ANALISIS SENTIMEN MULTI-PLATFORM UNTUK STRATEGI PEMASARAN DIGITAL PADA DESTINASI WISATA

Endah Septa Sintiya, Syahira Azizah Rendra Putri, Gunawan Budi Prasetyo, Candra Bella Vista
Teknologi Informasi, Politeknik Negeri Malang
Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang
e.septa@polinema.ac.id

ABSTRAK

Media sosial dan platform ulasan digital berperan penting dalam membentuk citra destinasi wisata, namun perbedaan karakteristik antar platform sering menghasilkan distribusi sentimen yang tidak merata. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap destinasi wisata melalui pendekatan multi-platform serta merumuskan strategi pemasaran digital yang efektif. Data dikumpulkan dari Google Maps, TikTok, dan Instagram menggunakan teknik scraping. Pelabelan sentimen dilakukan dengan pendekatan hybrid yang mengombinasikan metode lexicon-based dan Support Vector Machine (SVM). Proses klasifikasi menggunakan algoritma Random Forest dengan fitur TF-IDF, serta evaluasi model dilakukan menggunakan 10-Fold Cross Validation. Hasil penelitian menunjukkan bahwa model Random Forest mencapai akurasi sebesar 87,60%, precision 87,68%, recall 87,93%, dan F1-score 87,55%. Analisis distribusi sentimen mengungkap perbedaan karakteristik antar platform, di mana Google Maps didominasi sentimen negatif (47,4%), Instagram cenderung netral (46,1%), dan TikTok menunjukkan dominasi sentimen netral (53,5%). Selain itu, hasil analisis divisualisasikan menggunakan Priority Action Matrix yang mengidentifikasi strategi *Quick Wins* seperti peningkatan respons pada Google Maps dan kampanye *user-generated content* sebagai prioritas utama karena berdampak tinggi dengan upaya rendah. Penelitian ini memberikan kontribusi dalam pengembangan analisis sentimen multi-platform serta perumusan strategi pemasaran digital berbasis data pada destinasi wisata.

Kata kunci: *analisis sentimen; multi-platform; pemasaran digital; destinasi wisata; random forest*

1. PENDAHULUAN

Industri pariwisata merupakan salah satu sektor strategis dalam pembangunan ekonomi Indonesia. Data dari Kementerian Pariwisata dan Ekonomi Kreatif (2023) mencatat pergerakan wisatawan nusantara pada semester I-2023 mencapai 433,57 juta perjalanan, meningkat 12,57% dibandingkan tahun sebelumnya. Dalam era digital ini, pembentukan citra destinasi wisata mengalami transformasi signifikan melalui konten yang dibagikan di media sosial.

Besarnya jumlah pengguna media sosial di Indonesia membuka peluang besar bagi industri pariwisata. Data terbaru dari We Are Social dan Meltwater (2024) menunjukkan dari total 278,7 juta penduduk, sebanyak 139 juta aktif menggunakan media sosial. Khususnya untuk platform Instagram dan TikTok, Instagram digunakan oleh 85,3% pengguna internet Indonesia, sedangkan TikTok diakses oleh 73,5% pengguna.

User Generated Content (UGC) menjadi faktor penting dalam mempengaruhi keputusan calon wisatawan. UGC memiliki kredibilitas yang lebih tinggi dibandingkan konten promosi resmi karena berasal dari pengalaman nyata pengunjung [1].selain media sosial, 78% pengunjung lebih mempercayai ulasan pada Google Maps dibandingkan rekomendasi seseorang [2].

Penelitian sebelumnya menunjukkan manfaat analisis sentimen dalam mendukung pemasaran destinasi wisata, namun sebagian besar hanya berfokus pada aspek teknis tanpa mempertimbangkan elemen penting seperti daya tarik destinasi dan aktivitas wisatawan [3]. Penelitian lain menunjukkan bahwa kombinasi metode Lexicon-Based, SMOTE, dan algoritma pembelajaran mesin seperti SVM dan Random Forest efektif untuk menganalisis sentimen ulasan wisatawan [4].

2. TINJAUAN PUSTAKA

Sejumlah penelitian telah dilakukan terkait analisis sentimen pada berbagai domain dengan pendekatan metode dan teknik yang beragam. Penelitian [5] menggunakan algoritma Random Forest dengan ekstraksi fitur TF-IDF pada ulasan hotel dari TripAdvisor, serta menerapkan pelabelan otomatis berbasis rating. Hasil penelitian menunjukkan bahwa akurasi tertinggi sebesar 87,23% diperoleh tanpa proses stemming, yang mengindikasikan bahwa stemming tidak selalu meningkatkan performa model pada teks berbahasa Indonesia.

Dalam konteks pariwisata, penelitian [3] menegaskan bahwa analisis sentimen terhadap ulasan wisatawan di media sosial memiliki peran penting dalam mendukung optimalisasi strategi pemasaran destinasi. Sentimen positif terbukti mampu memengaruhi persepsi dan keputusan calon wisatawan. Hal ini diperkuat oleh penelitian [4] yang mengombinasikan pendekatan Lexicon-Based dengan algoritma Support Vector Machine (SVM) dan Random Forest pada data ulasan wisata di Labuan Bajo. Dengan penerapan SMOTE untuk mengatasi ketidakseimbangan data, performa model meningkat signifikan, di mana SVM mencapai akurasi 89% dan Random Forest sebesar 87%.

Selain itu, analisis sentimen juga banyak diterapkan pada domain e-commerce. Penelitian [6] menganalisis opini pengguna terhadap tiga marketplace besar di Indonesia menggunakan kombinasi metode Lexicon-Based dan Naïve Bayes. Hasilnya menunjukkan bahwa model TextBlob-Lexicon dengan Multinomial Naïve Bayes menghasilkan akurasi tertinggi sebesar 90%. Temuan serupa juga ditunjukkan oleh penelitian [7], yang mengombinasikan metode Naïve Bayes dan Lexicon-Based pada data Twitter terkait dompet elektronik. Penelitian tersebut menekankan pentingnya tahap preprocessing, seperti normalisasi kata tidak baku, tokenisasi, stopword removal, dan stemming, yang mampu meningkatkan akurasi hingga 88,56%.

Lebih lanjut, perbandingan performa algoritma juga dilakukan dalam penelitian [8], yang menunjukkan bahwa Random Forest memiliki keunggulan dibandingkan Support Vector Machine dalam menangani data berdimensi tinggi dan noise, dengan akurasi masing-masing sebesar 91% dan 90%. Hal ini menunjukkan bahwa Random Forest merupakan salah satu algoritma yang robust dalam analisis sentimen, terutama ketika dikombinasikan dengan teknik ekstraksi fitur seperti TF-IDF.

Beberapa penelitian sebelumnya menunjukkan bahwa analisis sentimen telah dilakukan dengan berbagai algoritma klasifikasi yang memiliki keunggulan masing-masing. Namun demikian, tantangan utama yang masih dihadapi adalah perbedaan karakteristik data pada masing-masing platform, khususnya dalam konteks dataset multi-platform. Perbedaan pola interaksi pengguna antara media sosial dan platform ulasan menyebabkan distribusi sentimen yang tidak merata, sehingga memengaruhi performa model klasifikasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengintegrasikan data dari tiga platform berbeda, yaitu Google Maps, TikTok, dan Instagram, dalam satu kerangka analisis sentimen. Metode yang digunakan adalah Random Forest, yang dikenal memiliki kemampuan dalam menangani data berdimensi tinggi, noise, serta ketidakseimbangan data. Selain itu, penelitian ini mengusulkan pendekatan hybrid labeling yang mengombinasikan metode lexicon-based dengan Support Vector Machine (SVM). Pendekatan ini diharapkan mampu meningkatkan akurasi pelabelan dengan memanfaatkan kekuatan representasi semantik kata serta kemampuan generalisasi dari machine learning.

Kombinasi antara analisis multi-platform, algoritma Random Forest, dan teknik hybrid labeling merupakan pendekatan yang relatif belum banyak diterapkan, khususnya dalam konteks analisis sentimen pada destinasi pariwisata di Indonesia. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen yang lebih adaptif terhadap heterogenitas data.

Santerra de Laponte sebagai salah satu destinasi wisata unggulan di Malang menghadapi tantangan dalam mempertahankan citra positif di era digital. Oleh karena itu, penelitian ini secara khusus bertujuan untuk menganalisis sentimen pengguna dari berbagai platform guna mengevaluasi efektivitas strategi pemasaran digital yang telah diterapkan. Fokus analisis diarahkan pada pengaruh konten media sosial serta ulasan pada Google Maps terhadap persepsi publik terhadap destinasi tersebut.

3. METODE PENELITIAN

Penelitian ini dimulai dengan pemilihan konten sebagai sumber data yang kemudian dilakukan proses pengumpulan (*scraping*) dan pembuatan dataset. Setelah itu, data melalui tahap preprocessing meliputi cleansing, case folding, stopwords removal, dan tokenisasi. Untuk pelabelan sentimen, digunakan pendekatan hybrid yang menggabungkan model Random Forest yang dilatih dengan data berlabel, skoring berdasarkan *lexicon*, dan *rule-based decision function* untuk menghasilkan label sentimen akhir. Data yang telah diberi label kemudian ditransformasikan menggunakan TF-IDF dan diklasifikasikan dengan algoritma Random Forest menggunakan teknik *K-Fold Cross Validation* untuk memperoleh rata-rata akurasi model. Tahap akhir adalah evaluasi model menggunakan metrik confusion matrix, accuracy, precision, recall, dan F1-score, yang kemudian dianalisis dan divisualisasikan sebagai hasil akhir penelitian.

3.1. Pengumpulan Data

Data dikumpulkan melalui teknik scraping menggunakan platform Apify dari tiga platform digital:

a. Google Maps: 896 ulasan (Januari 2021 - Mei 2025)
b. TikTok: 557 komentar (Januari 2024 - Mei 2025)
c. Instagram: 154 komentar (Januari 2024 - Mei 2025)

3.2. Preprocessing

Tahapan preprocessing meliputi:

a. *Cleansing*: Menghapus URL, mention, hashtag, emoji, dan karakter non-alfabet
b. *Case Folding*: Mengubah semua huruf menjadi lowercase
c. *Normalization*: Mengubah kata tidak baku menjadi bentuk standar
d. *Stopword Removal*: Menghapus kata-kata umum yang tidak bermakna
e. *Tokenizing*: Memecah teks menjadi token individual

3.3. Pelabelan Hybrid

Penelitian ini menggunakan pendekatan *hybrid* yang menggabungkan metode lexicon-based dengan Support Vector Machine (SVM). *Lexicon* disusun berdasarkan konteks ulasan destinasi wisata dan

divalidasi oleh dosen ahli Bahasa Indonesia. Algoritma hybrid bekerja dengan logika:

a. Analisis skor lexicon berdasarkan kata positif dan negatif
b. Jika perbedaan skor tidak signifikan, gunakan prediksi SVM
c. Label final ditentukan berdasarkan confidence score

3.4. TF-IDF

Metode TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensinya dalam sebuah dokumen dibandingkan dengan seluruh dokumen dalam dataset. Kata-kata yang sering muncul di satu dokumen tetapi jarang muncul di dokumen lain akan mendapatkan bobot yang lebih tinggi, sehingga dianggap lebih penting dalam menentukan karakteristik teks tersebut. TF mengukur seberapa sering suatu kata muncul dalam dokumen tertentu, sedangkan IDF mengukur seberapa langka kata tersebut di seluruh koleksi dokumen. Kombinasi TF dan IDF menghasilkan bobot yang membantu model machine learning memahami pentingnya setiap kata dalam konteks klasifikasi sentimen.

Perhitungan TF-IDF menggunakan rumus sebagai berikut:

$$Wt = tf_t \times idf_t \quad (1)$$

$$idf_t = \log \frac{N}{df_t} \quad (2)$$

di mana Wt adalah bobot TF-IDF untuk term t, tft adalah frekuensi term t dalam dokumen, N adalah jumlah total dokumen dalam koleksi, dan dft adalah jumlah dokumen yang mengandung term t.

3.5. Random Forest

Fitur ekstraksi menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) untuk mengubah teks menjadi representasi numerik. Klasifikasi sentimen menggunakan algoritma Random Forest dengan 100 estimators. Model Random Forest membangun sejumlah pohon keputusan secara paralel selama proses training, di mana setiap pohon memberikan prediksi dan hasil akhir ditentukan berdasarkan mekanisme voting mayoritas dari seluruh pohon keputusan. Keunggulan Random Forest terletak pada kemampuannya menangani dataset berskala besar dengan tingkat akurasi tinggi dan mengurangi risiko overfitting yang sering terjadi pada decision tree tunggal. [9]

Algoritma Random Forest menggunakan rumus ensemble sebagai berikut:

$$F(x) = \frac{1}{J} \sum_{j=1}^{J} hj(x) \quad (3)$$

di mana F(x) adalah output dari Random Forest, J adalah jumlah pohon dalam ensemble, dan $h_j(x)$ adalah output dari pohon ke-j.

3.6. Evaluasi K-Fold Cross Validation

Evaluasi model dilakukan menggunakan 10-Fold Cross Validation dengan metrik accuracy, precision, recall, dan F1-score. Teknik K-Fold Cross Validation membagi dataset menjadi 10 bagian yang saling terpisah, di mana pada setiap iterasi, model dilatih menggunakan 9 bagian dan diuji pada 1 bagian yang tersisa. Proses ini dilakukan secara bergantian hingga semua bagian mendapat giliran sebagai data uji. Dengan menggunakan seluruh data sebagai data latih dan uji secara bergantian, diperoleh gambaran yang lebih menyeluruh terhadap kinerja model dan dapat mengurangi risiko overfitting dan underfitting. [10]

3.7. Visualisasi

Untuk menyajikan hasil analisis sentimen secara komprehensif, penelitian ini menggunakan berbagai jenis visualisasi. Word cloud digunakan untuk menampilkan kata-kata yang paling sering muncul dalam setiap kategori sentimen (positif, negatif, netral), di mana ukuran kata mencerminkan frekuensi kemunculannya. Grafik distribusi sentimen berupa pie chart dan bar chart menunjukkan perbandingan persentase sentimen antar platform untuk memberikan gambaran karakteristik unik setiap platform. Tren sentimen temporal divisualisasikan menggunakan line chart untuk menunjukkan perubahan pola sentimen dari waktu ke waktu pada masing-masing platform. Confusion matrix digunakan untuk mengevaluasi performa klasifikasi dengan menampilkan distribusi prediksi yang benar dan salah. Selain itu, priority action matrix divisualisasikan untuk memberikan rekomendasi strategis berdasarkan tingkat upaya implementasi dan dampak bisnis.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang digunakan pada penelitian ini diperoleh dari media sosial Instagram, TikTok dan Google maps melalui aplikasi pihak ketiga **Apify**. Dimana scrapping data melalui **Apify** ini diambil dari konten TikTok dan Instagram sejak 2024 Januari hingga 2025 Mei. Proses data menghasilkan 1607 data dari ketiga platform tersebut.

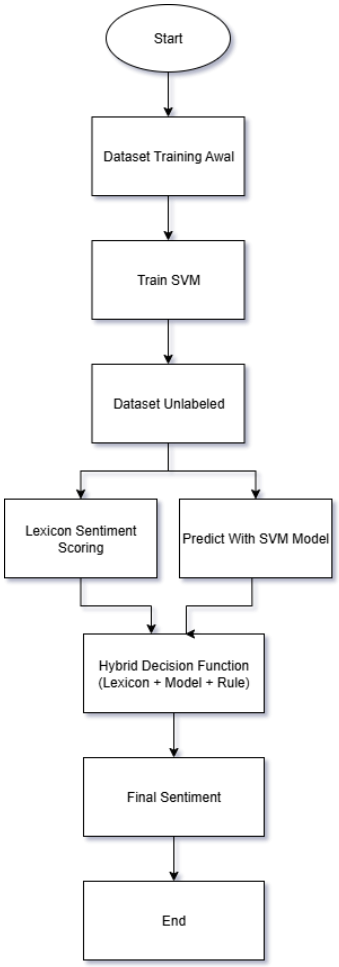
4.2. Preprocessing Data

Preprocessing data merupakan tahap awal yang sangat penting dalam analisis sentimen ini. Kumpulan tweet yang berisi opini terkait Santerra de Laponte telah diproses melalui berapa langkah untuk memastikan data yang digunakan bersih dan siap diolah oleh algoritma Random Forest. Proses tersebut mencakup cleansing, case folding, normalisasi, stopword removal dan tokenizing. Dengan mengurangi noise dan mengekstrak fitur linguistik yang relevan, diharapkan model mampu mengenali pola sentimen pada data lebih akurat. Adapun hasil dari data mentah dan sesudah preprocessing pada Tabel 1, sebagai berikut:

Tabel 1. Preprocessing Data

Data Mentah	Data Setelah Preprocessing
Taman bunga yang sangat luas dengan berbagai macam area foto, belajar dan bermain. Cukup terik di siang hari Banyak replika kota Cocok ajak anak anak kesini	['taman', 'bunga', 'luas', 'area', 'foto', 'belajar', 'bermain', 'terik', 'siang', 'replika', 'kota', 'cocok', 'ajak', 'anak']
Bagus banget tiap tahun selalu liburan kesini sama keluarga, wisata wahananya juga upgrade lebih update banyak baru baru dan menarik. top mantul cocok buat wisata bareng keluarga pasangan teman dll	['bagus', 'liburan', 'keluarga', 'wahananya','upgrade', 'diperbarui', 'menarik','mantap', 'cocok', 'bareng', 'keluarga', 'pasangan', 'teman']
Tempat lumayan bagus, tiket masuk murah tetapi harus bayar lagi untuk wahana nya, cocok untuk liburan keluarga	['lumayan', 'bagus', 'murah','bayar', 'cocok', 'liburan', 'keluarga']
Cukup lengkap cocok untuk berbagai kalangan usia, tiket terjangkau, pemandangan indah	['lengkap', 'cocok', 'kalangan', 'usia', 'terjangkau','pemandangan', 'indah']
Tempat Wisata Lumayan Bagus & Tiket Masuknya juga Murah	['lumayan', 'bagus', 'masuknya', 'murah']

4.3. Pelabelan Data



Gambar 1. Proses Pelabelan Data

Proses pelabelan data dilakukan menggunakan pendekatan hybrid pada Gambar 1 yang menggabungkan metode lexicon-based dengan algoritma Support Vector Machine (SVM). Lexicon sentimen disusun berdasarkan konteks ulasan destinasi wisata yang dikumpulkan dari berbagai ulasan nyata, kemudian divalidasi oleh dosen ahli Bahasa Indonesia dari Politeknik Negeri Malang yaitu Bapak Zulmi Faqihuddin Putera., S.Pd.,M.Pd untuk memastikan kesesuaian makna dan konteks dalam domain pariwisata.

Daftar lexicon terdiri dari tiga komponen utama: positive lexicon yang berisi kata-kata bernuansa positif seperti "bagus", "menyenangkan", "terjangkau", "ramah", "instagramable"; negative lexicon yang mencakup kata-kata negatif seperti "mahal", "kecewa", "ribet", "tidak ramah", "bising"; serta intensifier dan negation words yang digunakan untuk memodifikasi bobot sentimen, misalnya "sangat bagus" (lebih positif) atau "tidak bagus" (berubah menjadi negatif).

Algoritma hybrid bekerja dengan logika bertingkat. Pertama, sistem menganalisis skor lexicon berdasarkan kata positif dan negatif yang ditemukan dalam teks. Setiap kata positif atau negatif diberi bobot dasar 1 poin, lalu dimodifikasi jika didahului oleh intensifier (bobot dikalikan 2.0) atau didahului oleh negasi (bobot dibalik menjadi negatif). Kedua, jika perbedaan skor lexicon signifikan (memenuhi threshold 2.0), maka label ditentukan langsung dari analisis lexicon. Namun jika skor lexicon tidak konklusif, sistem menggunakan prediksi SVM dengan mempertimbangkan confidence score-nya. Jika confidence SVM ≥ 0.65, maka mengikuti prediksi SVM; jika confidence < 0.65, maka label berdasarkan dominasi skor lexicon.

Model SVM yang digunakan telah dilatih menggunakan dataset berlabel dengan konfigurasi Linear SVM untuk klasifikasi teks, menghasilkan prediksi label sentimen beserta confidence score yang menunjukkan tingkat keyakinan model. Pendekatan hybrid ini memberikan keuntungan ganda: memanfaatkan kemampuan generalisasi SVM terhadap struktur kalimat dan konteks, sekaligus menjaga akurasi semantik kata melalui validasi lexicon, sehingga menghasilkan akurasi pelabelan yang lebih tinggi dibandingkan metode tunggal.

4.4. TF-IDF

Setelah tahap preprocessing dan pelabelan, data teks diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam dokumen tertentu dibandingkan dengan frekuensi kemunculannya di seluruh koleksi dokumen. Kata-kata yang memiliki

frekuensi tinggi dalam satu dokumen tetapi jarang muncul di dokumen lain akan mendapat bobot TF-IDF yang tinggi, menandakan relevansi dan kekhususan kata tersebut terhadap dokumen. Implementasi TF-IDF pada Tabel 2 dilakukan menggunakan library Scikit-learn dengan TfidfVectorizer untuk menghasilkan matriks fitur yang akan digunakan dalam proses klasifikasi.

Tabel 2. Pengolahkan TF-IDF

Token	D1	D2	D3	D4	D5
Taman	0.0538	0	0	0	0
Bunga	0.0538	0	0	0	0
Luas	0.0538	0	0	0	0
Area	0.0538	0	0	0	0
Foto	0.0538	0	0	0	0
Bagus	0	0.0158	0.0317	0.0555	0
Cocok	0.0075	0.0069	0.0138	0	0.0138
Keluarga	0	0.0568	0.0568	0	0
Lumayan	0	0	0.0568	0.0995	0
Murah	0	0	0.0568	0.0995	0
Liburan	0	0.0284	0.0568	0	0
Lengkap	0	0	0	0	0.0999
indah	0	0	0	0	0.0999

4.5. Klasifikasi Random Forest

Dataset yang telah melalui tahap preprocessing dan representasi TF-IDF kemudian diklasifikasikan menggunakan algoritma Random Forest dengan 100 estimators. Model Random Forest dipilih karena kemampuannya dalam menangani dataset berskala besar dan mengurangi risiko overfitting melalui mekanisme ensemble learning. Evaluasi model dilakukan menggunakan 10-Fold Cross Validation untuk memastikan konsistensi dan reliabilitas hasil. Hasil evaluasi menunjukkan performa yang konsisten dengan akurasi rata-rata 87,60%, precision 87,68%, recall 87,93%, dan F1-score 87,55%. Performa tinggi pada Tabel 3 ini mengindikasikan bahwa model Random Forest efektif dalam mengklasifikasikan sentimen multi-platform dengan karakteristik data yang beragam.

Tabel 3. Hasil Klasifikasi Random Forest

Rata – Rata			
Accuracy	**Precision**	**Recall**	**F1_Score**
0.8760	0.8768	0.8793	0.8755

4.6. K-Fold Cross Validation

Evaluasi performa model Random Forest dilakukan menggunakan teknik 10-Fold Cross Validation untuk memastikan konsistensi dan reliabilitas hasil klasifikasi. Teknik ini membagi dataset menjadi 10 bagian (fold) yang berukuran sama secara acak, di mana pada setiap iterasi, 9 fold (90% data) digunakan sebagai data pelatihan dan 1 fold (10% data) digunakan sebagai data pengujian. Proses ini diulang sebanyak 10 kali sehingga setiap bagian data mendapat kesempatan tepat satu kali sebagai data uji, memastikan seluruh dataset digunakan baik untuk training maupun testing.

Hasil evaluasi menggunakan 10-Fold Cross Validation menunjukkan performa yang konsisten dengan variabilitas yang rendah antar fold. Akurasi model berkisar antara 83,02% hingga 91,82%, dengan nilai rata-rata 87,60% dan standar deviasi yang kecil, mengindikasikan stabilitas model. Precision rata-rata mencapai 87,68%, recall 87,93%, dan F1-score 87,55%, menunjukkan keseimbangan yang baik antara kemampuan model dalam mengidentifikasi sentimen dengan tepat dan kemampuannya menangkap seluruh data yang relevan.

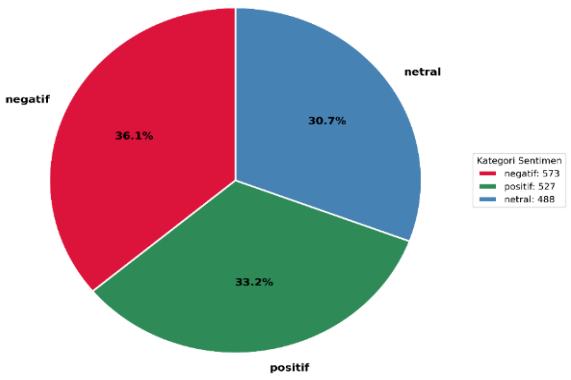
Penggunaan K-Fold Cross Validation memberikan beberapa keuntungan signifikan: pertama, mengurangi risiko overfitting dan underfitting karena model dilatih dan diuji pada berbagai subset data yang berbeda; kedua, memberikan estimasi performa yang lebih robust dan representatif dibandingkan single train-test split; ketiga, memaksimalkan penggunaan dataset yang tersedia karena setiap data digunakan baik untuk training maupun testing. Konsistensi hasil evaluasi di seluruh fold membuktikan bahwa model Random Forest memiliki kemampuan generalisasi yang baik terhadap karakteristik data sentimen multi-platform yang beragam.

Tabel 4. Hasil K-Fold Cross Validation

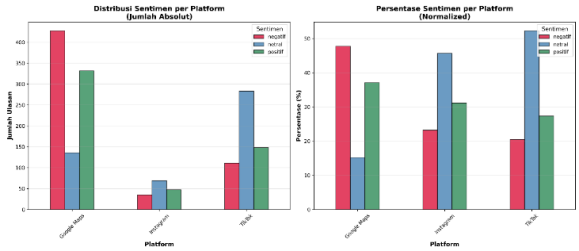
Fold	Accuracy	Precision	Recall	F1-Score
1	0.8868	0.8862	0.906	0.8866
2	0.8302	0.8302	0.8387	0.8305
3	0.8742	0.8753	0.8769	0.8715
4	0.8742	0.8708	0.8817	0.8743
5	0.8931	0.8893	0.8908	0.8900
6	0.8302	0.8342	0.8300	0.8301
7	0.8868	0.8871	0.8904	0.8878
8	0.9182	0.9198	0.9204	0.9181
9	0.8734	0.8746	0.8743	0.8726
10	0.8924	0.8928	0.8989	0.8937

4.7. Visualisasi

Berdasarkan Gambar 2 merupakan grafik hasil sentimen dari keseluruhan dataset yang ada. Dimana diperoleh total 36.1% (573 data) untuk sentimen negatif, kemudian 33.2% (527 data) untuk sentimen positif dan 30.7% (488) untuk sentimen netral. Dimana hal ini menunjukkan bahwa sentimen negatif paling dominan pada dataset ini.



Gambar 2. Distribusi Sentimen Keseluruhan



Gambar 3. Distribusi Sentimen Tiap Platform

Visualisasi pada Gambar 3 hasil penelitian disajikan dalam berbagai bentuk untuk memberikan insight yang komprehensif tentang distribusi dan karakteristik sentimen antar platform. Analisis distribusi sentimen mengungkapkan pola yang berbeda pada setiap platform digital. Google Maps menunjukkan volume data terbesar dengan total 1.391 ulasan yang terdiri dari 659 sentimen negatif, 518 sentimen positif, dan 214 sentimen netral. Instagram memiliki distribusi yang relatif seimbang dengan 48 sentimen positif, 35 sentimen negatif, dan 71 sentimen netral dari total 154 data. TikTok menampilkan dominasi sentimen netral dengan 298 data, diikuti 148 sentimen positif dan 111 sentimen negatif dari total 557 komentar.

Dalam perspektif persentase, karakteristik unik setiap platform menjadi lebih jelas. Google Maps didominasi sentimen negatif (47,4%) yang mengindikasikan fungsinya sebagai platform kritik dan evaluasi detail terkait fasilitas, pelayanan, dan aspek operasional destinasi wisata. Instagram menunjukkan keseimbangan dengan dominasi sentimen netral (46,1%), yang mencerminkan karakteristik platform visual dengan komentar deskriptif dan dokumentatif tanpa opini eksplisit. TikTok memperlihatkan dominasi sentimen netral (53,5%) yang signifikan, menunjukkan perannya sebagai media berbagi pengalaman visual yang menghibur dengan kecenderungan konten discovery yang implisit positif.

4.8. Visualisasi Wordcloud

Word cloud untuk setiap kategori sentimen memberikan wawasan mendalam tentang tema dominan dalam persepsi pengunjung.

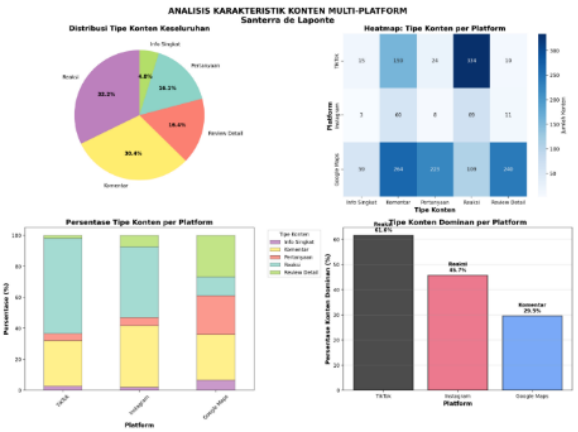
Gambar 4. Visualisasi Wordcloud

Sesuai Gambar 4 terdapat 3 sentimen, yang pertama sentimen positif (530 ulasan) didominasi kata-kata yang mencerminkan kepuasan pengunjung, dengan "bagus" sebagai kata terbesar, diikuti "seru", "foto", "spot", dan "area" yang mengindikasikan apresiasi terhadap daya tarik visual dan fasilitas fotogenik destinasi. Kata "anak" dan "keluarga" yang menonjol menunjukkan bahwa Santerra de Laponte dipersepsikan sebagai destinasi yang ramah keluarga. Munculnya kata "murah" dan "terjangkau" mengindikasikan persepsi positif terhadap value for money yang ditawarkan.

Sentimen negatif (571 ulasan) menampilkan keluhan operasional yang spesifik dan terstruktur. Kata "antre" mendominasi sebagai keluhan utama, menunjukkan masalah manajemen antrian yang perlu perhatian serius. Kata "weekend" yang menonjol mengindikasikan bahwa masalah kepadatan terutama terjadi pada akhir pekan. Keluhan "parkir", "toilet", dan "mahal" mencerminkan aspek infrastruktur dan pricing yang menjadi concern pengunjung. Kata "jalan", "macet", dan "panas" menunjukkan isu aksesibilitas dan kenyamanan lingkungan yang memerlukan solusi strategis dari pengelola.

Sentimen netral (487 ulasan) didominasi kata-kata informatif dan deskriptif tanpa muatan emosional. Kata "santerra" sebagai nama destinasi menjadi yang terbesar, diikuti "weekend", "minggu", dan "liburan" yang menunjukkan konteks waktu kunjungan. Kata "toilet", "parkir", dan "harga" muncul sebagai pertanyaan atau informasi faktual tanpa penilaian. Kata "anak" dan "keluarga" dalam konteks netral menunjukkan diskusi tentang kesesuaian destinasi untuk berbagai segmen pengunjung. Pola ini mengindikasikan bahwa sebagian besar konten netral berupa sharing informasi, dokumentasi kunjungan, atau pertanyaan dari calon pengunjung yang mencari informasi praktis.

4.9. Visualisasi Analisis Karakteristik Multi-Platform



Gambar 5. Visualisasi Analisis Karakteristik Multi-Platform

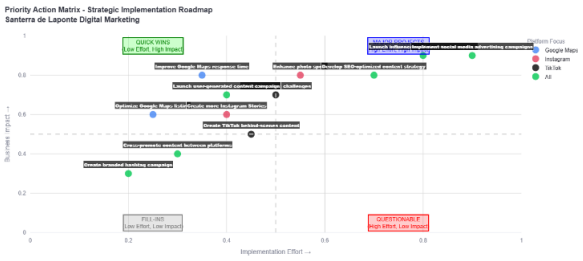
Analisis karakteristik konten multi-platform mengungkapkan pola distribusi yang menarik dalam sistem pemasaran digital Santerra de Laponte. Berdasarkan distribusi tipe konten keseluruhan, konten "Reaksi" mendominasi dengan 32,2% dari total konten, diikuti oleh "Komentar" sebesar 30,4%, sementara "Review Detail" dan "Pertanyaan" masing-masing mencapai 16,4% dan 15,9%. Dominasi konten reaksi dan komentar mengindikasikan bahwa Santerra

de Laponte telah berhasil menciptakan engagement tinggi yang memicu respons emosional dan interaksi aktif dari audiens.

Heatmap menunjukkan bahwa TikTok memiliki volume konten tertinggi dengan 349 reaksi dan 159 komentar, Google Maps dengan 264 komentar dan 240 review detail, sedangkan Instagram menunjukkan distribusi yang lebih seimbang namun dengan volume yang lebih rendah. Karakteristik platform individual menunjukkan perbedaan yang jelas dalam tipe konten yang dihasilkan. TikTok menunjukkan dominasi konten "Reaksi" sebesar 62,7%, menunjukkan efektivitas platform ini dalam menghasilkan konten viral yang memancing respons emosional masif dari pengguna. Instagram menampilkan karakteristik yang lebih seimbang dengan dominasi "Reaksi" sebesar 46,8%, diikuti proporsi komentar yang substantial, mengindikasikan efektivitas platform dalam membangun engagement melalui konten visual berkualitas.

Google Maps memiliki karakteristik unik dengan dominasi "Komentar" sebesar 29,5% dan proporsi review detail yang tinggi, sesuai dengan fungsinya sebagai platform pencarian informasi dan evaluasi destinasi wisata yang mendalam. Temuan ini menunjukkan bahwa TikTok terbukti paling efektif dalam membangun awareness dan viral engagement, Instagram menunjukkan efektivitas dalam membangun pertimbangan melalui konten visual, sedangkan Google Maps berperan kritis dalam tahap decision making. Kombinasi karakteristik ketiga platform ini menciptakan ekosistem pemasaran digital yang komprehensif dengan customer journey yang optimal.

4.10. Visualisasi Priority Action Matrix



Gambar 6. Visualisasi Priority Action Matrix

Visualisasi *Priority Action Matrix* pada Gambar 6 menggambarkan roadmap strategis dalam implementasi digital marketing untuk Santerra de Laponte. Matriks ini disusun berdasarkan dua dimensi utama, yaitu tingkat upaya implementasi (*effort*) dan dampak terhadap bisnis (*impact*), sehingga dapat digunakan sebagai dasar dalam menentukan prioritas strategi.

Strategi yang termasuk dalam kuadran *Quick Wins*, seperti peningkatan waktu respons pada Google Maps serta peluncuran kampanye konten buatan pengguna (*user-generated content*), menjadi prioritas utama karena memiliki tingkat upaya yang relatif rendah namun memberikan dampak yang tinggi terhadap bisnis. Sementara itu, strategi yang berada pada kuadran *Major Projects*, seperti kampanye iklan media sosial dan pengembangan strategi konten berbasis optimasi SEO, memerlukan upaya implementasi yang lebih besar, namun berpotensi memberikan dampak signifikan terhadap peningkatan visibilitas dan engagement.

Di sisi lain, beberapa strategi seperti pembuatan kampanye tagar bermerek (*branded hashtag campaign*) dan konten *behind-the-scenes* pada platform TikTok berada pada kategori dampak rendah, sehingga tidak menjadi prioritas dalam tahap awal implementasi.

Dengan demikian, *Priority Action Matrix* berperan sebagai alat bantu dalam pengambilan keputusan strategis, khususnya dalam mengidentifikasi dan memprioritaskan inisiatif yang paling efisien dan berdampak terhadap pengembangan pemasaran digital destinasi wisata.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem analisis sentimen multi-platform yang efektif untuk mengevaluasi persepsi publik terhadap destinasi wisata. Model Random Forest menunjukkan performa tinggi dengan akurasi 87,60% dalam mengklasifikasikan sentimen dari tiga platform digital. Setiap platform memiliki karakteristik sentimen yang berbeda, mengindikasikan perlunya strategi pemasaran digital yang disesuaikan dengan keunikan masing-masing platform.

Visualisasi dan analisis tren sentimen memberikan insight strategis yang dapat digunakan pengelola destinasi untuk optimalisasi pemasaran digital dengan visualisasi priority action matrix. Penelitian ini memberikan kontribusi pada pengembangan metodologi analisis sentimen multi-platform dalam domain pariwisata dan menyediakan framework untuk evaluasi efektivitas strategi pemasaran digital destinasi wisata.

Penelitian selanjutnya disarankan untuk menguji algoritma machine learning lain dan menggunakan teknik representasi teks yang lebih kaya seperti BERT embeddings untuk meningkatkan performa klasifikasi sentimen.

DAFTAR PUSTAKA

[1] H. Purba and Irwansyah, “User Generated Content dan Pemanfaatan Media Sosial Dalam Perkembangan Industri Pariwisata: Literature Review,” Dec. 2022.

[2] 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.

[3] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," *Journal of Information Systems and Informatics*, vol. 4, no. 1, Mar. 2022, [Online]. Available: http://journal-isi.org/index.php/isi

[4] A. J. Dahur, A. Wahyul Syafei, and T. Prahasto, "Analysis of Visitor Review Data Using Lexicon Based, Support Vector Machine, Random Forest in Determining the Priority Scale of Building Labuan Bajo Tourism Objects," in *E3S Web of Conferences*, EDP Sciences, Nov. 2023. doi: 10.1051/e3sconf/202344802043.

[5] B. Bayu Baskoro *et al.*, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," vol. 3, no. 2, pp. 21–029, 2021, doi: 10.20895/INISTA.V3I2.

[6] S. Juanita, K. Adiyarta, and M. Syafrullah, "Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naïve Bayes Model," Oct. 2022.

[7] A. D. Sanjaya, T. H. Pudjiantoro, A. K. Ningsih, and F. Renaldi, "Sentiment Analysis Of E-Wallets on Twitter social media With Naïve Bayes and Lexicon-Based Methods," Sep. 2022.

[8] P. Kumala Sari and R. Randy Suryono, "KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE," 2024.

[9] A. A. Firdaus, A. Id Hadiana, and A. K. Ningsih, "Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest," *R2J*, vol. 6, no. 5, 2024, doi: 10.38035/rrj.v6i5.

[10] M. Y. Aldean, Paradise, and N. A. S. Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," vol. 4, NO. 2, PP.064-072, May 2022, doi: 10.20895/INISTA.V4I2