

# ANALISIS SENTIMEN PADA POSTINGAN MEDIA SOSIAL TWITTER DAN INSTAGRAM TERKAIT MANGROVE SURABAYA

Noor Fitria Azzahra

Manajemen Informatika, Universitas Merdeka Madiun, Jl. Serayu 79, Madiun , 63133

E-mail: noor@unmer-madiun.ac.id

**Abstract**— *Social media has emerged as a significant platform for individuals to express opinions and share experiences, extending to discussions on tourist attractions like Surabaya's mangroves. Utilizing sentiment analysis, this study evaluates public perceptions of the mangrove tourist destination. Concurrently, topic modeling extracts prevalent themes from textual data to discern evaluation aspects of Surabaya's mangroves. Employing sentiment lexicons and shadow labeling techniques, the analysis reveals a balanced sentiment distribution but susceptibility to misclassifications due to limited manually labeled data. Conversely, Latent Dirichlet Allocation (LDA) identifies three dominant topics: the attraction's appeal with culinary exploration, concerns over plastic pollution alongside distressing incidents like infant mortality, and issues of plastic waste accumulation and deforestation. Furthermore, after experimenting with various training models including LSTM, B-LSTM, machine learning algorithms (SVM, GNB, LR, LDA, and DTREE), and ensemble learning techniques (max voting, averaging, weighted averaging), the most optimal performance was achieved with the LSTM model, yielding an accuracy of 0.78. This underscores the utility of computational techniques in understanding public sentiment and thematic concerns surrounding Surabaya's mangrove ecosystem for tourism management and environmental conservation stakeholders and policymakers.*

**Keywords**—: surabaya mangrove, social media sentiment analysis, topic extraction

## I. PENDAHULUAN

Dalam beberapa tahun terakhir, penggunaan platform media sosial secara signifikan telah mengubah lanskap wacana publik, memungkinkan individu untuk secara bebas mengungkapkan pendapat dan berbagi pengalaman tentang berbagai topik, termasuk pengalamannya dalam mengunjungi suatu lokasi wisata. Perubahan dinamika komunikasi ini menegaskan pentingnya memahami dan menganalisis sentimen publik terhadap lokasi wisata, karena hal ini secara langsung memengaruhi strategi manajemen pariwisata dan upaya konservasi lingkungan. Mangrove Surabaya telah menarik perhatian tidak hanya karena keanekaragaman hayatinya tetapi juga karena signifikansinya sebagai objek wisata. Memahami persepsi publik terhadap hutan bakau ini sangat penting bagi para pemangku kepentingan dalam merumuskan praktik pariwisata yang berkelanjutan dan langkah-langkah konservasi.

Media sosial memainkan peran penting dalam industri pariwisata, pengelola objek wisata dapat berinteraksi dengan calon wisatawan maupun melakukan evaluasi dari pengalaman yang disampaikan wisatawan selama berada di objek wisata tersebut (Hvass and Munar, 2012). Pemanfaatan media sosial dalam industri pariwisata adalah faktor penting dalam perkembangan industri yang dapat memengaruhi berbagai aspek, salah satunya adalah proses pengambilan keputusan di kalangan wisatawan (Nor Azazi and Mohammed Shaed, 2020). Hal serupa disampaikan oleh (Gaffar *et al.*, 2022) dalam penelitiannya yang membahas peran media sosial dalam industri pariwisata, dalam hal pemasaran untuk meningkatkan minat wisatawan mengunjungi destinasi pariwisata. Dengan demikian, pemodelan topik dari konten Instagram dan Twitter dapat memberikan wawasan berharga dalam memahami preferensi dan tren wisatawan, serta membantu dalam merencanakan strategi pemasaran pariwisata yang efektif.

Pemodelan topik dari konten pada platform Instagram dan Twitter untuk objek wisata merupakan sebuah pendekatan yang penting dalam memahami preferensi dan perilaku wisatawan. Dalam konteks ini, penelitian oleh (Egger and Yu, 2021) menyoroti identifikasi struktur semantik tersembunyi dalam data Instagram menggunakan pemodelan topik. Mereka melakukan penelitian empiris untuk memverifikasi topik yang diidentifikasi dalam konteks pariwisata. Selain itu, penelitian oleh (Santoso, Sulistyawati and Vydia, 2022) membahas peran Instagram dalam promosi pariwisata, menunjukkan bahwa platform ini memiliki peran penting dalam mempromosikan destinasi pariwisata.

Dalam mengidentifikasi topik dari konten pariwisata di media sosial, metode pemodelan topik seperti Latent Dirichlet Allocation (LDA) sering digunakan. Referensi oleh (Habibi *et al.*, 2021) dan (Egger and Yu, 2021) menggunakan LDA untuk menganalisis konten terkait pariwisata di Instagram. Mereka

Website : <http://pilar.unmermadiun.ac.id/index.php/pilarteknologi>

menyoroti pentingnya pemodelan topik dalam memahami beragam topik yang dibahas dalam konten pariwisata di platform media sosial.

Permasalahan penelitian utama yang diangkat dalam studi ini berkisar pada perlunya mengevaluasi secara komprehensif sentimen publik terhadap objek wisata Mangrove Surabaya yang dihasilkan dari media sosial, Instagram dan Twitter (X). Meskipun penelitian sebelumnya telah membentuk landasan dengan menggunakan metodologi analisis sentimen dan pemodelan topik, masih ada kesenjangan dalam memahami sentimen dan aspek evaluasi spesifik yang berkaitan dengan Mangrove Surabaya. Oleh karena itu, studi ini bertujuan untuk berkontribusi pada literatur yang ada dengan menggunakan kombinasi teknik analisis sentimen dan pemodelan topik untuk mengungkap berbagai persepsi masyarakat mengenai Mangrove Surabaya. Melalui studi ini, kami berharap dapat memberikan wawasan berharga bagi pembuat kebijakan dan para pemangku kepentingan untuk merumuskan strategi efektif bagi pengembangan pariwisata yang berkelanjutan dan pelestarian ekologis di ekosistem Mangrove Surabaya.

## II. METODE PENELITIAN

### A. Dataset

Dataset diperoleh melalui proses pengambilan data dari platform Instagram dan Twitter menggunakan kata kunci tertentu, di antaranya adalah sebagai berikut:

1. Untuk Instagram: "mangrove gunung anyar", "mangrove medokan ayu", "mangrove wonorejo", "mangrove surabaya", dan "lokasi wisata mangrove gunung anyar".
2. Untuk Twitter: "mangrove surabaya", "mangrove gunung anyar", "mangrove medokan ayu", dan "mangrove wonorejo".

Proses pencarian berdasarkan kata kunci menghasilkan tautan, kemudian dilakukan ekstraksi konten dari link tersebut, contoh dari konten ini ditunjukkan pada Gambar 1.

Awalnya nyaman, Ujung2 nya paylater #🤔 #🤔\n\n#wisatagratis #wisatasurabaya #wisatamangrove #mangrovegununganyar #ekowisatamangrove #mangrovesurabaya #bosemmewah #wisatadekatrumah #reelinstagram #video #photo\n\n#

Gambar 1. Konten Instagram yang Mengandung Kata Kunci

### B. Metode Penelitian

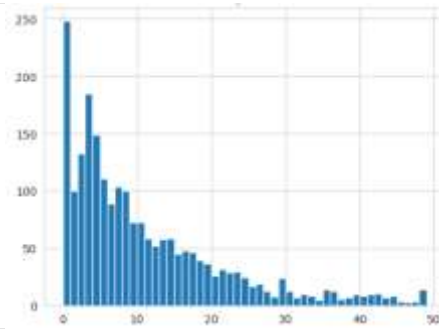
#### 1. Praproses Data

- **Konversi emoji**  
*Library emoji* digunakan untuk mengkonversi emoji menjadi teks dalam bahasa inggris. Sehingga tahap praproses berikutnya adalah menghapus *stopwords* bahasa inggris dan menggunakan daftar sentimen bahasa inggris.
- **Penghapusan karakter**  
Hasil *scraping* terlihat masih mengandung karakter yang tidak diinginkan, seperti yang terlihat dalam Gambar 1. Oleh karena itu, perlu dilakukan penghapusan karakter yang tidak diperlukan.
- **Konversi kata alay**  
Dilakukan konversi kata alay menjadi kata baku dengan memanfaatkan data dalam kamus alay. Hal ini diharapkan dapat memperkuat pemodelan untuk menangkap seluruh kata dalam data.
- **Penghapusan *stopwords***  
*Stopwords* merupakan kata-kata yang dianggap tidak memiliki makna yang signifikan, umumnya terdiri dari kata konjungsi atau kata depan. Contoh *stopwords* adalah "dari" dan "yang".
- **Analisis terhadap kata-kata yang sering muncul**  
Dilakukan analisis terhadap kata-kata yang sering muncul dalam dataset. Kata-kata yang tidak relevan dengan analisis dihapus, seperti nama tempat mangrove, "mangrove", "surabaya", "gunung", dan "wonorejo".
- **Penghapusan data tidak terkait**  
Dalam proses *scraping*, beberapa data tidak relevan didapatkan, terutama terkait dengan *online shop*. Pendekatan yang digunakan untuk menghapus data ini adalah dengan mendeteksi keberadaan nomor telepon dan beberapa kata yang merujuk pada promosi. Sebanyak 544 data terdeteksi sebagai data yang mengandung konten promosi.
- **Analisis kuantitatif terhadap data**

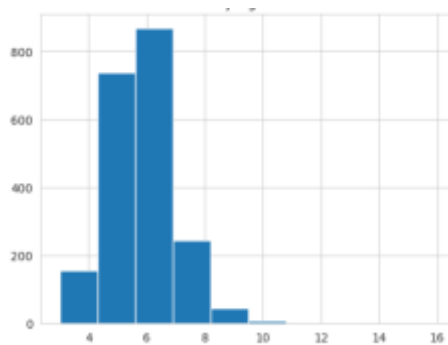
Website : <http://pilar.unmermadiun.ac.id/index.php/pilarteknologi>

Dilakukan analisis kuantitatif terhadap data, termasuk persebaran jumlah huruf, jumlah kata, dan panjang rata-rata kata. Distribusi hasil analisis dapat dilihat dalam grafik yang disertakan. Hasil analisis kuantitatif dataset antara lain sebagai berikut:

- Distribusi jumlah kata pada setiap data 0-49 kata.
- Distribusi Panjang kata rata-rata setiap data adalah 5-6 kata, yang artinya pendapat yang disampaikan cukup singkat.
- Pada analisis bigram yang divisualisasikan pada Gambar 4 dapat diketahui kata kunci berhubungan antara lain: “face”, “eyes”, “smiling”, “heart”, “face”, “tears”, “keren”.



Gambar 2. Distribusi Jumlah Kata dalam Satu Konten



Gambar 3. Distribusi Panjang Kata Rata-rata



Gambar 4. Visualisasi Wordcloud Kata yang Sering Muncul

## 2. Analisis Sentimen

### a. Pendekatan daftar sentimen kata

Dalam penilaian sentimen, digunakan dua bahasa, yakni bahasa Inggris dan Indonesia. Perhitungan sentimen didasarkan pada jumlah kemunculan kata dalam sentimen, yang dikalikan dengan nilai bobot yang telah ditentukan. Kata-kata dengan sentimen negatif akan diberi bobot negatif, sementara kata-kata dengan sentimen positif akan diberi bobot positif. Apabila nilai sentimen adalah 0, maka data memiliki sentimen netral. Jika nilai sentimen kurang dari 0, data memiliki sentimen negatif, sedangkan jika nilai sentimen lebih dari 0, data memiliki sentimen positif.

Website : <http://pilar.unmermadiun.ac.id/index.php/pilarteknologi>

b. Pendekatan *shadow labelling*

Dilakukan proses pemberian label secara manual terhadap 200 data. Model BLSTM (*Bidirectional LSTM*) dikembangkan menggunakan 200 data tersebut, yang akan digunakan sebagai model prediksi untuk data lainnya.

3. Pemodelan Topik

Digunakan metode LDA (*latent dirichlet allocation*) merupakan metode yang populer dan sering digunakan dalam pemodelan topik. Pada proses ini, dilakukan *hyperparameter tuning* terhadap parameter jumlah topik, alpha, beta, validation sets, dan corpus title. Proses ini diharapkan bisa mendapatkan topik dengan nilai koherensi terbaik.

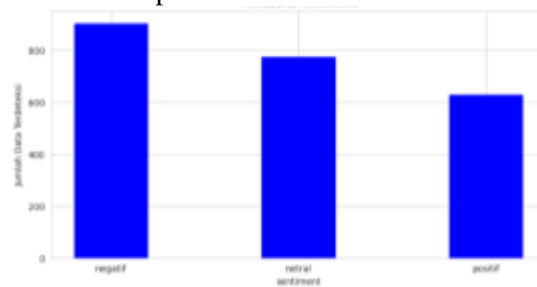
4. Pembuatan Model Prediksi

Model yang dibuat adalah LSTM, BLSTM, SVM, DTREE, GNB, LDA, LR, max voting, averaging, dan weighted averaging. Nilai evaluasi yang digunakan adalah akurasi.

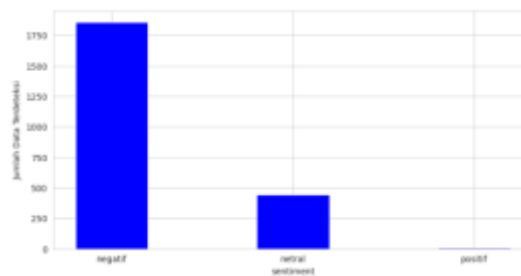
### III. HASIL DAN PEMBAHASAN

#### A. Distribusi Sentimen

Sebagaimana dijelaskan pada bab sebelumnya, dilakukan dua pendekatan dalam menganalisis sentimen, yakni dengan daftar sentimen dan *shadow labelling*. Distribusi sentimen dari masing-masing pendekatan divisualisasikan pada Gambar 4 dan Gambar 5. Terlihat bahwa distribusi sentimen dengan pendekatan daftar sentimen lebih merata. Sedangkan, pendekatan *shadow labelling* menghasilkan distribusi yang sangat tidak merata, sehingga pendekatan *shadow labelling* tidak layak digunakan dalam proses ekstraksi topik, maupun pembuatan model prediksi.



Gambar 5. Distribusi Sentimen dengan Pendekatan Daftar Sentimen



Gambar 6. Distribusi Sentimen dengan Pendekatan *Shadow Labelling*

Hasil analisis sentimen dengan pendekatan daftar sentimen divisualisasikan dengan grafik *wordcloud*, Gambar 6, untuk mengetahui kata apa saja yang muncul dalam data yang diidentifikasi memiliki sentimen negatif. Meski, memiliki distribusi yang lebih merata, terdapat beberapa kesalahan klasifikasi sentimen. Hal ini disebabkan oleh pemberian bobot pada kata yang kurang tepat. Sebagai contoh, data Gambar 7 diklasifikasikan sebagai negatif. Meski tidak mengandung kata yang bernilai negatif, data ini terklasifikasikan sebagai negatif dikarenakan kata “balai”. Kata tersebut dalam daftar sentimen diberikan bobot negatif -3, dimana nilai ini cukup besar dan mengakibatkan menjadi negatif.



Gambar 7. Visualisasi Wordcloud pada Data Teridentifikasi Memiliki Sentimen Negatif

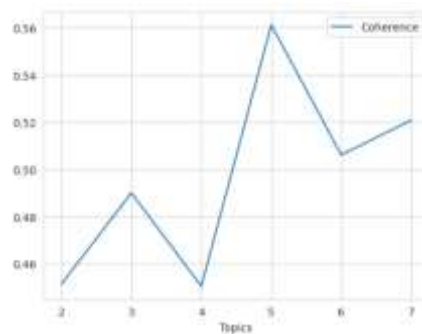
Tunjungan, Naik bus ssct di balai pemuda, alun-alun sby, north quey, hutan mangrove wonorejo, gunung anyar, romokalisari adventure, museum (iki seng paling recommend), kulineran di ampel, TP, angkirngan cak mis, Naik perahu malam Hari di kalimas, kenjeran, kbs, kodam, pusat thrift gembong

Gambar 8. Sampel Data Twitter yang Diklasifikasikan Bersentimen Negatif

### B. Ekstraksi Topik

Proses ekstraksi topik dimulai dengan menentukan jumlah topik. Jumlah topik yang ideal dapat ditentukan dengan menggunakan proses *hyperparameter tuning* yang menghasilkan nilai koherensi. Hasil dari proses tuning diilustrasikan pada Gambar 9. Dari grafik tersebut jumlah topik 5 memiliki koherensi paling tinggi, sehingga akan dilakukan ekstraksi topik sebanyak 5 topik. Hasil dari proses ekstraksi 5 topik diilustrasikan pada Gambar 10, antara lain:

- a. Topik 1: mengajak untuk menikmati lokasi wisata dengan makan, membuat dokumentasi.
- b. Topik 2: tentang perubahan iklim
- c. Topik 3: adanya sampah plastik dan penemuan jenazah bayi
- d. Topik 4: adanya sampah plastik dan penebangan pohon
- e. Topik 5: sulit untuk ditarik kesimpulan topik



Gambar 9. Nilai Koherensi Hasil *Hyperparameter*

```
[0,
 '0.022*wisata" + 0.014*face" + 0.013*ayo" + 0.011*kawasan" + '
 '0.010*coba" + 0.010*video" + 0.010*menarik" + 0.010*mencoba" + '
 '0.010*cek" + 0.009*makan"'),
(1,
 '0.028*minus" + 0.023*preservation" + 0.012*sih" + 0.012*looking" + '
 '0.012*support" + 0.012*denmark" + 0.012*carbon" + 0.012*climate" + '
 '0.012*works" + 0.012*change"'),
(2,
 '0.017*sampah" + 0.016*plastik" + 0.014*ditemukan" + 0.013*bayi" + '
 '0.013*jenazah" + 0.007*heart" + 0.006*fire" + 0.006*face" + '
 '0.004*bahan" + 0.004*sambal"'),
(3,
 '0.014*foto" + 0.014*take" + 0.014*sampah" + 0.013*plastik" + '
 '0.013*penebangan" + 0.012*terancam" + 0.012*nasib" + 0.008*face" + '
 '0.007*pohon" + 0.007*smiling"'),
(4,
 '0.009*wisata" + 0.008*gym" + 0.006*time" + 0.005*arah" + 0.005*pemkot" + '
 '+ 0.004*nby" + 0.004*kawasan" + 0.004*nsurabaya" + 0.004*kota" + '
 '0.004*xx"')]
```

Gambar 10. Ekstraksi 5 Topik



### C. Model Prediksi

Digunakan beberapa model prediksi sentimen dan dilakukan proses prediksi pada data yang telah dikumpulkan. Setiap model akan dievaluasi nilai akurasi yang dapat dilihat pada Grafik 1. Dari hasil tersebut, teridentifikasi tiga model dengan akurasi terbaik, yaitu LSTM, BLSTM, dan metode averaging. Meskipun model LSTM dan BLSTM menunjukkan akurasi terbaik, namun dapat dikategorikan sebagai model yang mengalami *overfitting*. Hal ini terlihat dari peningkatan nilai *loss* yang terjadi seiring dengan peningkatan kinerja model. Oleh karena itu, terdapat kekhawatiran bahwa model-model ini mungkin tidak dapat menangani analisis sentimen dengan baik pada kasus dunia nyata.

Model	Akurasi
LSTM	0,78
BLSTM	0,76
SVM	0,72
DTREE	0,72
LDA	0,66
LR	0,72
GNB	0,62
Max voting	0,71
Averaging	0,74
Weighted Averaging	0,74

Grafik 1. Hasil Akurasi Model Prediksi Sentimen

## VI. KESIMPULAN

Distribusi persepsi pengguna media sosial terhadap Mangrove Surabaya, yaitu 905 negatif, 631 netral, dan 776 positif, menunjukkan adanya persepsi negatif yang signifikan terhadap destinasi tersebut. Dari hasil ekstraksi topik yang dilakukan pada penelitian ini, dapat disimpulkan bahwa permasalahan utama yang dihadapi adalah sampah plastik dan berita penemuan jenazah bayi di lokasi wisata mangrove. Hal ini dapat menjadi bahan evaluasi bagi pihak pengelola wisata.

Proses pelabelan sentimen perlu mempertimbangkan pendekatan daftar sentimen kata-kata dalam bahasa Indonesia dan bahasa Inggris. Pada pendekatan ini, sering terjadi misklasifikasi, terutama pada data netral dan positif yang salah diklasifikasikan sebagai negatif. Pendekatan *shadow labelling* masih memiliki kekurangan karena jumlah pelabelan manual yang terlalu sedikit yang mengakibatkan ketidakseimbangan yang signifikan. Oleh karena itu, penelitian berikutnya dapat memperluas penggunaan *shadow labelling* dengan lebih banyak data yang telah dilabeli secara manual.

Model Prediksi dengan metode LSTM mencapai hasil terbaik dengan nilai akurasi sebesar 0.779697597026825. Meskipun demikian, perlu ditekankan bahwa model ini mengalami *overfitting*. Oleh karena itu, disarankan untuk menggunakan model averaging dengan harapan dapat mengatasi masalah ini secara lebih efektif dalam konteks dunia nyata.

## DAFTAR PUSTAKA

- Egger, R. and Yu, J. (2021) 'Identifying hidden semantic structures in Instagram data: a topic modelling comparison', *Tourism Review* [Preprint]. Available at: <https://doi.org/10.1108/TR-05-2021-0244>
- Gaffar, V. *et al.* (2022) 'Like, tag and share: bolstering social media marketing to improve intention to visit a nature-based tourism destination', *Tourism Review*, 77(2), pp. 451–470. Available at: <https://doi.org/10.1108/TR-05-2020-0215>.
- Habibi, M. *et al.* (2021) 'Topic Modelling of Germas Related Content on Instagram Using Latent Dirichlet Allocation (LDA)', in *Proceedings of the International Conference on Health and Medical Sciences (AHMS 2020)*. *International Conference on Health and Medical Sciences (AHMS 2020)*, Yogyakarta, Indonesia: Atlantis Press. Available at: <https://doi.org/10.2991/ahsr.k.210127.060>.

Website : <http://pilar.unmermadiun.ac.id/index.php/pilarteknologi>

- Hvass, K.A. and Munar, A.M. (2012) *The takeoff of social media in tourism*. Available at: <https://journals.sagepub.com/doi/10.1177/1356766711435978> (Accessed: 31 March 2024).
- Nor Azazi, N.A. and Mohammed Shaed, M. (2020) 'Social Media and Decision-Making Process among Tourist: A Systematic Review', *Jurnal Komunikasi: Malaysian Journal of Communication*, 36(4), p. 395–409. Available at: <https://doi.org/10.17576/JKMJC-2020-3604-24>.
- Santoso, A., Sulistyawati, A.I. and Vydia, V. (2022) 'Instagram as Social Media and His Role in The Tourism Promotion', *Indonesian Journal of Business and Entrepreneurship* [Preprint]. Available at: <https://doi.org/10.17358/ijbe.8.3.415>.