis bukti dengan memvalidasi sentiment score untuk alokasi sumber daya; memajukan inisiatif Pariwisata Aman Bencana melalui integrasi guardrail cuaca BMKG; serta mendorong pemerataan promosi destinasi non-mainstream dengan menyediakan justifikasi rekomendasi yang transparan dan up-to-date.

Meskipun kontribusi metodologisnya signifikan, penelitian ini memiliki keterbatasan yang membuka jalan bagi riset selanjutnya. Model yang ada saat ini

belum mengakomodasi preferensi cuaca yang spesifik dari pengguna dan menggunakan skema biner (layak/tidak layak) yang sederhana. Selain itu, sistem ini belum dilengkapi dengan mekanisme re-training otomatis untuk beradaptasi dengan perubahan pola ulasan jangka panjang. Oleh karena itu, penelitian di masa depan disarankan untuk fokus pada pengembangan personalisasi cuaca multikriteria dan mengintegrasikan model deep learning yang lebih canggih (seperti BERT) untuk analisis sentimen, serta merancang arsitektur sistem yang mampu melakukan pembaruan data dan bobot secara real-time guna menjaga adaptabilitas model.

Ucapan Terimakasih

Penelitian ini dengan nomor kontrak SPK/133/LPPMUNJAYA/VI/2025 telah terlaksana dengan baik, atas dukungan berbagai pihak. Peneliti mengucapkan terimakasih atas pendanaan hibah yang telah diberikan oleh Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) Kementerian Pendidikan Tinggi, Sains, dan Teknologi (Kemendiknasaintek) Republik Indonesia, serta fasilitas pendukung lainnya dalam penelitian yang disediakan oleh Universitas Jenderal Achmad Yani Yogyakarta.

Daftar Rujukan

[1] B. D. dan Sistem Informasi, "Perkembangan Perjalanan Wisatawan Nusantara Tahun 2019-2024," 2025. [Online]. Available: <https://kemenpar.go.id/statistik-wisatawan-nusantara/perkembangan-perjalanan-wisatawan-nusantara>

[2] L. I. Kirana, "Inilah Daftar Tipe Wisata Terfavorit Warga Indonesia 2024," 2024. [Online]. Available: <https://goodstats.id/article/pilihan-wisata-terfavorit-warga-ri-2024-wPjvq>

[3] W. Khoffah, D. Nur Rahayu, and A. Maulana Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 171–180, Jan. 2022, doi: 10.35969/INTERKOM.V16I4.192.

[4] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/MALCOM.V4I1.1066.

[5] M. Thoriq, F. Maulana, Y. S. Eirlangga, N. Hayati, M. A. Madani, and F. Maulana, "Implementasi Algoritma Naive Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzka," *Jurnal FASILKOM (teknologi inFormASi dan Ilmu KOMputer)*, 2025.

[6] N. D. H. Sadikin, S. Susanti, and A. R. Sanjaya, "Analisis Sentimen Publik Terhadap Kampanye Pengurangan Sampah Plastik Menggunakan Algoritma Naive Bayes," *Jurnal FASILKOM (teknologi inFormASi dan Ilmu KOMputer)*, vol. 15, no. 2, pp. 202–212, Aug. 2025, doi: <https://doi.org/10.37859/jf.v15i2.9574>.

[7] R. Sari, "ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA DUNIA FANTASI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)," *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 8, no. 1, Mar. 2020, doi: 10.31294/EVOLUSI.V8I1.7371.

[8] S. Widodo and B. Hartono, "Analisis Sentimen Pengguna Google Terhadap Destinasi Wisata Di Kota Semarang Menggunakan Metode K-Nearest Neighbor," *Progresif:*

Jurnal Ilmiah Komputer, vol. 19, no. 2, pp. 545–554, Aug. 2023, doi: 10.35889/PROGRESIF.V19I2.1364.

[9] M. S. Syahlan, D. Irmayanti, and S. Alam, "ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)," *Simtek : jurnal sistem informasi dan teknik komputer*, vol. 8, no. 2, pp. 315–319, Oct. 2023, doi: 10.51876/SIMTEK.V8I2.281.

[10] T. E. Tarigan, E. Faizal, and Sumiyatun, "Model Rekomendasi Wisata dengan Pendekatan Collaborative Filtering," *Fahma : Jurnal Informatika Komputer, Bisnis dan Manajemen*, vol. 21, no. 2, pp. 56–64, May 2023, doi: 10.61805/FAHMA.V21I2.18.

[11] F. F. Murzani and D. B. Arianto, "IMPLEMENTASI METODE COLLABORATIVE FILTERING PADA ALGORITMA SISTEM REKOMENDASI DESTINASI WISATA DI ACEH," *Jurnal Komputer, Informasi Teknologi, dan Elektro*, vol. 8, no. 3, p. 1, Dec. 2023, doi: 10.24815/KITEKTRO.V8I3.36168.

[12] L. Cahyani, N. Sephiana, M. Tahir, and J. Aisyiah, "Sistem Rekomendasi Wisata Kuliner Madura Menggunakan Content Based Filtering," *Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 16, no. 1, pp. 31–38, Jul. 2024, doi: 10.35891/EXPLORIT.V16I1.5366.

[13] K. Christofer, A. J. Santoso, and A. W. R. Emanuel, "Sistem Rekomendasi Objek Pariwisata di Pontianak Berbasis Android Menggunakan Metode Content-Based Filtering," *Jurnal Informatika Atma Jogja*, vol. 1, no. 1, pp. 42–49, Dec. 2020, [Online]. Available: <https://ojs.uajy.ac.id/index.php/jiaj/article/view/3837>

[14] D. Pratiwi, A. Asrianda, and L. Rosnita, "Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 2, pp. 85–96, 2024, doi: 10.54082/JIKI.169.

[15] A. D. Aryanto, A. Primadewi, N. Agung, and A. D. Aryanto, "Rekomendasi Wisata Kabupaten Magelang menggunakan Metode Content-Based Filtering dan Location-Based Service," *Jurnal FASILKOM (teknologi inFormASi dan Ilmu KOMputer)*, 2025, doi: <https://doi.org/10.37859/jf.v15i1.8156>.

[16] Musyrifah, Sulfiyanti, Irfan, Asmawati, and N. Zulkarnaim, "Sistem Rekomendasi Berbasis-Konten Untuk Pengembangan Web Smart Tourism," *Jurnal Komputer Terapan*, vol. 8, no. 1, pp. 143–150, Jun. 2022, doi: 10.35143/JKT.V8I1.5214.

[17] R. Faurina and E. Sitanggang, "Implementasi Metode Content-Based Filtering dan Collaborative Filtering pada Sistem Rekomendasi Wisata di Bali," *Techno.COM*, vol. 22, no. 4, pp. 870–881, 2023, [Online]. Available: <https://publikasi.dinus.ac.id/index.php/technoc/article/view/8556/4151>

[18] A. A. Daniswara and I. K. D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," *Journal of Informatics and Computer Science (JINACS)*, vol. 5, no. 01, pp. 97–100, Jul. 2023, doi: 10.26740/JINACS.V5N01.P97-100.

[19] K. Kharisma and U. S. Aesy, "ANALISIS TINGKAT KEBERMANFAATAN MYPERTAMINA MENGGUNAKAN K-MEANS CLUSTERING," *Journal of Information System Management (JOISM)*, vol. 4, no. 2, pp. 91–96, Jan. 2023, doi: 10.24076/JOISM.2023V4I2.967.

[20] Fajri Koto and Gemala Y. Rahmaningtyas, "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *2017 International Conference on Asian Language Processing*, IEEE, 2017. doi: 10.1109/IALP.2017.8300625.

[21] P. W. Cahyo, U. S. Aesy, and B. D. Santosa, "Topic Sentiment Using Logistic Regression and Latent Dirichlet Allocation as a Customer Satisfaction Analysis Model," *JURNAL INFOTEL*, vol. 16, no. 1, pp. 1–16, Jan. 2024, doi: 10.20895/INFOTEL.V16I1.1081.

[22] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naive Bayes Classifier (NBC) dalam Analisis

Sentimen Publik terhadap Konsep Child-free di Media Sosial
Twitter,” *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 1, pp.
93–104, Jun. 2023, doi: 10.21009/JSA.07109.