

PENGUNAAN METODE RANDOM FOREST UNTUK KLASIFIKASI SENTIMEN KOMENTAR MEDIA SOSIAL MENGENAI PENGELOLAAN SAMPAH DI KOTA TERNATE

SENTIMENT CLASSIFICATION OF SOCIAL MEDIA COMMENTS ON WASTE MANAGEMENT IN TERNATE CITY USING THE RANDOM FOREST METHOD

Rosihan¹, Alfianugrah A. Hi. Usman², Muhammad Fhadli³,
Muhammad Fikri Ramdhani⁴, Febrianti Darwin⁵
Fakultas Teknik, Program Studi Informatika
Universitas Khairun Maluku Utara
rosihan@unkhair.ac.id

Abstrak

Permasalahan pengelolaan sampah di Kota Ternate kerap menjadi pembahasan di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen publik dan menyediakan rekomendasi kebijakan berbasis data dengan mengembangkan model klasifikasi Random Forest. Penelitian ini menggunakan pendekatan kuantitatif melalui pengumpulan data komentar dari media sosial, preprocessing teks (tokenisasi, normalisasi, penghapusan stop word, stemming), ekstraksi fitur TF-IDF, dan klasifikasi sentimen dengan algoritma Random Forest. Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Model Random Forest yang dihasilkan mencapai akurasi keseluruhan 59,80%. Nilai presisi tertinggi diperoleh pada kelas positif (0,80), sedangkan recall tertinggi pada kelas negatif (0,9914). Analisis lebih lanjut mengidentifikasi kata kunci dominan seperti "laut", "buang", dan "sampah". Meskipun akurasi model masih dapat ditingkatkan, penelitian ini berhasil memetakan sentimen masyarakat. Temuan ini dapat menjadi dasar empiris bagi pemerintah daerah untuk merumuskan kebijakan pengelolaan sampah yang lebih responsif.

Kata Kunci: Sampah, *Random Forest*, Sentimen, Media Sosial, Kota Ternate

Abstract

Waste management issues in Ternate City are frequently discussed on social media. This study aims to analyze public sentiment and provide data-driven policy recommendations by developing a Random Forest classification model. A quantitative approach was employed, involving the collection of social media comment data, text preprocessing (tokenization, normalization, stop word removal, stemming), TF-IDF feature extraction, and sentiment classification using the

Random Forest algorithm. Model evaluation utilized accuracy, precision, recall, and F1-score metrics. The developed Random Forest model achieved an overall accuracy of 59.80%. The highest precision was obtained for the positive class (0.80), while the highest recall was for the negative class (0.9914). Further analysis identified dominant keywords such as "sea", "dispose", and "waste". Although the model's accuracy can be improved, this study successfully mapped public sentiment. These findings can serve as an empirical basis for the local government to formulate more responsive waste management policies.

Keywords: Trash, Random Forest, Sentiment, Social Media, Ternate City

PENDAHULUAN

Permasalahan sampah merupakan salah satu tantangan besar yang dihadapi oleh kota-kota di Indonesia, termasuk Kota Ternate. Sebagai kota pesisir yang dikelilingi oleh laut, Ternate menghadapi masalah serius terkait pengelolaan sampah, terutama sampah plastik yang mencemari ekosistem laut. Menurut data dari Kementerian Lingkungan Hidup dan Kehutanan (KLHK), volume sampah di Indonesia mencapai 67,8 juta ton per tahun, dengan sekitar 15% di antaranya merupakan sampah plastik yang tidak terkelola dengan baik [1]. Dampak dari sampah yang tidak terkelola ini tidak hanya merusak lingkungan, tetapi juga mengancam sektor pariwisata dan perikanan yang menjadi tulang punggung perekonomian Kota Ternate.

Media sosial telah menjadi platform yang efektif bagi masyarakat untuk menyampaikan pendapat, keluhan, dan saran terkait berbagai isu, termasuk pengelolaan sampah. Komentar-komentar di media sosial dapat menjadi sumber data yang kaya untuk memahami sentimen masyarakat

terhadap kebijakan dan praktik pengelolaan sampah. Namun, volume data yang besar dan beragamnya ekspresi bahasa dalam komentar media sosial membuat analisis manual menjadi tidak efisien. Oleh karena itu, diperlukan metode otomatis yang dapat mengklasifikasikan sentimen dari komentar-komentar tersebut.

Random Forest (RF) adalah salah satu algoritma machine learning yang telah terbukti efektif dalam berbagai tugas klasifikasi data teks [9], termasuk analisis sentimen. RF mampu menangani data dengan dimensi tinggi dan memiliki kemampuan untuk mengurangi overfitting, sehingga cocok digunakan untuk analisis data teks yang kompleks seperti komentar media sosial [2]. Dengan memanfaatkan RF, diharapkan dapat dilakukan klasifikasi sentimen secara akurat terhadap komentar-komentar masyarakat di media sosial mengenai pengelolaan sampah di Kota Ternate

Kebaruan penelitian ini terletak pada penerapan Random Forest untuk analisis sentimen pengelolaan sampah pada konteks lokal Kota Ternate. Penelitian sebelumnya masih terbatas pada metode tertentu dan belum banyak yang mengkaji penggunaan Random Forest pada konteks pengelolaan sampah di tingkat lokal seperti Kota Ternate.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pengelolaan sampah di Kota Ternate menggunakan metode Random Forest serta memberikan rekomendasi kebijakan berbasis data

LANDASAN TEORI

1. Analisis Sentimen

Analisis sentimen merupakan proses untuk mengidentifikasi opini, emosi, atau sikap seseorang terhadap suatu objek, layanan, atau kebijakan[6][8]. Analisis sentimen terdiri dari tiga level utama yaitu, Sentimen dokumen, Sentimen kalimat, Sentimen aspek. Model analisis sentimen biasanya memanfaatkan pendekatan machine learning, leksikon, atau gabungan keduanya

2. Media Sosial sebagai Sumber Data
Media sosial seperti Facebook, Instagram, dan Twitter merupakan platform tempat masyarakat mengekspresikan opini secara bebas dan terbuka. Menurut Kaplan &

Haenlein (2010), karakteristik media sosial yang dinamis menjadikannya sumber data yang kaya untuk analisis perilaku dan persepsi publik.

3. Preprocessing Teks

Preprocessing merupakan tahap penting dalam NLP untuk meningkatkan kualitas data sebelum diproses lebih lanjut.

4. Term Frequency–Inverse Document Frequency (TF-IDF), TF-IDF adalah teknik pembobotan kata yang digunakan untuk merepresentasikan teks dalam bentuk numerik[7].

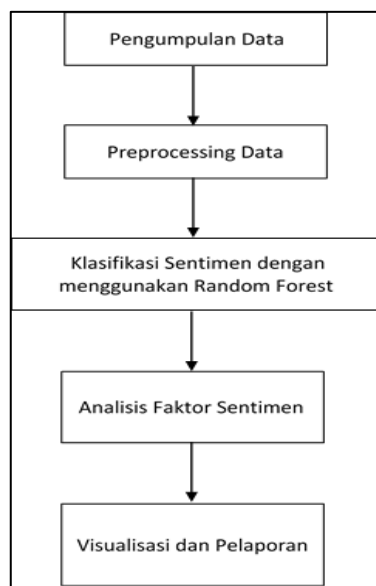
5. Machine Learning untuk Klasifikasi, Klasifikasi merupakan teknik untuk memetakan input ke dalam kategori tertentu dalam machine learning[5]. Dalam NLP, klasifikasi digunakan untuk menentukan apakah teks termasuk sentimen positif, negatif, atau netral.

6. Random Forest

Random Forest adalah algoritma ensemble learning yang diperkenalkan oleh Breiman (2001)[9]. Metode ini membangun banyak decision tree (hutan) dan hasil akhir ditentukan berdasarkan voting mayoritas (classification) atau rata-rata (regression).

7. Pengelolaan Sampah Kota Ternate
Kota Ternate termasuk daerah dengan volume sampah yang meningkat setiap tahun, dengan proporsi terbesar berasal dari rumah tangga dan aktivitas pesisir. Pengelolaan sampah yang kurang optimal sering menjadi sorotan masyarakat di media sosial, sehingga analisis sentimen perlu dilakukan untuk mendukung evaluasi kebijakan pemerintah. Analisis sentimen pada media sosial telah banyak dikembangkan menggunakan berbagai pendekatan machine learning dan deep learning[5]

Alur Penelitian



Gambar1. Diagram Alir Penelitian

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan sebagai berikut:

1. Pengumpulan data komentar media sosial (Twitter, Facebook, Instagram) menggunakan API.
2. Preprocessing data (pembersihan, tokenisasi, dan normalisasi).
3. Ekstraksi fitur menggunakan TF-IDF (Term Frequency-Inverse Document Frequency)[7].
4. Klasifikasi sentimen menggunakan algoritma *Random Forest*[5].
5. Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score.
6. Analisis hasil klasifikasi untuk mengidentifikasi pola sentimen dan faktor yang memengaruhinya.

Penelitian ini akan dilakukan melalui beberapa tahapan yang terstruktur untuk mencapai tujuan yang telah ditetapkan. Berikut adalah Gambar1 prosedur penelitian yang akan dilaksanakan.

Analisis Data/Sistem

Tahap 1: Pengumpulan Data

Tujuan: Mengumpulkan data komentar media sosial yang relevan dengan pengelolaan sampah di Kota Ternate.

Aktivitas:

- a. Identifikasi Platform Media Sosial: Memilih platform media sosial seperti Twitter, Facebook, dan Instagram yang banyak digunakan oleh masyarakat Ternate.

- b. Pengambilan Data: Menggunakan API (Application Programming Interface) dari platform media sosial untuk mengumpulkan komentar yang mengandung kata kunci terkait pengelolaan sampah (misalnya: "sampah Ternate", "kebersihan Ternate", "limbah plastik").
- c. Penyimpanan Data: Menyimpan data dalam format yang dapat diolah, seperti CSV atau JSON.

Tahap 2: Preprocessing Data

Tujuan: Membersihkan dan mempersiapkan data teks untuk proses klasifikasi.

Aktivitas:

- a. Tokenisasi: Memecah teks komentar menjadi kata-kata individual.
- b. Penghapusan Stop Words: Menghilangkan kata-kata umum yang tidak memiliki makna signifikan (misalnya: "dan", "atau", "di").
- c. Stemming: Mengubah kata-kata ke bentuk dasarnya (misalnya: "mengelola" menjadi "kelola").
- d. Labeling Sentimen: Memberikan label sentimen (positif, negatif, netral) pada data training dengan bantuan ahli atau menggunakan dataset berlabel yang sudah ada.

Tahap 3: Klasifikasi Sentimen Menggunakan *Random Forest*

Tujuan: Mengklasifikasikan sentimen komentar media sosial menggunakan metode *Random Forest*.

Aktivitas:

- a. Pembagian Data: Membagi data menjadi data training (80%) dan data testing (20%).
- b. Pembuatan Model: Membangun model *Random Forest* menggunakan data training.
- c. Optimasi Model: Melakukan tuning parameter seperti jumlah pohon (*n_estimators*) dan kedalaman pohon (*max_depth*) untuk meningkatkan akurasi model.
- d. Evaluasi Model: Menguji model menggunakan data testing dan mengevaluasi performanya dengan metrik seperti akurasi, presisi, recall, dan F1-score.

Tahap 4: Analisis Faktor Sentimen

Tujuan: Menganalisis faktor-faktor yang memengaruhi sentimen masyarakat terhadap pengelolaan sampah.

Aktivitas:

- a. Identifikasi Kata Kunci: Menganalisis kata-kata yang sering muncul dalam komentar positif, negatif, dan netral.
- b. Analisis Topik: Menggunakan teknik seperti Latent Dirichlet Allocation (LDA) untuk

mengidentifikasi topik-topik utama dalam komentar.

- c. Korelasi dengan Kebijakan: Menghubungkan hasil analisis sentimen dengan kebijakan pengelolaan sampah yang sedang berjalan di Kota Ternate.

Tahap 5: Visualisasi dan Pelaporan

Tujuan: Menyajikan hasil penelitian dalam bentuk yang mudah dipahami oleh pemangku kepentingan.

Aktivitas:

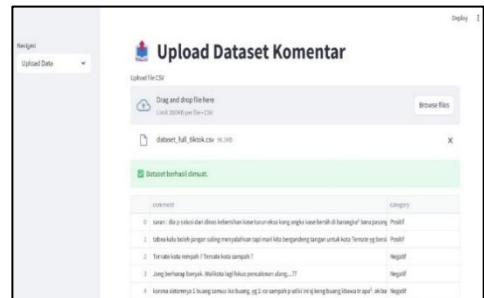
- a. Visualisasi Data: Membuat grafik dan dashboard interaktif untuk menampilkan distribusi sentimen dan faktor-faktor yang memengaruhinya.
- b. Pelaporan: Menyusun laporan penelitian yang mencakup metodologi, hasil, dan rekomendasi kebijakan

Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan akurasi sebesar 59,80%. Nilai ini menunjukkan bahwa performa model masih belum optimal. Hal ini disebabkan oleh karakteristik data media sosial yang tidak terstruktur dan menggunakan bahasa informal. Selain itu, kemungkinan ketidakseimbangan data antar kelas juga mempengaruhi hasil klasifikasi. Dibandingkan dengan penelitian sebelumnya yang menggunakan metode lain seperti SVM, performa Random Forest dalam penelitian ini masih relatif rendah

HASIL PEMBAHASAN

Tampilan Antarmuka Web Upload Dataset

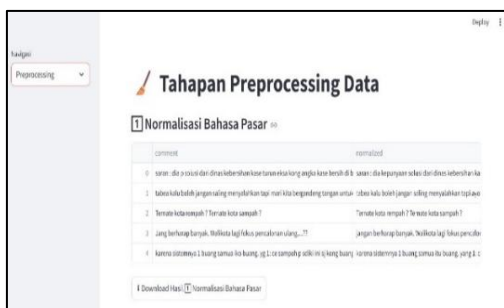
Halaman ini digunakan untuk mengunggah dataset komentar media sosial ke dalam sistem klasifikasi sentimen. Pengguna dapat mengunggah file berformat CSV melalui fitur Drag and Drop atau tombol Browse Files. Setelah file berhasil dimuat, data ditampilkan dalam tabel pratinjau yang berisi kolom komentar (comment) dan label sentimen (category), sebelum diproses lebih lanjut melalui tahap preprocessing. Berikut gambar 2 upload dataset.



Gambar 2. Tampilan Antarmuka Web untuk Fitur Upload Dataset Komentar

Tampilan Antarmuka Preprocessing Data (Normalisasi)

Halaman ini menunjukkan proses preprocessing data pada sistem klasifikasi sentimen komentar media sosial.



Gambar 3. Tampilan Antarmuka Web Tahap Preprocessing Data

Tahap Normalisasi Bahasa Pasar digunakan untuk mengubah kata atau frasa non baku menjadi bentuk baku agar lebih konsisten, sedangkan Case Folding digunakan untuk menyeragamkan seluruh huruf menjadi huruf kecil. Proses ini bertujuan untuk mempersiapkan data teks agar dapat dianalisis dengan lebih akurat menggunakan metode *Random Forest*. Gambar dapat dilihat pada gambar 3.

Tampilan Antarmuka Web Tahap Preprocessing Data (Case Folding)

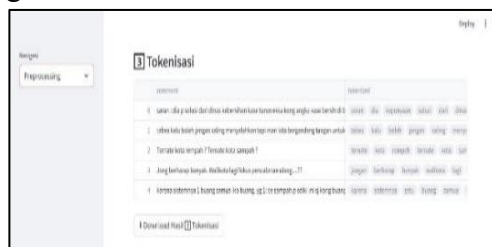
Halaman ini menampilkan tahap preprocessing data pada sistem klasifikasi sentimen komentar media sosial. Case Folding berfungsi menyeragamkan semua huruf menjadi huruf kecil untuk menghindari perbedaan penulisan kata yang sama. Gambar dapat dilihat pada gambar 4.



Gambar 4. Tampilan Antarmuka Web Tahap Preprocessing Data (Case Folding)

Tampilan Antarmuka Preprocessing Tokenisasi

Halaman ini menampilkan tahap preprocessing data pada sistem klasifikasi sentimen komentar media sosial. Tokenisasi digunakan untuk memecah teks komentar menjadi potongan kata (token) terpisah, sehingga memudahkan proses analisis selanjutnya. Kedua tahap ini membantu mempersiapkan data teks agar siap diproses oleh metode *Random Forest*. Dapat dilihat pada gambar 5.



Gambar 5. Tampilan Antarmuka Web Tahap Preprocessing Data (Tokenisasi)

Tampilan Antarmuka Tahap Preprocessing Data Final Preprocessed Text

Halaman ini menampilkan tahap terakhir preprocessing data pada sistem klasifikasi sentimen komentar media sosial. Bagian Final Preprocessed Text menampilkan hasil akhir preprocessing yang siap digunakan untuk tahap ekstraksi fitur

menggunakan metode TF-IDF dan klasifikasi dengan *Random Forest*. Dapat dilihat pada gambar 6.



Gambar 6. Tampilan Antarmuka Web Tahap Final Preprocessed Text

Tampilan Perhitungan TF-IDF

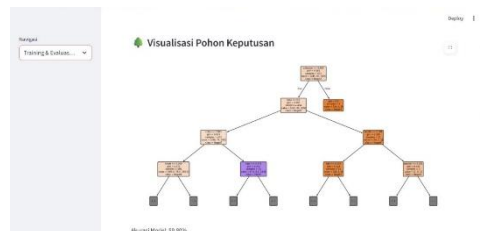
Halaman ini menunjukkan proses feature extraction menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) pada sistem klasifikasi sentimen komentar media sosial. TF-IDF mengubah teks hasil preprocessing menjadi representasi numerik berbentuk matriks, di mana setiap kolom mewakili kata unik dan setiap baris mewakili dokumen (komentar). Nilai pada matriks menunjukkan bobot pentingnya suatu kata dalam dokumen relatif terhadap keseluruhan dataset. Hasil ini digunakan sebagai input untuk model klasifikasi *Random Forest*. Dapat dilihat pada gambar 7.



Gambar 7. Tampilan Antarmuka Web Tahap Ekstraksi Fitur Menggunakan TF-IDF

Tampilan Antarmuka Visualisasi Pohon Keputusan

Gambar 8 menampilkan salah satu pohon keputusan yang dihasilkan dari model *Random Forest* pada sistem klasifikasi sentimen komentar media sosial. Setiap node menunjukkan kondisi pemisahan berdasarkan bobot kata dari hasil ekstraksi TF-IDF, disertai nilai gini impurity, jumlah samples, distribusi kelas (value), dan label kelas (class). Pohon ini membantu memvisualisasikan proses pengambilan keputusan oleh model dalam menentukan sentimen komentar.



Gambar 8. Visualisasi Pohon Keputusan Hasil Pelatihan Model *Random Forest*

Tampilan Antarmuka Hasil Evaluasi Model Klasifikasi Sentimen

Gambar 9 menampilkan hasil evaluasi model klasifikasi sentimen komentar media sosial menggunakan metrik precision, recall, f1-score, dan support untuk masing-masing kelas (Negatif, Netral, Positif). Model menghasilkan akurasi keseluruhan sebesar 59,80%. Nilai precision

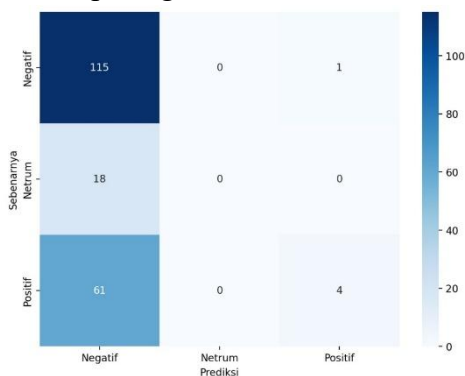
tertinggi diperoleh pada kelas Positif (0,8), sedangkan nilai recall tertinggi diperoleh pada kelas Negatif (0,9914). Bagian bawah gambar menampilkan confusion matrix yang memvisualisasikan distribusi prediksi model terhadap data uji.



Gambar 9. Hasil Evaluasi Model Klasifikasi Sentimen Menggunakan *Random Forest*

Tampilan Antarmuka *Confusion Matrix*

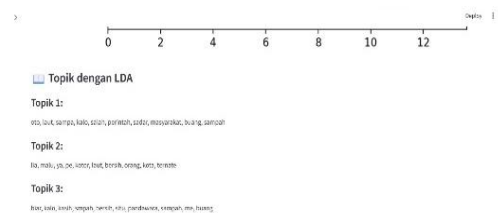
Evaluasi kinerja dilakukan dengan menggunakan confusion matrix untuk melihat distribusi prediksi model terhadap label sebenarnya. Dan menampilkan confusion matrix yang menunjukkan perbandingan antara kelas sebenarnya (sumbu vertikal) dan hasil prediksi (sumbu horizontal). Dapat dilihat pada gambar 10.



Gambar 10. Evaluasi Kinerja *Confusion Matrix*

Tampilan Antarmuka Topik dengan LDA

Gambar 11 merupakan tampilan web berjudul 'Topik dengan LDA' menampilkan hasil analisis topik yang terbagi menjadi tiga bagian utama: Topik 1 dengan kata kunci dominan oto, laut, sampa; Topik 2 yang menampilkan lia, malu, pe; serta Topik 3 yang menonjolkan biar, kalo, kasih. Setiap topik disajikan dalam format daftar teks sederhana tanpa grafik, dengan kata kunci dipisahkan oleh koma. Di bagian bawah halaman terdapat progress bar horizontal bernilai 0–12 yang menunjukkan indikator kemajuan, sementara tab browser terlihat di bagian atas antarmuka. Secara keseluruhan, antarmuka ini menggambarkan hasil pengelompokan topik otomatis menggunakan metode LDA (*Latent Dirichlet Allocation*) dengan fokus pada representasi kata kunci per kluster, tanpa visualisasi data kuantitatif.



Gambar 11. Topik dengan LDA

Hasil penelitian menunjukkan bahwa model *Random Forest* menghasilkan akurasi sebesar 59,80%. Meskipun akurasi model masih dapat ditingkatkan. Selain itu,

keterbatasan jumlah data dan proses labeling manual juga mempengaruhi kinerja model. Oleh karena itu, diperlukan pengembangan lebih lanjut untuk meningkatkan performa klasifikasi.

KESIMPULAN

Hasil dari Implementasi dan pembahasan yang dapat disimpulkan sebagai berikut:

1. *Random Forest* efektif untuk mengklasifikasi sentiment komentar media sosial. Model *Random Forest* yang dihasilkan mencapai akurasi keseluruhan 59,80%. Nilai presisi tertinggi diperoleh pada kelas positif (0,80), sedangkan recall tertinggi pada kelas negatif (0,9914)
2. Hasil analisis dapat menjadikan rekomendasi peningkatan kebijakan pengelolaan sampah dikota ternate.

Penelitian selanjutnya disarankan untuk menggunakan metode klasifikasi lain serta meningkatkan jumlah dan kualitas data guna memperoleh hasil yang lebih optimal

DAFTAR PUSTAKA

[1] D. Sebagai, S. Satu, P. Program, S. Teknik, I. Fakultas, and U. M. Ponorogo, "Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Jenjang Strata Satu (S1) Pada Program Studi Teknik

Informatika Fakultas Teknik Universitas Muhammadiyah Ponorogo," 2025.

- [2] V. No, D. Hal, and M. Alfidyah, "Analisis Sentimen Pengguna Aplikasi X terhadap Konflik antara Israel dan Palestina Menggunakan Algoritma Support Vector Machine Jurnal Statistika dan Data Sains," vol. 1, no. 1, pp. 8–14, 2025.
- [3] P. Studi, T. Industri, P. Sarjana, F. T. Industri, and U. I. Indonesia, "Person-Job Fit Menggunakan Decision Trees Algorithm (Studi Kasus PT PNR Perusahaan Jasa Pertambangan pada Divisi PL)," 2024.
- [4] M. Muharram, D. Hariyati, and L. R. Simanjuntak, "Efisiensi dan Efektivitas Pengadaan Barang / Jasa (E-Procurement) Pada LPSE dan E-Katalog di Dinas Lingkungan Hidup Provinsi," vol. 6, no. 3, pp. 1547–1559, 2025.
- [5] Yadav, A., & Vishwakarma, D. K. (2020). *Sentiment analysis using deep learning architectures: A review*. *Artificial Intelligence Review*, 53, 4335–4385. <https://doi.org/10.1007/s10462-020-09822-5>
- [6] Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment*

Analysis. Foundations and Trends in Information Retrieval.

- [7] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press
- [8] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- [9] Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45(1), 5–32, <https://doi.org/10.1023/A:1010933404324>