Jurnal Ilmiah Sinus (JIS) Vol : 23, No. 2, Juli 2025
ISSN (Print) : 1693-1173 , ISSN (Online): 2548-4028

Analisis Sentimen pada Ulasan Google Review Terhadap Ekowisata Sungai Mudal menggunakan Algoritme Stochastic Gradient Descent Classifier

Muhammad Ali Pratama[1)], Muhammad Zakariyah[2)*]
[1)] Program Studi Informatika, Universitas Teknologi Yogyakarta
[2)] Program Studi Informatika Medis, Universitas Teknologi Yogyakarta
[1)] muhammad.5220411416@student.uty.ac.id, [2)] muhammad.zakariyah@staff.uty.ac.id

ABSTRACT

Mudal River is an ecotourism developed by PT. PLN (Persero) located in Kulon Progo Regency, Special Region of Yogyakarta. As a tourism developer, visitors' impressions can be used as evaluation material to improve tourism facilities and services. This study aims to determine the distribution of positive and negative sentiments from visitor reviews that can be used as support in business decision making. A total of 3,663 data were obtained from Google Review, then pre-processing was carried out to obtain higher quality data. Labeling/annotation uses Large Language Models (chatGPT), while the classification process uses the Stochastic Gradient Descent Method. Model testing uses the K-Fold method to obtain average accuracy, recall, and AUC values. The classification performance results produce an accuracy of 85.5%, with a dominance of positive labels in the annotation results. Meanwhile, the wordcloud results show that words such as water quality, waterfalls, mudal river, and photo spots have positive affirmations. However, several areas are of concern in negative sentiment, especially related to roads and photo spot management, as well as the issue of entrance tickets and the potential for paid photos.

Keywords: *Sentiment Analysis, Text Mining, Mudal River Ecotourism.*

I. PENDAHULUAN

Sektor pariwisata merupakan sumber pendapatan asli daerah yang sangat besar di wilayah Yogyakarta, dengan wisata alam sebagai destinasi wisata yang sangat diminati oleh pengunjung (Wicaksono, 2020). Untuk melakukan pengembangan wisata, setidaknya ada empat potensi yang perlu diperhatikan, yaitu amenitas, aksesibilitas, atraksi, dukungan kelembagaan (Irsyad, 2020). Keempat potensi tersebut merupakan faktor kunci dalam pengambilan keputusan bisnis yang strategis untuk pengembangan pariwisata, serta mempengaruhi kepuasan pengunjung dan reputasi destinasi. Salah satu sumber data yang dapat menjadi bahan *business decision* adalah ulasan pengunjung yang salah satunya dapat dilihat pada Google Review. Pengaruh ulasan pengunjung terhadap pelayanan dan fasilitas akan berdampak pada citra destinasi wisata.

Jumlah wisatawan domestik maupun mancanegara yang mengunjungi daerah pariwisata di wilayah Yogyakarta, mengalami kenaikan selama 5 tahun terakhir seperti terlihat pada Gambar 1 (Badan Pusat Statistik DIY, 2024). Ekowisata Sungai Mudal merupakan salah satu destinasi wisata yang terletak di Desa Jatimulyo, Kecamatan Girimulyo, Kulon Progo, Daerah Istimewa Yogyakarta yang cukup banyak dikunjungi wisatawan. Menurut Riana (2023), pihak pengelola ekowisata Sungai Mudal masih belum

Gambar 1. Tren pengunjung wisatawan di daerah Yogyakarta
Sumber: Badan Pusat Statistik DIY, 2024

menyediakan kotak kritik dan saran di tempat wisata tersebut, sehingga pengunjung cukup kesulitan untuk memberikan *feedback* atas fasilitas dan pelayanan yang diberikan. Selain itu, pihak pengelola Ekowisata Sungai Mudal menjadi sulit untuk memahami preferensi, harapan, dan kebutuhan pengunjung.

Google LLC telah menyediakan platform *review* terhadap lokasi (termasuk objek pariwisata) melalui fitur Google Review yang terdapat pada aplikasi Google Maps. Hal ini tentunya menjadi peluang bagi pengelola Ekowisata Sungai Mudal untuk memanfaatkan teknologi pencatat ulasan lokasi tersebut. Ulasan dari para pengunjung tersebut kemudian dianalisis dengan menggunakan *Natural Language Processing* (NLP) dan algoritme *machine learning* untuk mendapatkan informasi yang lebih dalam. Hasil ulasan tersebut tentunya dapat digunakan untuk menganalisis perilaku pengunjung serta membantu dalam peningkatan pelayanan bagi pengelola wisata.

Untuk mengetahui sentimen positif dan negatif dari ulasan yang berbasis teks, dapat dilakukan analisis sentimen menggunakan algoritme klasifikasi seperti algoritme *Support Vector Machine* (Aulia dkk., 2021) dan Naive Bayes (Khofifah dkk., 2022). Kedua algoritme tersebut memiliki kelemahan dalam hal penanganan data dalam jumlah yang besar dan asumsi bahwa semua fitur independen satu sama lain (Nisa dkk., 2019). Namun pada kenyataannya, data teks seringkali memiliki ketergantungan dan korelasi yang kuat. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis NLP dengan menggunakan Algoritme *Stochastic Gradient Descent Classifier,* serta memberikan rangkuman data ulasan pengunjung yang lebih mudah dipahami oleh pengelola wisata. Tentunya dengan adanya teknologi ini, Ekowisata Sungai Mudal dapat melakukan evaluasi rutin dan mendapatkan peluang sebagai destinasi wisata dari pengunjung lokal hingga mancanegara.

II. TINJAUAN PUSTAKA

2.1 Pelabelan Sentimen menggunakan *Large Language Models* (LLMs)

Pelabelan data secara manual untuk analisis sentimen membutuhkan waktu dan sumber daya yang signifikan. Salah satu pendekatan untuk mengatasi hal ini adalah dengan pelabelan berbasis rating, di mana rating bintang yang diberikan pengunjung dikonversi menjadi label sentimen. Namun, pendekatan ini mungkin tidak selalu menangkap nuansa sentimen yang sebenarnya, terutama pada ulasan yang kompleks (Pangakis & Wolken, 2024). Perkembangan terkini dalam NLP menunjukkan potensi LLMs seperti model *Generative Pre-trained Transformer* (GPT) untuk berbagai tugas, termasuk pelabelan data. Teknik *prompt engineering* (khususnya *zero-shot prompting*) dapat diinstruksikan untuk memberikan label sentimen pada teks dengan memberikan beberapa contoh (*shots*) ulasan beserta labelnya (Ghatora dkk., 2024). Pendekatan ini diharapkan dapat menghasilkan label yang lebih akurat dan kontekstual dibandingkan pelabelan berbasis rating sederhana.

2.2 *Term Frequency-Inverse Document Frequency* (TF-IDF)

Metode TF-IDF adalah sebuah metode yang dapat menilai tingkat pentingnya suatu kata dalam sebuah dokumen, terutama ketika dokumen tersebut merupakan bagian dari koleksi atau korpus yang besar. Manfaat utama TF-IDF adalah kemampuannya untuk menyoroti kata-kata yang paling representatif dan khas untuk suatu dokumen (Liu dkk., 2022). Hal ini dilakukan dengan cara menyeimbangkan frekuensi kemunculan sebuah kata dalam dokumen tersebut (*Term Frequency*) dengan seberapa umum kata itu muncul di seluruh dokumen dalam korpus (*Inverse Document Frequency*). TF-IDF secara efektif mengurangi bobot kata-kata yang sering muncul secara umum, membantu mengidentifikasi kata kunci yang signifikan dan meningkatkan relevansi dalam analisis teks (Raza dkk., 2021). Persamaan 1, 2, dan 3 menunjukkan formulasi TF-IDF.

$$TF(t,d) = \frac{jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{jumlah\ total\ kata\ dalam\ dokumen\ d} \quad (1)$$

 DOI : https://doi.org/10.30646/sinus.v23i2.938

$$IDF(t,D) = \log\left(\frac{jumlah\ total\ dokumen\ dalam\ korpus\ (|D|)}{jumlah\ dokumen\ yamg\ mengandung\ kata\ t\ (df_t)}\right) \quad (2)$$

$$TFIDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (3)$$

2.3 *Stochastic Gradient Descent* (SGD)

Algoritme ini sederhana dan efisien, yaitu dengan menggunakan pendekatan turunan pada klasifikasi linear. Klasifikasi SGD melibatkan beberapa iterasi untuk mencari nilai *minimum functionality points*. Untuk perhitungan pada setiap iterasi dapat dilihat pada Persamaan 4.

$$\omega_i + 1 = \omega_i - \eta\nabla_{\omega_i} L(\omega_i) \quad (4)$$

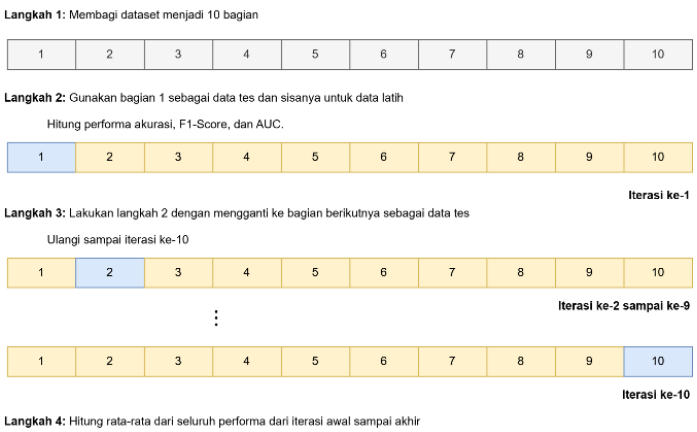
$\omega_i + 1$ merupakan parameter model prediksi, ω_i sebagai parameter model pada iterasi sebelumnya, η adalah *learning rate*, dan L merupakan *loss function*. Untuk formulasi *loss function* dapat dilihat pada persamaan 5.

$$L(X_j, Y_j) = max(0, 1 - Y_j . (\omega X_j + b) \quad (5)$$

ω dan b merupakan parameter model untuk prediksi, X_j adalah sampel input, dan Y_j adalah target kelas. Persamaan ini digunakan untuk perhitungan *loss function* pada setiap iterasi ketika sedang dalam pelatihan data (Antonio dkk., 2022).

2.4 K-Fold Cross-Validation

Evaluasi terhadap algoritme yang dipakai menggunakan k-Fold Cross Validation. Gambar 2 menjelaskan tahapan k-Fold Cross Validation untuk melakukan evaluasi model dengan data yang terbatas (Zakariyah & Zaky, 2022). Pada penelitian ini, parameter yang digunakan yaitu 10 jumlah kelompok (k) untuk melakukan evaluasi. Setiap iterasi akan dikalkulasi skor *accuracy*, *recall*, dan *Area Under Curve* (AUC) lalu diambil rata-rata dari masing-masing model.



Gambar 2. Langkah metode k-Fold cross-validation dengan k=10

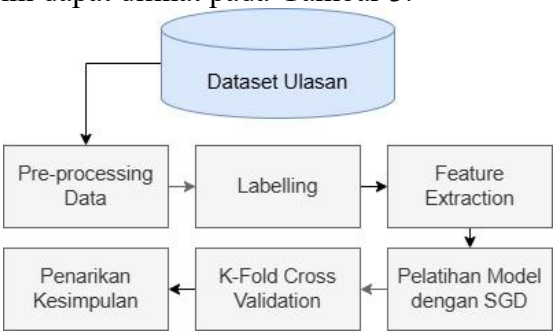
Persamaan 6 dan 7 merupakan formula dari nilai *accuracy* dan *recall*. Dimana nilai *true positives* (TN), *true negatives* (TN), *false positives* (FP), *false negatives* (FN) didapatkan dari *confusion matrix* (Onan, 2021).

$$ACC = \frac{TN + TP}{TP + FP + FN + TN} \quad (6)$$

$$REC = \frac{TP}{TP + FN} \quad (7)$$

III. METODE PENELITIAN

Penelitian ini termasuk dalam penelitian *Knowledge Discovery in Databases* (KDD) yang bertujuan untuk memudahkan dalam pencarian informasi data (Marquis dkk., 2020). Tahapan dalam penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Tahapan penelitian

3.1. Pengumpulan data

Pada tahapan ini, data hasil ulasan/review diambil dari halaman Google Review. Metode pengambilan data yang dilakukan, yaitu melalui *web scraping* dengan menggunakan Modul Selenium pada bahasa pemrograman Python (Yuan, 2023). Penelitian ini memiliki kebutuhan fitur diantaranya sebagai berikut pada Tabel 1.

Tabel 1. Kebutuhan fitur data

Fitur	Kegunaan
msg_id	Mengidentifikasi baris data
date	Mengetahui tanggal ulasan dibuat
name	Mengetahui nama *username* pengunjung
rating	Jumlah bintang kepuasan yang diberikan pengunjung
msg_review	Berisi ulasan yang terdiri dari beberapa kata

Berikut merupakan sumber data dari *website* Google Review dengan kata kunci "Ekowisata Sungai Mudal" yang diakses pada tanggal 10 Mei 2025 (Gambar 4).

msg_id	date	name	rating	msg_review
ChZDSUhNMG9nS0VJQ0FnSUNUNzY2d0FREAE	7 jam lalu	Kopet Kepet	5 bintang	Kemarin sudah kesana, tempat nya benar2 indah ...
ChZDSUhNMG9nS0VJQ0FnSUNUNzY2d0FREAE	7 jam lalu	Kopet Kepet	5 bintang	Kemarin sudah kesana, tempat nya benar2 indah ...
ChZDSUhNMG9nS0VJQ0FnSUNUdDhhcU1REAE	16 jam lalu	nur auliya	5 bintang	Ekowisata yang harus bgt dikunjungi karena ben...
ChZDSUhNMG9nS0VJQ0FnSUNUdDhhcU1REAE	16 jam lalu	nur auliya	5 bintang	Ekowisata yang harus bgt dikunjungi karena ben...
ChZDSUhNMG9nS0VJQ0FnSUNUdDVyUWRBEAE	16 jam lalu	Keceriaan Ishanaishan	5 bintang	Rekomendasi wisata alam, masih alami, air sung...

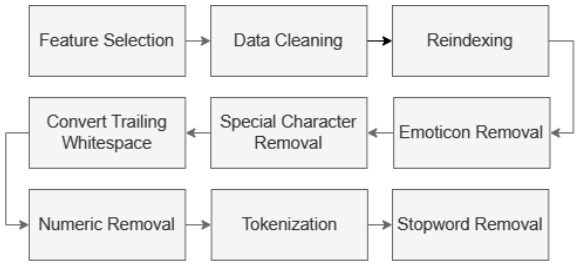
Gambar 4. Contoh data hasil metode *web scraping*

Data yang didapatkan masih banyak duplikat dan masih belum dapat dipahami konteksnya dengan baik. Untuk mengubah data tersebut menjadi bentuk data yang diinginkan, maka perlu dilakukan *preprocessing data.*

 DOI : https://doi.org/10.30646/sinus.v23i2.938

3.2. *Pre-processing* data

Setelah mendapatkan *raw data*, harus dilakukan proses pengelolaan data agar dapat dilatih dan dievaluasi. Tahap ini sangat berpengaruh terhadap performa data yang akan dilatih (Khairunnisa dkk., 2021).



Gambar 5. Langkah *preprocessing* data

Setiap tahapan yang dilakukan pada Gambar 5 tersebut, dijelaskan secara singkat pada Tabel 2 berikut ini.

Tabel 2. Tahapan *pre-processing* data

Langkah	Deskripsi
Feature Selection	Melakukan seleksi fitur yang hanya dibutuhkan saja
Data Cleaning	Membersihkan data dari duplikat dan data kosong
Reindexing	Membuat nomor urut ulang dari baris pertama sampai akhir
Emoticon Removal	Proses menghilangkan karakter emoji dari teks **Data sebelum diproses**: Baguss buat ngadem 😍 Kalo weekend rame **Data sesudah diproses**: Baguss buat ngadem Kalo weekend rame
Special Character Removal	Proses menghilangkan dari karakter &, *, #, atau lainnya **Data sebelum diproses**: Tiket masuk: 10k + parkir 2k **Data sesudah diproses**: Tiket masuk 10k parkir 2k
Convert Trailing Whitespace	Melakukan standarisasi spasi yang terlalu lebar **Data sebelum diproses**: bagus dan keren **Data sesudah diproses**: bagus dan keren
Numeric Removal	Proses menghapus karakter numerik dari teks **Data sebelum diproses**: Parkir motor Rp 2000 Tiket masuk Rp 10000 per orang **Data sesudah diproses**: Parkir motor Rp Tiket masuk Rp per orang
Tokenization	Proses memotong kata dalam kalimat berdasarkan spasi
Stopword Removal	Proses menghapus kata sambung yang tidak efektif **Data sebelum diproses**: Tempatnya bagus dan terawat dengan baik alamatnya juga jelas di Map **Data sesudah diproses**: Tempatnya bagus terawat baik alamatnya jelas Map

3.3. *Labelling*

Setelah tahap *pre-processing data*, setiap ulasan diberi label sentimen positif atau negatif menggunakan LLMs model GPT-4o-mini dari OpenAI. Model ini dipilih karena kecerdasannya yang setara dengan GPT-4 Turbo namun lebih cepat dan efisien (Beno, 2025), serta memiliki pemahaman Bahasa Indonesia yang baik, yang relevan untuk dataset penelitian. Proses pelabelan menerapkan teknik *zero-shot prompting* (Shu dkk., 2022). Teknik ini menginstruksikan GPT-4o-mini untuk menentukan polaritas sentimen ulasan

tanpa memerlukan contoh spesifik sebelumnya, melainkan mengandalkan pemahaman internalnya dari proses *pre-training*.

Prompt dirancang untuk meminta model mengklasifikasikan setiap ulasan ke dalam kategori "Positif" atau "Negatif". Dari total 3.663 ulasan yang diproses, tahap pelabelan ini menghasilkan 2.965 ulasan dengan sentimen positif dan 698 ulasan dengan sentimen negatif. Distribusi label ini kemudian digunakan untuk analisis data eksploratif dan tahap pemodelan klasifikasi. Gambar 6 merupakan contoh dari teknik *zero-shot prompting* yang digunakan.

```
response = client.chat.completions.create(
    messages=[
        {
            "role": "system",
            "content": "Kamu ahli sentiment analysis.",
        },
        {
            "role": "user",
            "content": f"""
            Kamu adalah asisten yang akan membaca ulasan wisata dan mengembalikan satu dari dua label:
            - Positif,
            - Negatif
            Beri jawabannya hanya label saja, tanpa penjelasan.

            Ulasan: "{text}"
            Label:
            """,
        }
    ],
    temperature=0.0,
    model=deployment
)
```

Gambar 6. Teknik *zero-shot prompting*

3.4. *Feature Extraction*

Setelah tahap pelabelan, langkah selanjutnya adalah ekstraksi fitur untuk mengubah data teks ulasan menjadi representasi numerik yang dapat diproses oleh algoritme klasifikasi. Proses ini diawali dengan tokenisasi, di mana setiap ulasan dipecah menjadi unit-unit kata individual atau token. Penelitian ini menggunakan modul NLTK untuk mengolah token kata dari dataset, sehingga dataset akan berbentuk *wordlist*. Token kata yang banyak digunakan pada masing-masing sentimen diurutkan dari 10 terbanyak.

Metode TF-IDF diterapkan untuk menghitung bobot setiap token dalam setiap ulasan. TF-IDF bekerja dengan memberikan bobot yang tinggi kepada kata-kata yang sering muncul dalam satu ulasan tertentu (*Term Frequency*), namun jarang muncul di ulasan-ulasan lain dalam keseluruhan dataset (*Inverse Document Frequency*). Representasi numerik dari TF-IDF ini kemudian menjadi vektor fitur yang akan digunakan sebagai *input* untuk melatih model klasifikasi *Stochastic Gradient Descent* (SGD).

3.5. Pelatihan Model dengan SGD *Classifier*

Algoritme klasifikasi Model SGD dilatih menggunakan vektor fitur yang dihasilkan dari TF-IDF untuk mempelajari pola yang dapat membedakan antara sentimen positif dan negatif. Kinerja model dievaluasi menggunakan skema K-Fold *cross-validation* guna memastikan hasil evaluasi yang lebih *robust* dan tidak bias terhadap pembagian data latih dan uji. Tabel 3 memberikan detail *hyperparameter* yang digunakan untuk melatih model sentimen.

Tabel 3. *Hyperparameter* yang digunakan

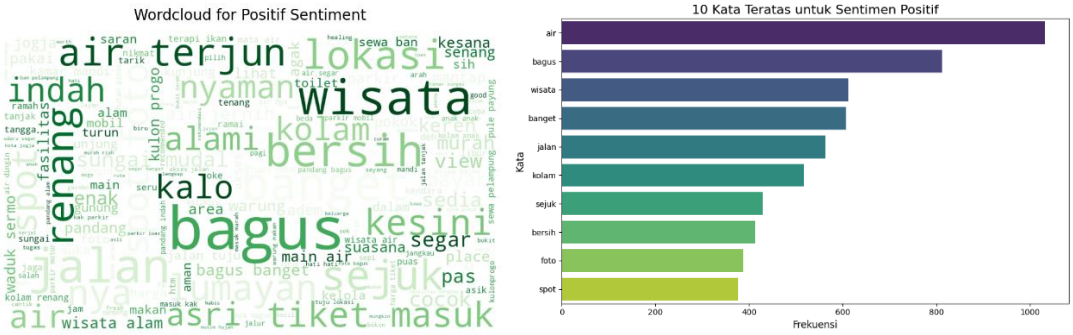
Hyperparameter	*Value*
loss	log_loss
max iteration	1000
tolerance	1×10^{-3}
random state	42

 DOI : https://doi.org/10.30646/sinus.v23i2.938

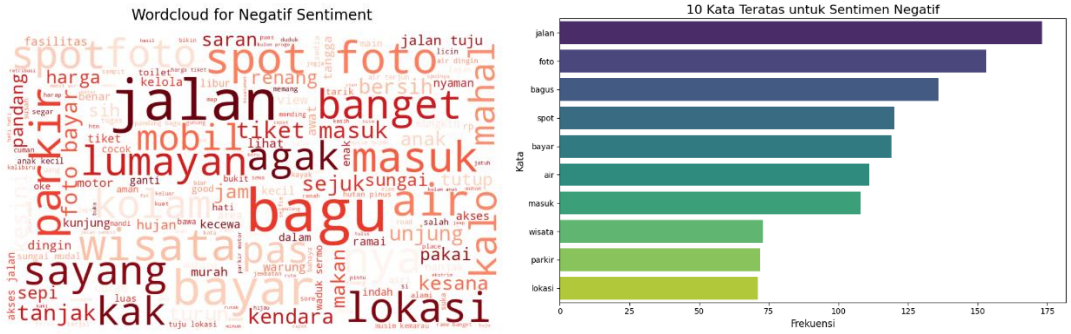
IV. HASIL DAN PEMBAHASAN

4.1. Hasil Pengolahan Data

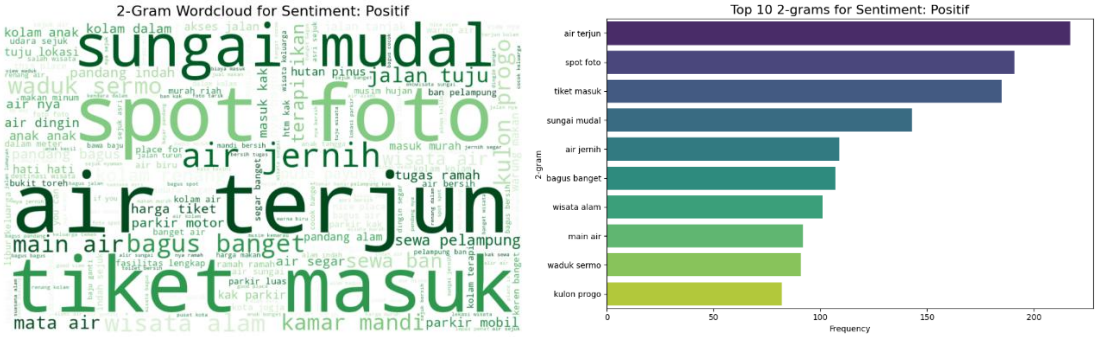
Setelah melakukan proses *feature extraction*, didapatkan daftar kata yang sering ditemukan pada masing-masing sentimen yang disajikan dalam gambar *word cloud* berikut.



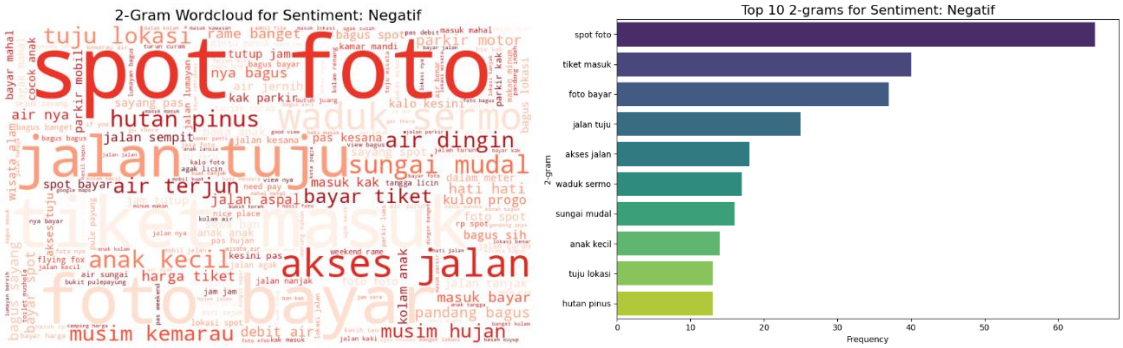
Gambar 7. Daftar 10 kata unigram yang mendominasi untuk sentimen positif



Gambar 8. Daftar 10 kata unigram yang mendominasi untuk sentimen negative



Gambar 9. Daftar 10 kata bigram yang mendominasi untuk sentimen positif

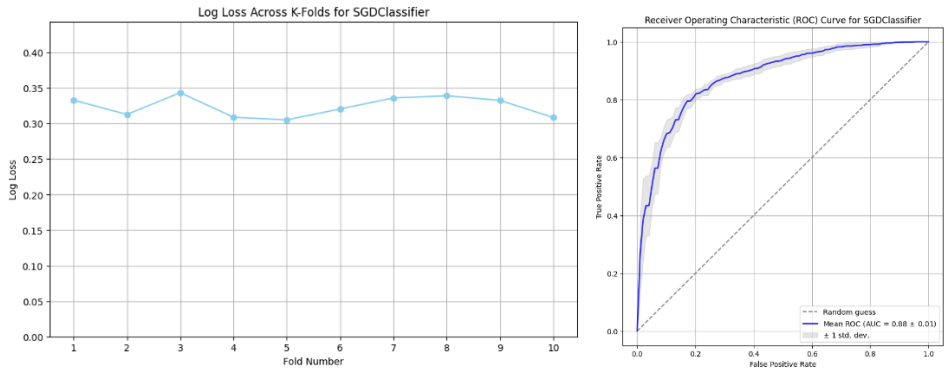


Gambar 10. Daftar 10 kata bigram yang mendominasi untuk sentimen negatif

Gambar 7 dan 8 menampilkan *word cloud* dari kata 1-gram (unigram) yang mendominasi *dataset* masing-masing sentimen. Gambar 9 dan 10 menampilkan *word cloud* dari kata 2-gram (bigram) yang mendominasi *dataset* masing-masing sentimen.

4.2. Perbandingan Skor Evaluasi Model

Model SDG dievaluasi dengan menggunakan 10-K Fold Cross Validation. Detail hasil pengujian berupa rata-rata skor Log Loss dan kurva ROC dari keseluruhan iterasi 10-fold (Gambar 11).



Gambar 11. Skor evaluasi log loss dan model dengan kurva ROC dari setiap fold

Dari semua fold yang dijalankan, model SGD Classifier menunjukkan detail hasil performa seperti pada tabel 3.

Tabel 3. Hasil evaluasi model

Fold	Akurasi (%)	Recall (%)	AUC
1/10	0.8583	0.8583	0.8690
2/10	0.8583	0.8583	0.8876
3/10	0.8365	0.8365	0.8640
4/10	0.8579	0.8579	0.8954
5/10	0.8743	0.8743	0.8887
6/10	0.8497	0.8497	0.8886
7/10	0.8470	0.8470	0.8780
8/10	0.8525	0.8525	0.8614
9/10	0.8415	0.8415	0.8815
10/10	0.8743	0.8743	0.8904
Rata-rata	**0.8550 ± 0.0118**	**0.8550 ± 0.0118**	**0.8805 ± 0.0113**

4.3. Pembahasan

Analisis eksploratif data melalui wordcloud dan daftar kata teratas memberikan gambaran yang jelas mengenai persepsi pengunjung terhadap Ekowisata Sungai Mudal. Pada sentimen positif, kata-kata unigram seperti "air", "bagus", "wisata", "banget", dan "jalan" mendominasi, menunjukkan apresiasi terhadap kualitas air, keindahan umum, dan pengalaman wisata secara keseluruhan. Lebih lanjut, analisis bigram positif memperkuat temuan ini dengan frasa seperti "air terjun", "spot foto", dan "sungai mudal", yang menyoroti atraksi spesifik yang dinikmati pengunjung. Sebaliknya, pada sentimen negatif, unigram yang sering muncul adalah "jalan", "foto", "bagus", dan "spot", yang menariknya beberapa kata juga muncul di sentimen positif, mengindikasikan kemungkinan konteks penggunaan yang berbeda atau aspek tertentu dari "jalan" dan "spot foto" yang dikeluhkan. Analisis bigram negatif seperti "spot foto", "tiket masuk", dan "foto bayar" memberikan petunjuk lebih spesifik mengenai area yang mungkin memerlukan perhatian, seperti pengelolaan spot

foto, informasi tiket, atau adanya biaya tambahan untuk berfoto yang mungkin tidak disukai sebagian pengunjung.

Evaluasi model klasifikasi sentimen menggunakan SGD Classifier menunjukkan kinerja yang cukup baik dan stabil. Berdasarkan hasil validasi 10-fold cross validation, model mencapai rata-rata akurasi sebesar 0.8550 ± 0.0118, rata-rata recall 0.8550 ± 0.0118, dan rata-rata AUC 0.8805 ± 0.0113. Kurva ROC juga mengonfirmasi kemampuan diskriminatif model dengan nilai Mean ROC (AUC) sebesar 0.88 ± 0.03, yang menandakan bahwa model memiliki kemampuan yang baik dalam membedakan antara ulasan positif dan negatif. Kestabilan performa antar fold, seperti yang terlihat pada grafik Log Loss Across K-Folds, menunjukkan bahwa model tidak mengalami *overfitting* yang signifikan pada salah satu *fold* data dan mampu melakukan generalisasi dengan cukup konsisten. Hasil ini mengindikasikan bahwa model SGD yang dilatih cukup efektif untuk tugas klasifikasi sentimen pada dataset ulasan Ekowisata Sungai Mudal.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan analisis yang dilakukan, dapat disimpulkan bahwa pengunjung Ekowisata Sungai Mudal secara umum mengapresiasi aspek-aspek alamiah seperti kualitas "air", "air terjun", dan keindahan "sungai mudal" itu sendiri, serta fasilitas pendukung seperti "spot foto". Namun, beberapa area menjadi perhatian dalam sentimen negatif, terutama terkait dengan "jalan", pengelolaan "spot foto", serta isu "tiket masuk" dan potensi "foto bayar". Model klasifikasi sentimen menggunakan SGD Classifier menunjukkan kinerja yang baik dan stabil dengan rata-rata akurasi sekitar 85.5%, mengindikasikan efektivitasnya dalam mengidentifikasi dan membedakan sentimen positif dan negatif dari ulasan pengunjung untuk Ekowisata Sungai Mudal.

5.2. Saran

Bagi pengelola Ekowisata Sungai Mudal, disarankan untuk mempertahankan dan terus menonjolkan kualitas atraksi utama, seperti keindahan alam dan air yang diapresiasi pengunjung. Perhatian lebih lanjut sebaiknya diberikan pada aspek yang sering muncul dalam sentimen negatif, seperti perbaikan atau klarifikasi mengenai kondisi "jalan", optimalisasi pengelolaan "spot foto", serta transparansi terkait "tiket masuk" dan biaya tambahan lainnya untuk meningkatkan kepuasan pengunjung. Untuk penelitian selanjutnya, dapat dieksplorasi penggunaan algoritme klasifikasi lain, optimasi hyperparameter, atau analisis lebih mendalam terhadap kata-kata dengan konteks ganda yang muncul di kedua sentimen guna mendapatkan pemahaman yang lebih kaya.

DAFTAR PUSTAKA

Antonio, V., Efendi, S., & Mawengkang, H. (2022). Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent. *International Journal of Nonlinear Analysis and Applications*, *13*(1). https://doi.org/10.22075/ijnaa.2021.5735

Aulia, T.M.P., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (SVM) dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *Science and Information Technology Journal*, *4*(2), 139–145. https://doi.org/10.31598/sintechjournal.v4i2.762

Badan Pusat Statistik DIY. (2024). *Perkembangan Pariwisata Daerah Istimewa Yogyakarta, Agustus 2024*. https://jogjakota.bps.go.id/id/pressrelease/2024/10/01/902/perkembangan-pariwisata-daerah-istimewa-yogyakarta--agustus-2024.html

Beno, J. P. (2025). *ELECTRA and GPT-4o: Cost-Effective Partners for Sentiment Analysis* (No. arXiv:2501.00062). arXiv. https://doi.org/10.48550/arXiv.2501.00062

Ghatora, P. S., Hosseini, S. E., Pervez, S., Iqbal, M. J., & Shaukat, N. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. *Big Data and Cognitive Computing*, *8*(12), 199. https://doi.org/10.3390/bdcc8120199

Irsyad, M. (2020). Kondisi Potensi Wisata di Ekowisata Sungai Mudal Kabupaten Kulon Progo. *Jurnal Kepariwisataan: Destinasi, Hospitalitas Dan Perjalanan*, *4*(1), 29–39. https://doi.org/10.34013/jk.v4i1.36

Khairunnisa, S., Adiwijaya, A., & Faraby, S. A. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, *5*(2), 406. https://doi.org/10.30865/mib.v5i2.2835

Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). 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*, *16*(4), 28–38. https://doi.org/10.35969/interkom.v16i4.192

Liu, H., Chen, X., & Liu, X. (2022). A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis. *IEEE Access*, *10*, 32280–32289. https://doi.org/10.1109/ACCESS.2022.3160172

Marquis, P., Papini, O., & Prade, H. (Ed.). (2020). *A Guided Tour of Artificial Intelligence Research: Volume II: AI Algorithms*. Springer International Publishing. https://doi.org/10.1007/978-3-030-06167-8

Nisa, A., Darwiyanto, E., & Asror, I. 2019. Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi. *eProceedings of Engineering*, *6*(2), 8650-8658.

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education*, *29*(3), 572–589. https://doi.org/10.1002/cae.22253

Pangakis, N., & Wolken, S. (2024). *Knowledge Distillation in Automated Annotation: Supervised Text Classification with LLM-Generated Training Labels* (No. arXiv:2406.17633). arXiv. https://doi.org/10.48550/arXiv.2406.17633

Raza, G. M., Butt, Z. S., Latif, S., & Wahid, A. (2021). Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models. *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 1–6. https://doi.org/10.1109/ICoDT252288.2021.9441508

Riana, A. (2023). *Kualitas Layanan pada Ekowisata Sungai Mudal.* Diploma Thesis. STIM YKPN Yogyakarta

Shu, L., Xu, H., Liu, B., & Chen, J. (2022). *Zero-Shot Aspect-Based Sentiment Analysis* (No. arXiv:2202.01924). arXiv. https://doi.org/10.48550/arXiv.2202.01924

Wicaksono, A. (2020). New Normal Pariwisata Yogyakarta. *Kepariwisataan: Jurnal Ilmiah*, *14*(03), 139–150. https://doi.org/10.47256/kepariwisataan.v14i03.59

Yuan, S. (2023). Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools. *Academic Journal of Computing & Information Science*, *6*(9). https://doi.org/10.25236/AJCIS.2023.060904

Zakariyah, M., & Zaky, U. (2022). Analysis of Machine Learning Algorithm for Sleep Apnea Detection Based on Heart Rate Variability. *JUITA : Jurnal Informatika*, *10*(2), 173. https://doi.org/10.30595/juita.v10i2.14575