

ANALISIS SENTIMEN TEMPAT WISATA DI KABUPATEN BANDUNG JAWA BARAT DENGAN VISUALISASI WEBSITE

1st Haris Tri Nugroho

School of Electrical Engineering
Telkom University
Bandung, Indonesia

haristri@student.telkomuniversity.ac.id

2nd Rita Purnamasari

School of Electrical Engineering
Telkom University
Bandung, Indonesia

ritapurnamasari@telkomuniversity.ac.id

3rd Efri Suhartono

School of Electrical Engineering
Telkom University
Bandung, Indonesia

esuhartono@telkomuniversity.ac.id

Abstrak — Sektor pariwisata Kabupaten Bandung adalah pilar ekonomi utama dengan berbagai objek wisata. Meskipun kunjungan wisatawan meningkat pasca pandemi, ulasan di Google Maps yang sering tidak terstruktur dan beragam kualitasnya menyulitkan pengelola destinasi dalam pengambilan keputusan berbasis data. Untuk mengatasi hal tersebut, penelitian ini merancang sistem analisis sentimen ulasan wisata secara otomatis berbasis machine learning. Sistem mengumpulkan data ulasan wisatawan dari Google Maps menggunakan teknik web scraping, kemudian melakukan *preprocessing* teks (pembersihan, tokenisasi, penghapusan *stopwords*, dan *stemming*), ekstraksi fitur menggunakan TF-IDF, serta klasifikasi sentimen ke dalam kategori positif, negatif, dan netral menggunakan tiga algoritma pembelajaran mesin (Naive Bayes, SVM, dan K-NN). Hasil analisis disajikan melalui antarmuka web interaktif berbasis Flask (backend) dan Next.js (frontend) yang menampilkan visualisasi distribusi sentimen, analisis komentar negatif, serta fitur uji sentimen untuk kalimat tunggal. Hasil evaluasi menunjukkan model SVM memberikan akurasi tertinggi (88%), diikuti Naive Bayes (85%) dan K-NN (67%). Sebagai contoh, analisis ulasan di destinasi Tangkuban Perahu (474 ulasan) menghasilkan distribusi sentimen 74% positif, 17% negatif, dan 9% netral. Temuan ini membuktikan bahwa penerapan analisis sentimen berbasis machine learning dapat memberikan wawasan objektif bagi pengelola pariwisata dalam memahami persepsi pengunjung, sehingga mendukung pengambilan kebijakan yang lebih tepat untuk meningkatkan kualitas layanan wisata di Kabupaten Bandung.

Kata kunci: analisis sentiment, google maps, klasifikasi teks, machine learning, pariwisata, visualisasi web

I. PENDAHULUAN

Kabupaten Bandung, Jawa Barat, memiliki potensi pariwisata yang besar dengan kekayaan objek wisata alam, budaya, dan kuliner yang menarik wisatawan domestik maupun mancanegara. Meskipun sektor pariwisata menggerakkan ekonomi daerah, peningkatan pengunjung membawa tantangan menjaga kualitas pengalaman. Ulasan digital para wisatawan di platform online seperti Google Maps kini sangat memengaruhi citra destinasi. Ulasan negatif yang tersebar di media digital dapat berdampak pada menurunnya minat wisatawan untuk berkunjung ke suatu objek wisata [1][2]. Di sisi lain, informasi yang tersedia dalam ulasan sering kali tidak lengkap dan tidak terstruktur dengan baik [3][4], sehingga wisatawan kesulitan menentukan destinasi yang sesuai minat dan pengelola kesulitan memahami persepsi pengunjung secara menyeluruh. Selain itu, belum ada standar evaluasi kualitas destinasi yang jelas

dan objektif, sehingga pengelola merasa sudah memberikan yang terbaik meskipun tidak ada acuan baku sebagai perbandingan [5][6]. Kondisi ini menunjukkan perlunya metode yang efektif untuk merangkum opini wisatawan yang tersebar di internet menjadi informasi yang bernilai bagi pemangku kepentingan pariwisata.

Berbagai upaya telah dilakukan untuk memahami kepuasan wisatawan, seperti survei kepuasan secara konvensional dan pemantauan komentar di media sosial. Namun, survei manual berbasis formulir memiliki partisipasi rendah dan lambat dalam pengolahan data, sementara pemantauan ulasan online secara sederhana (misal hanya menghitung jumlah komentar positif vs negatif) kurang mampu menangkap konteks dan nuansa bahasa dalam ulasan [7][8]. Analisis sentimen otomatis menggunakan teknik text mining pada ulasan wisata mengubah data teks tidak terstruktur menjadi wawasan objektif tentang persepsi pengunjung. Penelitian terdahulu telah menerapkan analisis sentimen berbasis machine learning untuk domain pariwisata. Somantri dan Dairoh [1] melakukan analisis sentimen ulasan wisata di Kota Tegal menggunakan *text mining*, sedangkan Ipmawati et al. [2] mengklasifikasikan sentimen tempat wisata berdasarkan ulasan Google Maps dengan algoritma SVM. Hasil-hasil studi tersebut menunjukkan bahwa pendekatan *data-driven* mampu memetakan opini publik secara lebih baik dibanding metode tradisional. Selain itu, tinjauan sistematis oleh Manosso dan Ruiz [3] mengungkapkan bahwa sentimen positif wisatawan di media online berkorelasi kuat dengan peningkatan kunjungan wisatawan, menegaskan pentingnya pemantauan sentimen dalam industri pariwisata. Akan tetapi, studi-studi sebelumnya terbatas pada analisis dan belum terintegrasi ke dalam sebuah platform yang dapat digunakan langsung oleh pengelola wisata secara interaktif.

Penelitian ini bertujuan membangun sistem analisis sentimen ulasan wisata di Kabupaten Bandung yang otomatis dan terpadu, dengan mengumpulkan data dari Google Maps, memproses teks, dan mengklasifikasikan sentimen menggunakan machine learning. Sistem ini juga menyediakan website interaktif untuk memvisualisasikan hasil analisis secara *real-time*, guna membantu pemerintah dan pengelola destinasi dalam memantau umpan balik pengunjung, mengidentifikasi area perbaikan, serta mendukung pengambilan keputusan berbasis data untuk meningkatkan kualitas pariwisata.

II. KAJIAN TEORI

A. Analisis Sentimen dan NLP

Analisis sentimen adalah bagian dari bidang *Natural Language Processing* (NLP) yang berfokus pada penggalian opini, emosi, atau kecenderungan sentimen dari teks[9]. Dengan teknik ini, teks ulasan pengguna dapat diklasifikasikan ke kategori seperti positif, negatif, atau netral secara otomatis. Penerapan analisis sentimen di sektor pariwisata memungkinkan eksplorasi persepsi wisatawan terhadap destinasi melalui ulasan dan komentar yang mereka tinggalkan di *platform* digital[10]. Ulasan pada layanan seperti Google Maps telah terbukti menyediakan wawasan berharga bagi pengelola destinasi untuk memahami preferensi dan tingkat kepuasan pengunjung[11]. Penelitian Rutba dan Pramana [10] menunjukkan bahwa peningkatan sentimen positif wisatawan di media sosial berhubungan erat dengan kenaikan angka kunjungan wisatawan mancanegara pascapandemi. Hal ini menegaskan pentingnya analisis sentimen sebagai indikator kinerja destinasi wisata. Secara umum, analisis sentimen teks dilakukan dengan kombinasi teknik text mining dan machine learning, di mana model dilatih untuk mengenali pola bahasa yang menunjukkan sentimen tertentu [9].

B. Preprocessing Teks

Sebelum data teks ulasan dianalisis sentimennya, perlu dilakukan preprocessing agar data lebih bersih dan siap diolah oleh algoritma. Tahapan *preprocessing* utama dalam penelitian ini meliputi:

1. Pembersihan teks (*text cleaning*): Menghapus karakter-karakter khusus, tanda baca, simbol, serta melakukan *case folding* (mengubah semua huruf menjadi kecil) untuk memastikan konsistensi format teks. Langkah ini membersihkan noise pada data seperti emotikon atau tanda baca yang tidak diperlukan dalam analisis.
2. Tokenisasi: Memecah kalimat atau paragraf ulasan menjadi unit-unit kata (token) individual. Setiap kata atau frasa diperlakukan sebagai token yang akan dianalisis lebih lanjut. Proses ini biasanya menggunakan pemisah seperti spasi atau tanda baca untuk mengidentifikasi batas token.
3. Penghapusan *stopwords*: Membuang kata-kata umum yang sering muncul namun tidak memiliki makna penting, misalnya "yang", "dan", "di", "untuk" dalam Bahasa Indonesia. Daftar *stopwords* Bahasa Indonesia digunakan sebagai acuan pada tahap ini, sehingga hanya kata-kata bermakna yang dipertahankan untuk analisis.
4. *Stemming*: Mengubah kata berimbuhan atau varian kata ke bentuk dasarnya (akar kata). Teknik *stemming* (misal dengan algoritma Nazief & Adriani atau pustaka Sastrawi) menghilangkan imbuhan seperti prefix/suffix, sehingga kata "mengunjungi" menjadi "kunjung". Hal ini mengurangi keanekaragaman bentuk kata yang sebenarnya memiliki arti dasar sama, meningkatkan akurasi dalam pengelompokan kata.

Tahapan di atas memastikan data ulasan siap untuk tahap ekstraksi fitur dan pemodelan, dengan menghilangkan elemen-elemen yang tidak relevan dan menyamakan format teks. Hasil dari preprocessing adalah kumpulan token kata dasar yang mewakili isi ulasan wisatawan secara konsisten.

C. Ekstraksi Fitur dengan TF-IDF

Setelah teks bersih dan ditokenisasi, langkah selanjutnya adalah ekstraksi fitur untuk mengubah data teks menjadi representasi numerik yang dapat diolah oleh algoritma *machine learning*. Penelitian ini menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF) untuk pembobotan kata. TF-IDF terbukti efektif merepresentasikan pentingnya sebuah kata dalam korpus teks secara matematis. Metode ini memberikan bobot tinggi pada kata yang sering muncul dalam sebuah dokumen (ulasan) tetapi jarang muncul di dokumen lain, sehingga kata tersebut dianggap memiliki informasi penting. Sebagai contoh, jika kata "murah" sangat sering muncul pada ulasan destinasi A dibanding destinasi lain, TF-IDF akan memberi bobot tinggi pada kata "murah" untuk destinasi A, menunjukkan bahwa aspek murah (harga) menonjol di sana.

Parameter umum TF-IDF antara lain jumlah fitur maksimum (misal mengambil 1.000 kata dengan bobot tertinggi) dan pilihan *n-gram* (hanya unigram, atau termasuk bi-gram, tri-gram) yang digunakan dalam membentuk vektor fitur. Dalam penelitian ini digunakan TF-IDF unigram dengan mengambil seluruh kata unik yang ada di korpus ulasan sebagai fitur. Keunggulan TF-IDF telah ditunjukkan di mana model sentiment analysis dengan TF-IDF mencapai akurasi 96–99% pada berbagai *dataset benchmark*, lebih baik dibanding beberapa teknik ekstraksi lain. Selain TF-IDF, metode lain seperti *Bag-of-Words*, *Word2Vec*, atau *GloVe* juga umum digunakan sebagai pembanding, tetapi TF-IDF dipilih karena kesederhanaan implementasi dan kinerjanya yang baik untuk teks ulasan yang tidak terlalu panjang.

D. Algoritma Klasifikasi Sentimen Menggunakan SVM

Penelitian ini menerapkan algoritma Support Vector Machine (SVM) untuk tugas klasifikasi sentimen. Algoritma ini dipilih karena performanya yang baik untuk menganalisis teks.

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang kuat untuk data berdimensi tinggi dan sering digunakan dalam analisis sentimen karena kemampuannya menghasilkan akurasi yang konsisten. SVM berupaya mencari *hyperplane* (bidang pemisah) optimal yang memaksimalkan margin antara kelas positif dan negatif dalam ruang vektor fitur. Penelitian terkini menunjukkan SVM mampu mencapai akurasi 85% pada dataset ulasan film IMDb dan mengungguli metode klasik seperti Naive Bayes. Dalam konteks pariwisata, SVM terbukti efektif mengklasifikasikan sentimen ulasan wisatawan terhadap kebijakan pembangunan destinasi. Parameter utama SVM meliputi pilihan kernel (linear, RBF, polynomial), parameter regularisasi C, dan *kernel coefficient* (gamma untuk RBF) yang perlu dituning agar model optimal.

Parameter utama SVM meliputi pilihan *kernel* (linear, RBF, polynomial), parameter regularisasi C, dan *kernel coefficient* (gamma untuk RBF) yang perlu dituning agar model mencapai performa optimal.

E. Teknologi Web: Flask dan Next.js

Sistem yang dikembangkan dilengkapi dengan aplikasi web untuk visualisasi hasil, sehingga melibatkan teknologi frontend dan backend. Pada sisi backend, penelitian ini menggunakan kerangka kerja Flask berbasis Python. Flask adalah micro-framework yang ringan untuk pembangunan web dan REST API, cocok untuk prototipe analisis data karena strukturnya sederhana dan fleksibel. Flask menangani

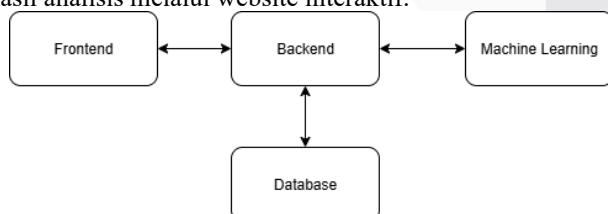
logika bisnis di server, termasuk proses preprocessing teks, inferensi model machine learning, serta penyediaan endpoints API yang akan diakses oleh frontend. Kelebihan Flask adalah kemudahan dalam mengintegrasikan berbagai pustaka Python (scikit-learn, pandas, dll.) serta dukungan komunitas yang luas. Framework ini telah banyak digunakan dalam penelitian untuk cepat membangun layanan analitik karena mendukung routing dan manajemen request-response yang efisien.

Sisi frontend dibangun menggunakan Next.js, yakni framework JavaScript berbasis React yang mendukung server-side rendering (SSR) dan static site generation. Next.js dipilih untuk menghasilkan antarmuka pengguna yang interaktif, responsif, dan optimal dari segi performa. Dengan SSR, halaman web dapat dimuat lebih cepat dan ramah SEO, yang penting untuk aplikasi web publik. Next.js memudahkan pengelolaan routing, pemisahan komponen UI, dan integrasi dengan backend melalui REST API. Dalam sistem ini, Next.js digunakan untuk menampilkan visualisasi hasil analisis sentimen secara dinamis. Hasil klasifikasi sentimen yang dikirim oleh API Flask divisualisasikan dalam bentuk grafik dan tabel interaktif di halaman web Next.js. Kombinasi Flask-Next.js ini memberikan arsitektur client-server yang terpisah (decoupled), di mana Flask berperan sebagai penyedia data dan Next.js sebagai penyaji antarmuka.

Untuk penyimpanan data sementara, sistem ini menggunakan format CSV dan memori, tidak memakai database khusus mengingat data ulasan dapat langsung diproses. Proses deployment website dilakukan pada platform Railway dengan URL publik, sehingga stakeholder dapat mengaksesnya melalui browser tanpa perlu menyiapkan lingkungan eksekusi Python secara mandiri.

III. METODE

Analisis sentimen adalah proses komputasi yang digunakan untuk menentukan kategori emosi dalam teks, dengan tahapan mulai dari pra-pemrosesan hingga klasifikasi, yang dalam penelitian ini diterapkan menggunakan *Natural Language Processing* (NLP) dan algoritma machine learning[12]. Sistem analisis sentimen ini dirancang dengan beberapa komponen utama yang saling terintegrasi, sebagaimana terlihat pada **Gambar 3.1** desain sistem website. Arsitektur sistem terdiri dari tahap pengumpulan data, pra-pemrosesan, klasifikasi sentimen menggunakan algoritma machine learning, serta visualisasi hasil analisis melalui website interaktif.



Gambar 3.1 Desain Sistem Website

A. Dataset dan Akuisisi Data

Data yang digunakan dalam penelitian ini berupa ulasan tekstual wisatawan yang diambil dari platform Google Maps untuk berbagai destinasi wisata di Kabupaten Bandung. Pengambilan data dilakukan dengan teknik *web scraping* secara otomatis agar dapat mengumpulkan ulasan dalam

jumlah besar dengan cepat. Alur pengumpulan data sebagai berikut:

- *Web Scraping Google Maps*: Peneliti memanfaatkan tool *Instant Data Scraper* (Chrome extension) dan skrip kustom berbasis Python/Node.js untuk mengekstraksi teks ulasan beserta meta-data (nama pengulas, tanggal ulasan, dan rating bintang) dari halaman Google Maps setiap destinasi wisata. Teknik scraping dipilih karena Google Maps tidak menyediakan API publik untuk ulasan. Untuk menghindari limitasi, scraping dilakukan secara bertahap per destinasi dengan jeda waktu tertentu.
- *Batasan Waktu dan Lokasi*: Dataset dibatasi pada ulasan yang dipublikasikan dalam periode satu tahun terakhir (agar data relevan dan mutakhir) dan berfokus pada destinasi di wilayah Kabupaten Bandung saja. Ulasan dari destinasi di luar area ini tidak diikutsertakan. Selain itu, hanya destinasi yang memiliki jumlah ulasan cukup (puluhan hingga ratusan) yang diambil, guna memastikan ketersediaan data yang representatif.
- *Kuantitas Data*: Secara total terkumpul ribuan ulasan wisatawan dari belasan objek wisata populer di Kabupaten Bandung. Setiap ulasan terdiri dari teks komentar dan rating bintang 1–5. Data mentah ini kemudian disimpan dalam format CSV untuk diproses lebih lanjut. Sebagai gambaran, salah satu destinasi terkenal (Gunung Tangkuban Perahu) memiliki 474 ulasan yang berhasil di-scrape, sementara destinasi lain seperti kawah atau situ di wilayah Pangalengan dan Ciwidey memiliki ratusan ulasan masing-masing.

B. Pelabelan Sentimen

Karena data ulasan yang diperoleh tidak memiliki label sentimen (positif/negatif/netral) secara langsung, perlu dilakukan pelabelan data agar dapat digunakan untuk melatih dan menguji model *supervised learning*. Penelitian ini menerapkan metode pelabelan otomatis berbasis kamus sentimen (lexicon) Bahasa Indonesia. Tahapan pelabelan:

1. *Kamus Sentimen*: Dibangun kamus kata positif dan negatif beserta skor polaritasnya. Kamus ini diadaptasi dari lexicon publik (misal *sentimenanalysis.id* atau hasil riset terdahulu) yang berisi ratusan kata positif (contoh: "indah", "murah", "ramah") dan negatif ("mahal", "macet", "kotor") dengan bobot +1 untuk kata positif dan -1 untuk kata negatif.
2. *Skoring Ulasan*: Setiap teks ulasan yang sudah melalui preprocessing dihitung skor sentimennya dengan menjumlahkan skor semua kata yang ditemukan dalam kamus. Jika sebuah ulasan mengandung 3 kata positif dan 1 kata negatif, maka skor total = $(+13) + (-11) = +2$.
3. *Menentukan Label*: Aturan penentuan label berdasarkan skor polaritas: jika skor > 0 , maka ulasan dilabel positif; jika skor < 0 dilabel negatif; dan jika skor $= 0$ dilabel netral. Misalnya, ulasan "*tempatnyanya sangat indah dan bersih*" mungkin mendapat skor +2 (positif), sedangkan "*harga tiket mahal dan parkirannya sempit*" bisa mendapat skor -2 (negatif). Ulasan dengan skor nol (jumlah kata positif dan negatif seimbang, atau tidak mengandung kata bersentimen) dianggap netral.

Metode lexicon-based labeling ini dipilih untuk efisiensi, mengingat pelabelan manual oleh manusia terhadap ribuan

data akan memakan waktu lama. Meskipun akurasi pelabelan otomatis ini tidak sempurna, pendekatan tersebut cukup untuk menyediakan *training set* bagi model ML tahap awal. Seluruh proses pelabelan dijalankan dalam environment Google Colab menggunakan Python, sehingga dapat ditangani secara cepat dan memungkinkan peninjauan manual secara sampling untuk memastikan kualitas label. Pada tahap akhir, diperoleh dataset terlabel dengan proporsi kelas $\pm 60\%$ positif, $\pm 20\%$ negatif, dan $\pm 20\%$ netral (setiap destinasi bervariasi proporsinya). Dataset ini kemudian dibagi menjadi data latih (training) dan uji (testing) dengan perbandingan 80:20 secara acak.

C. Pelatihan Model Machine Learning

Langkah berikutnya adalah melatih model klasifikasi sentimen menggunakan data latih terlabel. Algoritma SVM diterapkan sesuai penjelasan di Kajian Teori. Proses pelatihan dilakukan dengan bantuan pustaka Scikit-learn:

- SVM: Menggunakan kelas SVC (Support Vector Classifier) dengan kernel linear. Dipilih kernel linear karena teks cenderung dapat dipisahkan secara linier dalam ruang dimensi tinggi setelah transformasi TF-IDF. Parameter C (regularisasi) dioptimalkan melalui percobaan grid search pada nilai {0.1, 1, 10}, dan terbaik diperoleh C=1 tanpa menggunakan kernel non-linear (RBF/poly) karena linear sudah memberikan hasil akurat dan lebih efisien.

Selama pelatihan, dilakukan validasi silang (*cross-validation*) 5-fold pada data latih untuk memastikan model tidak overfitting dan parameter yang dipilih cukup general. Proses komputasi dilakukan di Google Colab dengan bantuan GPU untuk mempercepat terutama pada model SVM yang trainingnya lebih lama dibanding NB.

D. Evaluasi dan Pengujian Model

Setelah model dilatih, kinerja ketiga model dievaluasi menggunakan data uji (20% yang disisihkan). Metrik evaluasi utama adalah akurasi, yaitu persentase prediksi yang benar terhadap seluruh data uji. Selain itu, dihitung juga *precision*, *recall*, dan *F1-score* untuk masing-masing kelas sentimen, guna memperoleh gambaran yang lebih detail mengenai performa model. Confusion matrix dibuat untuk melihat distribusi prediksi vs aktual, sehingga dapat diidentifikasi jenis kesalahan klasifikasi yang terjadi (misal banyak ulasan netral diprediksi positif, dll.).

Berdasarkan hasil pengujian, model SVM menunjukkan akurasi tertinggi dalam mengklasifikasikan sentimen ulasan wisata, diikuti oleh Naive Bayes dan terakhir K-NN. Hasil lengkap evaluasi ditunjukkan pada Tabel 1. Tabel tersebut menyajikan metrik akurasi, *precision*, *recall*, dan *F1-score* rata-rata untuk kelas positif dan negatif (karena kelas netral berada di antara keduanya):

Tabel 3.1 Hasil Evaluasi Performa Model

Algorithm	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.85	0.82	0.85	0.80
SVM	0.88	0.90	0.88	0.88
K-NN	0.67	0.86	0.67	0.71

Dari Tabel 3.1 dapat dilihat bahwa SVM sedikit lebih unggul daripada NB, terutama dalam menjaga keseimbangan *precision-recall* untuk kedua kelas. K-NN memiliki performa paling rendah; hal ini sejalan dengan ekspektasi karena K-NN kurang mumpuni untuk data teks berdimensi tinggi. Selisih akurasi yang cukup besar menunjukkan model berbasis probabilistik dan hyperplane lebih efektif menangkap pola sentimen dibanding pendekatan berbasis tetangga terdekat.

Setelah evaluasi, model SVM dipilih untuk diimplementasikan dalam sistem web, mengingat akurasinya paling tinggi. Model NB juga disediakan sebagai pembandingan di backend, sedangkan model K-NN tidak digunakan lebih lanjut karena kinerjanya kurang memadai (akurasi di bawah ambang 80% yang diharapkan).

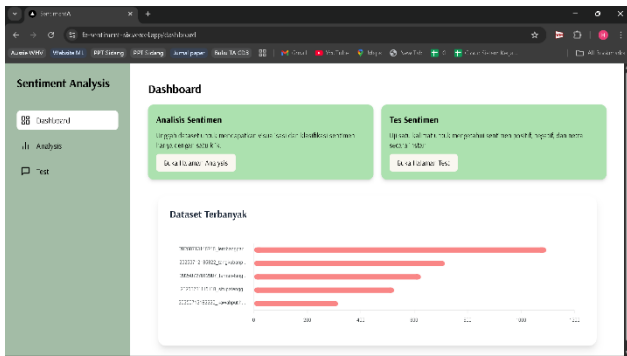
E. Arsitektur Sistem dan Antarmuka Web

Sistem yang dibangun memiliki arsitektur *client-server* tiga komponen: Frontend (Next.js), Backend (Flask), dan Modul Machine Learning. Gambar 1 menunjukkan diagram arsitektur sistem secara ringkas. Frontend Next.js bertugas menampilkan halaman antarmuka ke pengguna dan mengambil data hasil analisis melalui permintaan API ke backend. Backend Flask menjadi penghubung antara frontend dan modul machine learning, menangani permintaan (request) dari frontend, menjalankan proses analisis (preprocessing + prediksi model), lalu mengembalikan hasil dalam format JSON. Modul machine learning di sini merujuk pada model klasifikasi (SVM/NB) yang telah dilatih dan di-load di memori server Flask untuk melakukan prediksi saat ada input baru.

Alur operasi sistem sebagai berikut: pengguna mengakses website melalui browser, melakukan login terlebih dahulu (untuk memastikan hanya pengguna terotorisasi yang dapat mengakses data). Setelah login, pengguna masuk ke halaman utama (dashboard) aplikasi. Pada dashboard, terdapat dua fitur utama: Analisis Sentimen Dataset dan Uji Kalimat Tunggal.

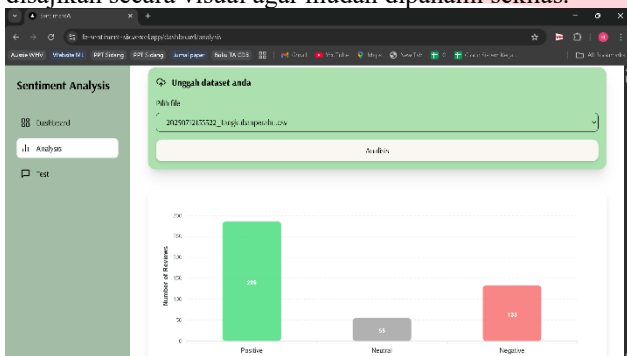
- Fitur Analisis Sentimen Dataset memungkinkan pengguna mengunggah file CSV ulasan Google Maps yang diproses di backend Flask untuk pra-pemrosesan dan prediksi sentimen. Hasil distribusi sentimen dan statistik komentar negatif berdasarkan kategori serta tren bulanan dikirim dalam format JSON ke frontend Next.js untuk visualisasi interaktif.
- Diagram pie atau bar untuk komposisi persentase sentimen positif, netral, negatif.
- Grafik batang untuk 5 kategori keluhan terbanyak dari komentar negatif (contoh: kebersihan, harga mahal, kemacetan, dll.), disertai trend per bulannya. Hal ini membantu mengidentifikasi isu utama yang sering dikeluhkan wisatawan.
- Tabel ringkasan yang menampilkan beberapa contoh ulasan negatif beserta kategorinya, sebagai ilustrasi konkret.
- Uji Sentimen Kalimat: Fitur ini menampilkan tabel contoh ulasan negatif beserta kategorinya sebagai ilustrasi. Selain itu, terdapat fitur uji sentimen kalimat tunggal di website, dimana pengguna memasukkan kalimat yang kemudian diproses untuk prediksi sentimen secara *real-time* menggunakan model yang sama, dengan hasil langsung ditampilkan di layar.

Antarmuka web didesain agar *user-friendly* dan informatif. Halaman dashboard menampilkan petunjuk singkat dan statistik umum.

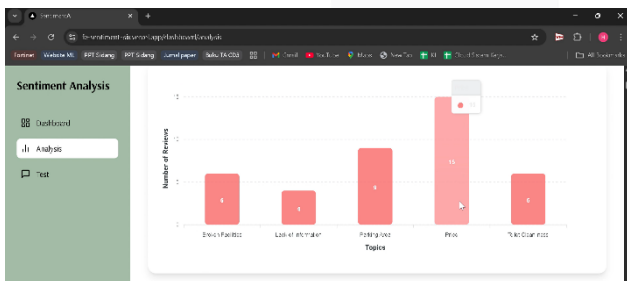


Gambar 3.2 Halaman Dashboard Website

Setelah pengguna mengunggah dataset dan analisis selesai, grafik-grafik akan muncul di halaman *Analysis*. Pengguna dapat melihat misalnya: "Dari 300 ulasan pada Destinasi X: 65% positif, 25% netral, 10% negatif". Grafik selanjutnya mungkin menunjukkan bahwa dari ulasan negatif, 40% mengeluhkan harga, 30% fasilitas, 20% akses jalan, dan 10% kebersihan, dengan tren komentar negatif paling banyak di bulan liburan tertentu. Semua informasi ini disajikan secara visual agar mudah dipahami sekilas.



Gambar 3.3 Halaman Analysis Website



Gambar 3.4 Halaman Analysis Website 2

Sistem web ini telah di-deploy sehingga dapat diakses melalui internet. Mekanisme autentikasi login sederhana diterapkan untuk mengamankan data (menghindari sembarang pengguna mengunggah data tidak relevan). Secara keseluruhan, integrasi frontend-backend-modul ML ini memungkinkan automasi penuh: mulai dari pengambilan data ulasan, analisis sentimen, hingga visualisasi hasil dalam satu platform terpadu yang dapat digunakan oleh pengelola pariwisata Kabupaten Bandung.

F. Classification with SVM

Salah satu tugas klasifikasi sentimen dilakukan menggunakan Support Vector Machine (SVM), yaitu algoritma *supervised learning* yang banyak digunakan untuk klasifikasi teks karena ketahanannya pada ruang fitur berdimensi tinggi. SVM bekerja dengan cara mencari *hyperplane* optimal yang memisahkan titik data dari kelas

yang berbeda dengan margin maksimum. Dalam formulasi standarnya (*hard-margin*), permasalahan optimasi didefinisikan sebagai berikut:

$$f(x) = \text{sign}(w \cdot x + b) \quad (2)$$

Dengan keterangan:

- w adalah *weight vector* (vektor bobot)
- x adalah *feature vector* (vektor fitur)
- b adalah *bias*

IV. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian analisis sentimen beserta interpretasinya terkait berbagai aspek layanan pariwisata. Analisis dilakukan pada kumpulan ulasan pengguna yang diperoleh dari sumber-sumber terkait pariwisata, dengan fokus pada Tangkuban Perahu sebagai studi kasus. Ulasan-ulasan tersebut dikategorikan ke dalam beberapa sektor yang telah ditentukan sebelumnya. Sentimen dari setiap ulasan diklasifikasikan ke dalam tiga kategori yaitu negatif, netral, dan positif. Dari penelitian sebelumnya yaitu [12] PLN dapat memahami ulasan konsumen secara lebih cepat dan akurat, sehingga dapat menanggapi umpan balik pengguna dengan lebih baik dan meningkatkan mutu layanan aplikasi PLN Mobile.

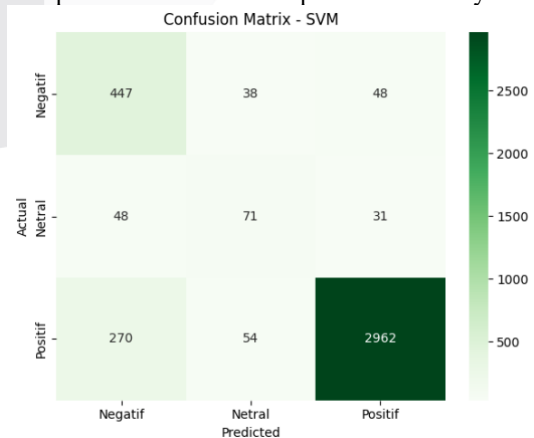
A. SVM Evaluation Metrics

Pada bagian ini, dilakukan evaluasi kinerja algoritma SVM berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score*.

Evaluasi Model SVM:				
Akurasi Model SVM: 0.88				
Model Precision: 0.90				
Model Recall : 0.88				
Model F1-Score : 0.88				
	precision	recall	f1-score	support
Negatif	0.58	0.84	0.69	533
Netral	0.44	0.47	0.45	150
Positif	0.97	0.90	0.94	3286
accuracy			0.88	3969
macro avg	0.66	0.74	0.69	3969
weighted avg	0.90	0.88	0.88	3969

Gambar 4.1 Evaluasi Model SVM

Visualisasi hasil klasifikasi ditampilkan dalam bentuk *confusion matrix* pada Gambar 3, yang menggambarkan distribusi prediksi model terhadap label sebenarnya.



Gambar 4.2 Confusion Matrix SVM

V. KESIMPULAN

Penelitian ini telah berhasil merancang dan mengimplementasikan sistem analisis sentimen ulasan wisata

berbasis *machine learning* untuk destinasi wisata di Kabupaten Bandung. Data ribuan ulasan Google Maps dikumpulkan dan diproses melalui tahapan *text preprocessing*, kemudian diklasifikasikan ke dalam sentimen positif, netral, atau negatif menggunakan tiga algoritma (Naive Bayes, SVM, K-NN). Dari evaluasi model, algoritma SVM memberikan hasil paling akurat dengan akurasi 88%, sedikit mengungguli Naive Bayes (85%), sementara K-NN kurang efektif (67%). Sistem ini dilengkapi dengan antarmuka website interaktif berbasis Flask-Next.js yang menampilkan visualisasi distribusi sentimen dan analisis mendalam atas ulasan negatif, serta menyediakan fitur uji kalimat untuk prediksi sentimen secara instan.

Melalui studi kasus data nyata, sistem mampu mengungkap proporsi sentimen wisatawan pada berbagai objek wisata dan mengidentifikasi aspek-aspek yang sering dikeluhkan (seperti harga, fasilitas, akses, kebersihan). Informasi ini memberikan *insight* berharga bagi pengelola destinasi dan dinas pariwisata setempat untuk mengambil tindakan perbaikan yang tepat sasaran. Dengan kata lain, analisis sentimen otomatis dapat berperan sebagai “*early warning system*” terhadap isu layanan, sekaligus alat evaluasi objektif kinerja destinasi wisata dari perspektif pengunjung. Implementasi sistem ini diharapkan membantu terwujudnya pengambilan keputusan berbasis data (*data-driven decision making*) dalam pengelolaan pariwisata Kabupaten Bandung, sehingga program pengembangan ke depan dapat lebih responsif, efisien, dan sesuai ekspektasi wisatawan.

Sebagai saran, pengembangan lebih lanjut dapat dilakukan dengan memperluas cakupan data (misalnya ulasan dari platform lain), menerapkan teknik *sentiment analysis* yang lebih canggih seperti BERT untuk meningkatkan akurasi, serta menambahkan analisis aspek (*aspect-based sentiment analysis*) agar dapat menggali opini wisatawan pada aspek spesifik dengan lebih detail. Integrasi sistem dengan dashboard pemerintah atau aplikasi mobile juga bisa dipertimbangkan untuk mempermudah akses bagi para pemangku kepentingan. Dengan demikian, inovasi ini dapat terus disempurnakan untuk mendukung pariwisata yang berkelanjutan dan berbasis teknologi di masa mendatang.

REFERENSI

- [1] O. Somantri and D. Dairoh, “Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining,” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.
- [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, 2024, doi: 10.57152/malcom.v4i1.1066.
- [3] F. C. Manosso and T. C. D. Ruiz, “Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review,” *Journal of Tourism, Heritage and Services Marketing*, vol. 7, no. 2, pp. 16–27, 2021, doi: 10.5281/zenodo.5548426.
- [4] W. Rafdinal, “Is smart tourism technology important in predicting visiting tourism destination? Lessons from West Java, Indonesia,” *Journal of Tourism Sustainability*, vol. 1, no. 2, pp. 102–115, 2021, doi: 10.35313/jtos.v1i2.20.
- [5] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, “Analisis Sentimen Review Wisatawan pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine,” *Jurnal Informatika Teknologi dan Sains*, vol. 5, no. 1, pp. 133–140, 2023.
- [6] A. Taneja, “Sentiment Analysis Using Machine Learning : a Comprehensive Sentiment Analysis Using Machine Learning : a Comprehensive Review,” vol. 3, no. July, pp. 181–189, 2024.
- [7] S. A. Rutba and S. Pramana, “Aspect-based Sentiment Analysis and Topic Modelling of International Media on Indonesia Tourism Sector Recovery,” 2025, doi: 10.36256/ijtl.v6i1.502.
- [8] N. A. Semary, W. Ahmed, K. Amin, P. Plawiak, and M. Hammad, “Enhancing machine learning-based sentiment analysis through feature extraction techniques,” *PLoS One*, vol. 19, no. 2 February, 2024, doi: 10.1371/journal.pone.0294968.
- [9] N. Wulandari, Y. Cahyana, and H. H. Handayani, “Sentiment Analysis on the Relocation of the National Capital (IKN) on Social Media X Using Naive Bayes and K-Nearest Neighbor (KNN) Methods,” vol. 9, no. 3, pp. 724–731, 2025.
- [10] R. Kurniawan, H. O. L. Wijaya, and R. P. Aprisusanti, “Sentiment Analysis of Google Play Store User Reviews on Digital Population Identity App Using K-Nearest Neighbors,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 2, pp. 170–178, 2024, doi: 10.32736/sisfokom.v13i2.2071.
- [11] D. J. Evan and P. O. N. Saian, “Implementasi Python Framework Flask Pada Modul Transfer Out Toko Di Pt Xyz,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1121–1131, 2023, doi: 10.29100/jipi.v8i4.4020.
- [12] “Analisis Sentimen Menggunakan Algoritma Naive Bayes Classifier Pada Ulasan Aplikasi PLN Mobile di Google Play Store”.