

KLASIFIKASI SENTIMEN ULASAN PENGUNJUNG LAWANG SEWU DI GOOGLE MAPS MENGGUNAKAN K-NEAREST NEIGHBOR

Joko Bambang Wiranto, Eri Zuliarso

Teknik Informatika, Universitas STIKUBANK Unisbank Semarang
Jalan Raya Tri Lomba Juang 1 Semarang, Indonesia
jokobambangwiranto@mhs.unisbank.ac.id

ABSTRAK

Analisis sentimen digunakan untuk mengidentifikasi opini dalam ulasan daring. Lawang Sewu sebagai destinasi wisata bersejarah di Semarang memiliki banyak ulasan di Google Maps yang mencerminkan pengalaman pengunjung. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan menggunakan K-Nearest Neighbor (KNN). Data dikumpulkan dengan web scraping menggunakan SerpAPI dan diproses melalui text preprocessing seperti cleaning, case folding, stemming, dan stopword removal. Label sentimen ditentukan berdasarkan rating, yaitu positif, negatif, dan netral. Model KNN diterapkan untuk klasifikasi dan dievaluasi dengan confusion matrix. Hasil penelitian menunjukkan bahwa 52,8% ulasan bersentimen positif, 27,4% negatif, dan 19,8% netral. Model mencapai akurasi 64% dengan F1-score rata-rata 0,60, tetapi mengalami kendala dalam membedakan sentimen netral. Oleh karena itu, disarankan penggunaan teknik feature weighting seperti TF-IDF atau eksplorasi algoritma lain seperti Support Vector Machine (SVM) untuk meningkatkan akurasi klasifikasi.

Kata kunci : Analisis Sentimen, K-Nearest Neighbor, Google Maps, Lawang Sewu, Text Mining

1. PENDAHULUAN

Teknologi digital telah memudahkan wisatawan dalam berbagi pengalaman mereka melalui platform ulasan daring seperti Google Maps. Ulasan ini berisi opini subjektif yang dapat digunakan untuk menilai kepuasan pengunjung terhadap suatu destinasi wisata [4]. Lawang Sewu, sebagai salah satu ikon wisata bersejarah di Semarang, memiliki banyak ulasan yang mencerminkan pengalaman pengunjung. Namun, informasi dalam ulasan tersebut masih belum terstruktur, sehingga diperlukan metode analisis yang tepat untuk mengekstraknya menjadi wawasan yang bermanfaat bagi pengelola wisata.

Analisis sentimen adalah metode dalam text mining yang digunakan untuk mengelompokkan opini dalam teks menjadi kategori positif, negatif, atau netral [5]. Salah satu algoritma yang sering digunakan dalam analisis sentimen adalah K-Nearest Neighbor (KNN), yang bekerja dengan mengklasifikasikan data berdasarkan kedekatan antar sampel [1]. Algoritma ini dikenal sederhana namun efektif dalam menangani klasifikasi berbasis teks, termasuk dalam analisis ulasan wisata.

Beberapa penelitian sebelumnya telah menerapkan KNN dalam analisis sentimen ulasan wisatawan. [3] mengimplementasikan KNN untuk menganalisis sentimen wisata Dieng, dengan akurasi mencapai 86%. [2] menggunakan KNN dalam klasifikasi ulasan aplikasi transportasi dan memperoleh akurasi terbaik sebesar 90,23%. Namun, penelitian terkait analisis sentimen terhadap ulasan wisata Lawang Sewu masih terbatas. Selain itu, tantangan utama dalam klasifikasi sentimen adalah ketidakseimbangan distribusi data, terutama dalam membedakan sentimen netral yang sering memiliki kesamaan pola dengan kategori lainnya [6].

Penelitian ini bertujuan untuk menerapkan algoritma KNN dalam klasifikasi sentimen ulasan Lawang Sewu di Google Maps serta mengevaluasi performanya menggunakan metrik akurasi dan F1-score [7]. Dengan analisis ini, diharapkan hasil penelitian dapat memberikan wawasan bagi pengelola destinasi wisata dalam meningkatkan kualitas layanan berdasarkan opini pengunjung.

2. TINJAUAN PUSTAKA

Penelitian terkait analisis sentimen dalam sektor pariwisata telah berkembang pesat. [3] menggunakan algoritma K-Nearest Neighbor (KNN) untuk menganalisis ulasan wisata Dieng dan mencapai akurasi sebesar 86%. Hasil ini menunjukkan kemampuan KNN dalam mengelompokkan sentimen positif dan negatif dari ulasan wisatawan. Dalam penelitian lain, [5] memanfaatkan KNN untuk klasifikasi ulasan aplikasi transportasi Maxim dan mendapatkan akurasi tertinggi sebesar 90,23% dengan pembobotan fitur TF-IDF.

Tantangan utama dalam analisis sentimen ulasan adalah ketidakseimbangan distribusi data sentimen, khususnya dalam membedakan sentimen netral. [2] menyoroti pentingnya teknik pra-proses seperti normalisasi dan stopword removal untuk meningkatkan akurasi model KNN. Penelitian tersebut juga menggarisbawahi perlunya eksplorasi algoritma lain dan pembobotan fitur seperti TF-IDF untuk meningkatkan performa klasifikasi.

Text mining menjadi elemen penting dalam analisis sentimen. Teknik ini mampu mengekstraksi informasi dari teks yang tidak terstruktur dan memberikan wawasan berharga bagi pengelola wisata [4]. Dalam konteks analisis sentimen ulasan wisata, algoritma KNN sering digunakan karena

kesederhanaannya dan efektivitasnya dalam klasifikasi berbasis teks [7].

Penelitian ini memberikan dasar teoretis yang mendukung pemilihan metode KNN dalam analisis sentimen ulasan Lawang Sewu. Dengan memanfaatkan teknik text preprocessing dan evaluasi performa model menggunakan confusion matrix, penelitian ini diharapkan dapat memberikan hasil yang bermanfaat dalam memahami persepsi pengunjung serta meningkatkan layanan destinasi wisata.

3. METODE PENELITIAN

Penelitian ini bertujuan untuk melakukan klasifikasi sentimen terhadap ulasan pengunjung Lawang Sewu yang diperoleh dari platform Google Maps. Penelitian menggunakan algoritma K-Nearest Neighbor (KNN) serta teknik text preprocessing untuk pengolahan data teks. Berikut adalah tahapan yang dilakukan:

3.1. Pengumpulan Data

Data dikumpulkan secara otomatis dengan menggunakan teknologi web scraping berbasis SerpAPI. Proses ini memungkinkan peneliti memperoleh ulasan terkait Lawang Sewu, termasuk nama pengguna, rating, tanggal ulasan, dan isi teks ulasan. Data kemudian disimpan dalam format CSV agar mudah diolah lebih lanjut dalam proses analisis.

3.2. Preprocessing Data

Data yang telah dikumpulkan perlu melalui tahap praproses agar lebih bersih dan relevan untuk analisis. Adapun tahapan praproses yang diterapkan adalah sebagai berikut:

- Cleaning:** Menghapus karakter khusus, tanda baca, angka, emoji, dan URL yang tidak relevan dengan analisis sentimen.
- Case Folding:** Mengubah semua huruf dalam teks menjadi huruf kecil untuk konsistensi pemrosesan.
- Stemming:** Mengubah kata berimbuhan ke bentuk dasarnya, misalnya "berjalan" menjadi "jalan". Teknik ini menggunakan library NLP seperti Sastrawi untuk bahasa Indonesia.
- Stopword Removal:** Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis sentimen, seperti "dan", "yang", atau "adalah".

3.3. Pelabelan Data

Penentuan label sentimen didasarkan pada rating ulasan yang diberikan oleh pengguna:

- Negatif:** Rating 1 dan 2
- Netral:** Rating 3
- Positif:** Rating 4 dan 5

Tahap ini memastikan bahwa setiap data memiliki kategori sentimen yang jelas untuk proses pelatihan model KNN.

3.4. Pemodelan Klasifikasi

Algoritma K-Nearest Neighbor (KNN) dipilih karena kesederhanaannya dan kinerjanya yang baik dalam klasifikasi berbasis teks. Model ini bekerja dengan menentukan kategori sentimen berdasarkan kedekatan antar sampel dalam ruang multidimensi. Parameter utama yang digunakan adalah jumlah tetangga terdekat $k=5$, sedangkan jarak antar sampel dihitung menggunakan metode Euclidean Distance. Rumus Euclidean Distance yang digunakan adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

3.5. Evaluasi Model

Evaluasi model dilakukan dengan menggunakan confusion matrix untuk menghitung metrik berikut:

- Accuracy:** Persentase prediksi yang benar terhadap keseluruhan data.
- Precision:** Kemampuan model dalam memprediksi sentimen positif dengan tepat.
- Recall:** Kemampuan model dalam mendeteksi semua data positif yang sebenarnya.
- F1-Score:** Rata-rata harmonis antara precision dan recall, digunakan untuk menilai kinerja model secara keseluruhan.

Rumus evaluasi metrik:

$$\begin{aligned} \text{e. Accuracy} &= \frac{\text{Jumlah prediksi benar}}{\text{Total data}} \\ &= \frac{\text{TPpos} + \text{TPneg} + \text{TPnet}}{\text{Total data}} \end{aligned} \quad (2)$$

$$\begin{aligned} \text{f. Precision} & \\ \text{Precisionpos} &= \frac{\text{TPpos}}{\text{TPpos} + \text{FP}_+} \\ \text{Precisionneg} &= \frac{\text{TPneg}}{\text{TPneg} + \text{FP}_-} \\ \text{Precisionnet} &= \frac{\text{TPnet}}{\text{TPnet} + \text{FP}_0} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{g. Recall} & \\ \text{h. Recallpos} &= \frac{\text{TPpos}}{\text{Total data netral}} \\ \text{i. Recallneg} &= \frac{\text{TPneg}}{\text{Total data netral}} \\ \text{j. Recallnet} &= \frac{\text{TPnet}}{\text{Total data netral}} \end{aligned} \quad (4)$$

$$\begin{aligned} \text{k. F1-Score} & \\ \text{F1pos} &= 2 \times \frac{\text{Precisionpos} \times \text{Recallpos}}{\text{Precisionpos} + \text{Recallpos}} \\ \text{F1neg} &= 2 \times \frac{\text{Precisionneg} \times \text{Recallneg}}{\text{Precisionneg} + \text{Recallneg}} \\ \text{F1net} &= 2 \times \frac{\text{Precisionnet} \times \text{Recallnet}}{\text{Precisionnet} + \text{Recallnet}} \end{aligned}$$

3.6. Evaluasi Model

Hasil analisis menunjukkan distribusi sentimen ulasan sebagai berikut: 52,8% positif, 27,4% negatif, dan 19,8% netral. Evaluasi performa model menghasilkan akurasi sebesar 64% dengan F1-score rata-rata 0,60. Tantangan utama ditemukan pada klasifikasi sentimen netral yang memiliki pola serupa dengan sentimen lain, yang menyebabkan penurunan performa model.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan wawasan bagi pengelola Lawang Sewu terkait persepsi pengunjung dan menjadi referensi untuk peningkatan layanan di masa depan.

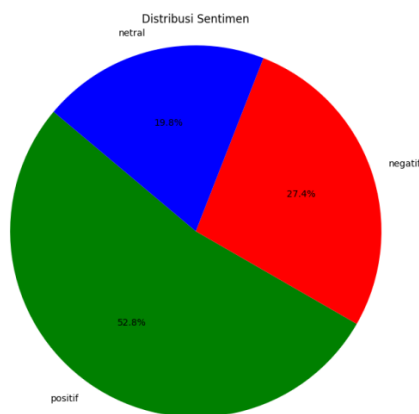
4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan dan Praproses Data

Data ulasan Lawang Sewu dikumpulkan dengan teknik web scraping menggunakan SerpAPI, menghasilkan total 567 ulasan yang terdiri dari nama pengguna, rating, tanggal ulasan, serta isi teks ulasan. Setelah melalui proses praproses seperti cleaning, case folding, stemming, dan penghapusan stopword, data yang telah bersih dan siap digunakan sebanyak 500 ulasan dengan distribusi sentimen sebagai berikut

- Positif (Rating 4-5): 52,8%
- Netral (Rating 3): 27,4%
- Negatif (Rating 1-2): 19,8%

Distribusi ini divisualisasikan dalam diagram pada Gambar 1.



Gambar 1 Distribusi Sentimen Ulasan Pengunjung Lawang Sewu

4.2. Klasifikasi Sentimen Menggunakan KNN

Algoritma K-Nearest Neighbor (KNN) diterapkan dengan parameter $k=5$. Pemilihan nilai k dilakukan berdasarkan evaluasi performa model dengan berbagai nilai tetangga terdekat. Jarak antar sampel dihitung menggunakan metode Euclidean Distance.

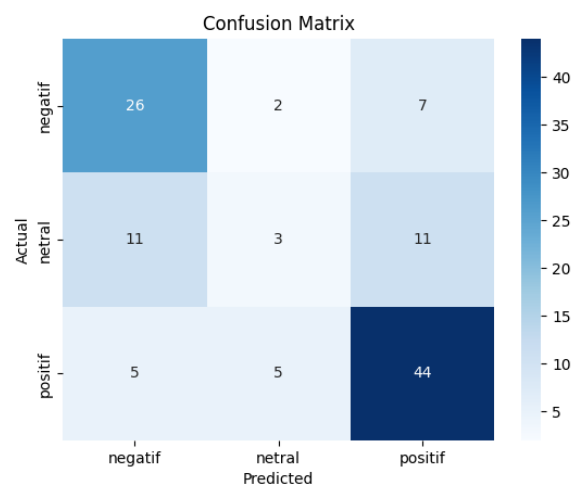
4.3. Evaluasi Kinerja Model

Evaluasi model dengan confusion matrix menghasilkan metrik sebagai berikut:

Tabel 1. Tabel Model

Metrik	Nilai
Accuracy	64%
Precision	0,62
Recall	0,57
F1-Score	0,60

Meskipun akurasi mencapai 64%, terdapat tantangan dalam membedakan sentimen netral yang sering kali memiliki pola bahasa serupa dengan kategori positif dan negatif.

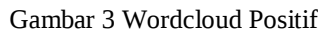


Gambar 2 Confusion Matrix

Pada Gambar 2 Confusion Matrix menggambarkan perbandingan antara Actual dan Predicted dalam analisis sentimen ulasan Lawang Sewu menggunakan K-Nearest Neighbor (KNN). Pada matrix ini, baris menunjukkan jumlah data berdasarkan kategori sentimen sebenarnya, sementara kolom merepresentasikan hasil prediksi model. Nilai yang berada pada diagonal matrix menandakan prediksi yang sesuai dengan data asli, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi, seperti sentimen netral yang sering terprediksi sebagai positif atau negatif. Hal ini menunjukkan bahwa model masih menghadapi kesulitan dalam mengidentifikasi sentimen netral dengan baik karena kemiripan pola bahasa. Oleh karena itu, evaluasi ini memberikan wawasan mengenai aspek yang perlu diperbaiki untuk meningkatkan akurasi model dalam klasifikasi sentimen.

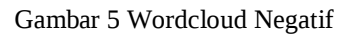
4.4. Analisis Wordcloud

Untuk memahami kata-kata yang sering muncul dalam setiap sentimen, visualisasi wordcloud digunakan dan dapat dilihat pada gambar berikut.



Gambar 4 Wordcloud Netral

Pada Gambar 5 menampilkan wordcloud untuk sentimen negatif, di mana kata-kata seperti gelap, kurang penerangan, dan mahal sering muncul dalam ulasan. Hal ini mengindikasikan adanya keluhan dari pengunjung terkait pencahayaan yang kurang memadai serta biaya yang dianggap tinggi. Kata-kata ini memberikan wawasan mengenai aspek yang masih perlu diperbaiki agar pengalaman wisata di Lawang Sewu menjadi lebih nyaman bagi semua pengunjung.



Evaluasi model dengan Confusion Matrix menunjukkan bahwa algoritma KNN memiliki tantangan dalam mengklasifikasikan sentimen netral, yang sering kali dikategorikan sebagai positif atau negatif. Hal ini dapat terjadi karena kemiripan pola bahasa dalam ulasan netral dengan kategori lainnya. Selain itu, berdasarkan analisis wordcloud, ulasan positif didominasi oleh kata-kata seperti bagus, bersejarah, dan spot selfie, sementara ulasan negatif sering mengandung kata gelap, kurang penerangan, dan mahal. Ulasan netral umumnya berisi informasi tanpa ekspresi emosional yang jelas.

Evaluasi model dengan Confusion Matrix menunjukkan bahwa model masih mengalami kesulitan dalam membedakan sentimen netral, yang sering kali diklasifikasikan sebagai positif atau negatif akibat kesamaan pola bahasa. Selain itu, analisis wordcloud memperlihatkan kata-kata yang sering muncul dalam setiap kategori sentimen, yang mencerminkan aspek-aspek utama yang diperhatikan oleh pengunjung.

Berdasarkan temuan ini, pengelola Lawang Sewu disarankan untuk meningkatkan fasilitas yang sering menjadi keluhan, seperti pencahayaan dan kenyamanan pengunjung, guna meningkatkan pengalaman wisata secara keseluruhan. Selain itu, informasi dari ulasan netral dapat dimanfaatkan untuk memahami aspek yang masih perlu dikembangkan, meskipun tidak secara eksplisit mengandung sentimen tertentu. Dengan memahami persepsi pengunjung melalui analisis sentimen ini, diharapkan pengelola dapat mengambil langkah-langkah yang lebih tepat dalam meningkatkan layanan wisata.

DAFTAR PUSTAKA

- [1] R. Ramadhan, M. Afdal, I. Permana, and M. Jazman, "Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor," *J. Ris. Komputer*, vol. 10, no. 3, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i3.6396.
- [2] M. P. R. Putra and K. R. N. Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 1, pp. 63–71, 2020, doi: 10.32767/jutim.v5i1.791.
- [3] Y. W. Syaifudin and R. A. Irawan, "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means," *J. Inform. Polinema*, vol. 4, no. 3, pp. 189–194, 2018, doi: 10.33795/jip.v4i3.205.
- [4] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 2, no. 1, pp. 32–41, 2017, doi: 10.31284/j.integer.2017.v2i1.95.
- [5] I. N. T. I. Nabawi, Rudiman, and F. Yulianto, "Analisis Sentimen Pada Ulasan Aplikasi Google Maps Terhadap Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Samarinda Menggunakan Metode K-Nearest Neighbor Dengan Fitur Ekstraksi TF-IDF," *J. Teknol. Inf.*, vol. Vol 18, no. 2, p. 17, 2024.
- [6] Sarah Anis, S. Saad, and M. Aref, "A survey on sentiment analysis in tourism," *Int. J. Intell. Comput. Inf. Sci.*, vol. 0, no. 0, pp. 1–20, 2020, doi: 10.21608/ijicis.2020.22851.1014.
- [7] I. A. Safra and E. Zuliarso, "Analisa Sentimen Persepsi Masyarakat Terhadap Pemindahan Ibukota Baru di Kalimantan Timur pada Media Sosial Twitter," *Proceeding SENDIU 2020*, pp. 978–979, 2020, [Online]. Available: <https://unisbank.ac.id/ojs/index.php/sendiu/article/view/7984/2947>