

ANALISIS SENTIMEN PENGGUNA TIKTOK TENTANG PROGRES PEMBANGUNAN IKN DENGAN METODE RANDOM FOREST

Siti Rihastuti¹, Afnan Rosyidi²

^{1,2}STMIK Amikom Surakarta, Sukoharjo,

¹siti@dosen.amikomsolo.ac.id

²afnan@dosen.amikomsolo.ac.id

Abstrak

Klasifikasi sentimen merupakan teknik analisis teks yang digunakan untuk mengidentifikasi dan mengkategorikan opini pengguna terhadap suatu aplikasi atau layanan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat tentang progres pembangunan IKN (Ibukota Nusantara) dengan algoritma Random Forest berdasarkan komentar pengguna platform Tiktok. Dataset diambil dari kaggle sebanyak 1472 komentar berbahasa Indonesia. Dataset yang digunakan terdiri dari komentar pengguna yang dikategorikan ke dalam sentimen positif dan negatif. Evaluasi dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score untuk mengetahui hasil dari klasifikasi sentimen pengguna. Pengujian metode Random Forest di Google Colab menunjukkan nilai akurasi sebesar 77%, precision 78%, recall 77% dan F1-score 77%. Dari hasil pengujian tersebut, metode Random Forest dinilai cukup baik dalam mengklasifikasikan sentimen pengguna Tiktok dalam menanggapi progres pemindahan IKN.

Kata Kunci: *Klasifikasi Sentimen, Tiktok, Random Forest, Komentar*

Abstract

Sentiment classification is a text analysis technique used to identify and categorize user opinions about an application or service. This study aims to classify public sentiment about the progress of the development of the IKN (Indonesian Capital) with the Random Forest algorithm based on comments from users of the Tiktok platform. The dataset was taken from Kaggle with 1472 comments in Indonesian. The dataset used consists of user comments categorized into positive and negative sentiments. The evaluation was carried out based on the accuracy, precision, recall, and F1-score metrics to determine the results of the user sentiment classification. Testing the Random Forest method on Google Colab showed an accuracy value of 77%, precision 78%, recall 77% and F1-score 77%. From these values, the Random Forest method is considered quite good in classifying Tiktok user sentiment in responding to the progress of the IKN relocation.

Keywords: *Sentiment Classification, Tiktok, Random Forest, Comments.*

1. Pendahuluan

Pemindahan Ibukota dari Jakarta ke Ibu Kota Nusantara (IKN) yang berada di wilayah Kalimantan Timur telah ditetapkan oleh Presiden Joko Widodo pada tanggal 29 April 2019 (Supian et al., 2024). Dalam proses pemindahannya, muncul beragam reaksi dari masyarakat terutama pengguna media sosial. Pemberitaan terkait pemindahan ibukota menimbulkan opini positif dan negatif. Salah satu media sosial yang umum digunakan di Indonesia adalah Tiktok. Pengguna Tiktok bebas menuliskan komentar atau pendapatnya pada suatu konten berisi

layanan, topik atau lembaga. Analisis sentimen masyarakat terhadap pemindahan ibukota yang sampai saat ini masih berjalan proses pembangunannya menarik untuk diteliti, untuk mengetahui klasifikasi sentimen masyarakat lebih dominan mendukung (positif) atau kurang setuju (negatif).

Analisis sentimen adalah teknik dalam *Natural Language Processing* (NLP) yang difungsikan untuk menentukan pendapat atau perasaan pengguna terhadap suatu produk atau layanan, salah satunya berupa aplikasi. Dalam konteks ini, metode *machine learning* bisa

digunakan untuk melakukan proses klasifikasi sentimen pengguna menjadi beberapa kategori yaitu positif atau negatif. Algoritma yang akan digunakan dalam klasifikasi sentimen adalah *Random Forest* (RF). *Random Forest* adalah metode berbasis *ensemble learning* yang membangun banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting*.

Penelitian ini bertujuan untuk menerapkan metode *Random Forest* dalam mengklasifikasikan sentimen atau opini terhadap komentar pengguna platform Tiktok tentang progres pembangunan IKN. Evaluasi dilakukan berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-score* guna mengetahui sentimen pengguna.

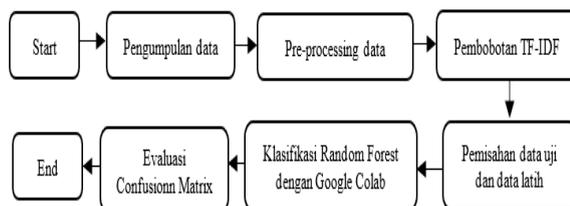
Metode *Naïve Bayes* dan *Support Vector Machine* (SVM) diterapkan untuk analisis sentimen tentang isu pemindahan IKN (Ibu Kota Nusantara). Dari hasil pengujian, SVM menghasilkan nilai akurasi lebih tinggi (94%) dibandingkan menggunakan metode *Naïve Bayes* (91%), ini menunjukkan bahwa algoritma SVM lebih baik dalam klasifikasi teks yang sifatnya kompleks (Supian et al., 2024). Teknik web scraping diterapkan dalam proses pengumpulan data penjualan handphone di marketplace Tokopedia. Data yg terkumpul meliputi jumlah unit terjual, total penjualan, harga rata-rata, serta merek dan lokasi penjualan terbanyak. Data tersebut kemudian dilakukan analisis menggunakan metode regresi linier untuk memprediksi penjualan. Hasil pengujian mengindikasikan nilai akurasi mencapai 90%, yang menunjukkan metode regresi linier cukup efektif dalam memprediksi penjualan (Guntara et al., 2024). Algoritma *Decision Tree* dan *KNN* digunakan untuk klasifikasi *cyberbullying* (perundungan melalui media digital) pada komentar platform X (Twitter) dengan dataset yang bersumber dari Kaggle. Evaluasi model yang digunakan meliputi *confusion matrix*, *ROC curve*, dan metrik evaluasi seperti akurasi, *precision*, dan *recall*. Hasil pengujian menunjukkan bahwa algoritma *KNN* menghasilkan akurasi lebih tinggi (75,99%) dibandingkan menggunakan algoritma *Decision Tree* (65%). *ROC AUC* (*Receiver Operating Characteristic - Area Under the Curve*) ini menunjukkan seberapa baik model dalam membedakan antara kategori *cyberbullying* dan *non-cyberbullying*) untuk *KNN* mencapai 81,51%, mengindikasikan performa klasifikasi yang lebih baik (Maradona et al., 2023). Menurut

(Hendrawan & Sela, 2024) analisis sentimen diperlukan untuk mengetahui bagaimana masyarakat merespon isu resesi global tahun 2023. Sebanyak 500 komentar di platform Youtube dikumpulkan dan diuji dengan metode *Long Short-Term Memory* (LSTM). Hasil pengujian menunjukkan LSTM dengan 50 epoch dan aktivasi *sigmoid* menghasilkan akurasi tertinggi pada 90% pada data latih dan 76% pada data uji. Hal tersebut menunjukkan metode LSTM dinilai cukup efektif dalam analisis sentimen komentar YouTube (Hendrawan & Sela, 2024). Ulasan pelanggan yang berada di platform digital sangat berpengaruh terhadap kelangsungan bisnis, salah satunya bisnis kecantikan. Analisis sentimen diterapkan untuk mengklasifikasikan ulasan pada *Google Review* klinik kecantikan Natasha. Ulasan dibedakan sebagai positif atau negatif. Pengujian data menggunakan algoritma SVM. Hasilnya, SVM menunjukkan akurasi tinggi dalam mengklasifikasikan sentimen ulasan pelanggan (Rizqi et al., 2024). Analisis sentimen diperlukan untuk memberikan gambaran umum tentang bagaimana masyarakat menilai kualitas konten televisi nasional. Teknik web scraping dengan *library Harvest1* diterapkan untuk mengumpulkan data dari Twitter yang menghasilkan 515.492 data sentimen. Data dibedakan menjadi sentimen positif, negatif dan netral. Pengujian menggunakan metode *Bi-LSTM* dengan *Word2Vec*. Hasil pengujian metode yang digunakan cukup efektif dalam klasifikasi sentimen pada data media sosial (Bouchra et al., 2018). Analisis sentimen di sosial media terus berlangsung, salah satunya di dunia olahraga. Setelah pertandingan besar, pemain sepak bola sering menjadi sasaran kritik tajam, ujaran kebencian, dan pelecehan dari netizen. Sejumlah 9.060 tweet dikumpulkan dan dipisah menjadi 80% data latih, 10% data validasi dan 10% data uji. Metode yang digunakan yaitu LSTM. Model terbaik mencapai nilai akurasi 78% setelah optimasi dengan *precision*, *recall*, dan *F1-score* masing-masing sebesar 78%. Hasil pengujian mengindikasikan bahwa model LSTM cukup efektif untuk klasifikasi sentimen (Widiyantoro & Prasetyo, 2025). Aplikasi web scraping dikembangkan untuk otomatisasi pengumpulan data di marketplace Tokopedia untuk membantu penjual memahami dinamika pasar. Hasil menunjukkan aplikasi dapat bekerja dengan baik dalam mengumpulkan data harga produk di Tokopedia secara otomatis (Alhafiz & Adiguna,

2024). Implementasi algoritma Naive Bayes (NB) dan Support Vector Machine (SVM) untuk menganalisis sentimen tentang isu cyberbullying dalam bahasa Indonesia dan Inggris di platform X (Twitter). Hasil pengujian menunjukkan metode Naive Bayes (NB) lebih tinggi akurasi (87%) dalam klasifikasi cyberbullying dibandingkan SVM (86%)(Machine, 2025). Penerapan LSTM dengan Word2Vec dalam menganalisis dataset berupa komentar masyarakat terhadap kebijakan dan layanan pemerintah daerah pada platform Instagram. Hasil pengujian menunjukkan metode yang digunakan cukup efektif dalam klasifikasi opini publik di Instagram pemerintah(Arwindarti et al., 2023). Aplikasi PeduliLindungi menerima beragam komentar dari pengguna di ulasan platform Google Play Store. Analisis sentimen dibuat untuk membantu pengembangan kualitas pelayanan berdasarkan komentar pengguna. Metode yang digunakan yaitu LSTM. Hasil pengujian menunjukkan nilai akurasi sebesar 92,51%, mengindikasikan LSTM performa tinggi dalam mengklasifikasikan sentimen pengguna(Mutmatimah et al., 2024).

2. Metode Penelitian

Tahapan penelitian ini menggunakan metodologi *Knowledge Discovery in Database* (KDD), yang meliputi tahapan yaitu pemilihan data, *pre-processing data*, *transformation* (pembobotan), *data mining* dan evaluasi, seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. Pengumpulan data

Pada tahap ini dilakukan pengumpulan data dengan mengambil data di situs <https://www.kaggle.com/datasets/najmarafifah/analisis-sentimen-komentar-tiktok-progres-ikn>. Sebanyak 1472 komentar terdiri dari komentar berupa teks dan label (terdiri dari 780 komentar negatif dan 692 komentar positif).

b. Pre-processing data

Langkah berikutnya yaitu pre-processing yaitu pembersihan data meliputi *cleansing*, *case folding*, *tokenizing*, *stopwords*, *stemming*, dan normalisasi(Herjanto & Carudin, 2024).

c. Pembobotan TF-IDF

Metode Tf (*Term Frequency*) adalah metode untuk menentukan bobot suatu dokumen dengan menghitung berapa kali istilah atau kata tertentu muncul dalam dokumen tersebut. Sedangkan ldf (*Inverse Document Frequency*) berfokus pada kemunculan suatu *term* pada seluruh koleksi teks dalam komentar. Dalam hal ini mengubah data bertipe nominal menjadi data numerik agar lebih mudah dalam memproses data mining.

d. Pemisahan data latih (data train) dan data uji (data test)

Setelah data siap, selanjutnya data dibedakan menjadi 2 (dua) bagian, yaitu data latih (data train) sebanyak 80% dan data uji (data test) sebanyak 20%.

e. Klasifikasi Random Forest

Random Forest merupakan teknologi pre-processing dalam klasifikasi fitur yang meningkatkan skalabilitas, akurasi, dan efisiensi, terutama dalam proses klasifikasi teks seperti pohon. Metode Random Forest akan diuji menggunakan Google Colab.

f. Evaluasi Confusion Matrix

Evaluasi dilakukan untuk mengukur efektivitas dan akurasi kinerja dari model Random Forest yang telah diterapkan. *Confusion matrix* digunakan untuk menyajikan informasi tentang *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)*, yang sangat membantu dalam mengevaluasi tingkat keakuratan hasil klasifikasi.

3. Hasil Penelitian dan Pembahasan

Penelitian ini berfokus pada komentar pengguna platform Tiktok dalam menanggapi progres pembangunan IKN. Dalam penelitian ini digunakan algoritma Forest Random sebagai model dalam klasifikasi.

4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat ditarik kesimpulan :

- a. Tahapan klasifikasi sentimen menggunakan random forest meliputi : pengumpulan data, pre-processing data, pembobotan TF-IDF, pemisahan data uji dan data latih, klasifikasi dengan random forest dan evaluasi confusion matrix.
- b. Hasil pengujian di Google Colab dengan metode Random Forest apada 1472 komentar yang terdiri dari 780 komentar negatif dan 692 komentar positif, menghasilkan akurasi sebesar 77%, precision 78%, recall 77% dan F1-score 77%. Dari hasil tersebut metode Random Forest dinilai cukup baik dalam mengklasifikasikan sentimen negatif yang memang benar negatif berdasarka komentar pengguna Tiktok dalam menanggapi proses pemindahan IKN.
- c. Untuk mendapatkan nilai akurasi yang lebih baik, disarankan menggunakan metode yang berbeda dan jumlah dataset antara label positif dan negatif berimbang

5. Daftar Pustaka

- Alhafiz, V., & Adiguna, M. A. (2024). PENGEMBANGAN APLIKASI WEB SCRAPING UNTUK PENGUMPULAN DATA HARGA PRODUK E-COMMERCE MENGGUNAKAN PYTHON (STUDI KASUS : TOKOPEDIA). *Jurnal Penelitian Ilmu Komputer*, 2(2), 1–13.
- Arwindarti, T., Setiawan, E. I., & Imron, S. (2023). Klasifikasi Sentimen Opini Publik Pada Instagram Pemerintah Kabupaten Bojonegoro Menggunakan LSTM. *Teknika*, 13(1), 1–9. <https://doi.org/10.34148/teknika.v13i1.699>
- Bouchra, F., Agus, I. M., Suarjaya, D., Kadek, N., & Rusjyanthi, D. (2018). Analisis Sentimen Masyarakat terhadap Tayangan Televisi Nasional menggunakan Metode Deep Learning. *Jurnal Buana Informatika*, 89–99.
- Guntara, R. G., Kashira, F. B., Amri, T. K., Restu, L. B., & Susanto, F. R. (2024). Analisis Penjualan Handphone di Tokopedia dengan Teknik Web Scraping Menggunakan Python pada Google Colab. *ULIL ALBAB : Jurnal Ilmiah Multidisiplin*, 3(4), 69–75. <https://journal-nusantara.com/index.php/JIM/article/view/3200>
- Hendrawan, A., & Sela, E. I. (2024). Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan LSTM. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(1), 587–593. <https://doi.org/10.35870/jimik.v5i1.526>
- Herjanto, M. F. Y., & Carudin, C. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2), 1204–1210. <https://doi.org/10.23960/jitet.v12i2.4192>
- Machine, S. V. (2025). Penerapan Algoritma Klasifikasi Naive Bayes dan Support Vector Machine untuk Analisis Sentimen Cyberbullying Bilingual di Aplikasi X Implementation of Naive Bayes and Support Vector Machine Classification Algorithms for Sentiment Analysis of Bilingual Cyb. *Sistemasi: Jurnal Sistem Informasi*, 14, 211–224.
- Maradona, M., Kusrini, K., & Alva Hendi Muhammad. (2023). Analisis Perbandingan Metode Decision Tree Dan K-Nearest Neighbor Untuk Klasifikasi Cyberbullying Pada Sosial Media Twitter. *Metik Jurnal*, 7(2), 47–61. <https://doi.org/10.47002/metik.v7i2.591>
- Mutmatimah, S., Khairunnas, & Khairunnisa. (2024). Metode Deep Learning LSTM dalam Analisis Sentimen Aplikasi PeduliLindungi. *Journal of Computers Sciences and Informatics*, 1(1), 9–19. <https://doi.org/10.34304/scientific.v1i1.231>
- Rizqi, M., Rustiawan, A., & Prasetyaningrum, P. T. (2024). Analisis Sentimen Terhadap Klinik Natasha Skincare di Yogyakarta Dengan Metode Google Review. *Journal of Information Technology Ampera*, 5(1), 2774–2121. <https://doi.org/10.51519/journalita.v5i1.556>
- Supian, A., Tri Revaldo, B., Marhadhi, N., Efrizoni, L., & Rahmaddeni, R. (2024). Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara. *Jurnal Ilmiah Informatika*, 12(01), 15–21. <https://doi.org/10.33884/jif.v12i01.8721>
- Widiyantoro, P., & Prasetyo, Y. D. (2025). Deteksi Cyberbullying pada Pemain Sepak Bola di Platform Media Sosial “ X ” Menggunakan Metode Long Short-Term Memory (LSTM). *Repeater : Publikasi Teknik Informatika Dan Jaringan*.