

ANALISIS SENTIMEN PUBLIK TERHADAP PEMINDAHAN IBU KOTA NEGARA DI TWITTER MENGGUNAKAN METODE KLASIFIKASI RANDOM FOREST DAN SMOTE

Agung Kharisma Hidayah¹, Yetman Erwadi², Sri Handayani³

¹ Teknik Informatika, Universitas Muhammadiyah Bengkulu

^{2,3} Sistem Informasi, Universitas Muhammadiyah Bengkulu

Jalan Bali Kota Bengkulu, Indonesia

kharisma@umb.ac.id

ABSTRAK

Pemindahan Ibu Kota Negara (IKN) ke Kalimantan Timur merupakan kebijakan nasional strategis yang memicu beragam respons dari masyarakat, khususnya di media sosial seperti Twitter. Berbagai opini publik tersebut penting untuk dianalisis guna mengetahui persepsi masyarakat terhadap kebijakan tersebut. Namun, karakteristik data Twitter yang tidak terstruktur dan ketidakseimbangan distribusi sentimen menjadi tantangan dalam proses klasifikasi sentimen secara otomatis. Penelitian ini bertujuan untuk membangun model analisis sentimen publik terhadap pemindahan IKN dengan pendekatan machine learning. Metode yang digunakan meliputi preprocessing data teks, ekstraksi fitur menggunakan TF-IDF, penyeimbangan data dengan teknik Synthetic Minority Over-sampling Technique (SMOTE), serta klasifikasi menggunakan algoritma Random Forest. Dataset terdiri dari 2.177 tweet berbahasa Indonesia yang telah diberi label sentimen positif dan negatif. Hasil pengujian menunjukkan bahwa model Random Forest mampu mengklasifikasikan sentimen dengan akurasi sebesar 95,35%, serta precision dan recall yang tinggi dan seimbang. Temuan ini menunjukkan bahwa kombinasi preprocessing, SMOTE, dan Random Forest efektif digunakan dalam mengklasifikasikan opini publik di media sosial terhadap isu strategis seperti pemindahan IKN.

Kata kunci : Analisis Sentimen, Twitter, Ibu Kota Negara, SMOTE, Random Forest

1. PENDAHULUAN

Pemindahan Ibu Kota Negara (IKN) Indonesia merupakan isu nasional yang telah menjadi perhatian publik sejak diumumkan secara resmi oleh Presiden Joko Widodo pada tanggal 26 Agustus 2019. Rencana pemindahan ibu kota dari Jakarta ke Kalimantan Timur, yang diberi nama “Nusantara”, menuai berbagai tanggapan dari masyarakat baik yang mendukung maupun menolak kebijakan tersebut [1][2][3].

Perbedaan pandangan ini banyak disuarakan melalui berbagai platform media sosial, terutama Twitter yang dikenal sebagai salah satu media real-time paling aktif di Indonesia [4]. Twitter memungkinkan masyarakat untuk mengemukakan opini secara terbuka dengan cepat, menggunakan hashtag seperti #IbukotaBaru, #IKN, atau #PindahIbuKota. Dengan demikian, platform ini menjadi sumber data yang sangat kaya untuk menganalisis persepsi publik terhadap isu-isu kebijakan nasional [5].

Pengguna internet sekarang bebas berbagi informasi hingga membentuk sebuah opini publik yang memiliki berbagai makna. Salah satu pendekatan yang banyak digunakan untuk memahami opini publik dari media sosial adalah analisis sentimen, yaitu proses otomatis untuk mengklasifikasikan teks ke dalam kategori sentimen seperti positif, negatif, dan netral [3][4]. Berbagai penelitian sebelumnya telah menerapkan metode Naïve Bayes [1][5], Support Vector Machine [6], dan K-Nearest Neighbor [7]

dalam melakukan klasifikasi sentimen terhadap opini masyarakat terkait pemindahan IKN.

Namun, sebagian besar dari penelitian-penelitian tersebut belum mempertimbangkan permasalahan ketidakseimbangan data (imbalanced data), yang mana jumlah data dalam kelas negatif biasanya jauh lebih besar dibandingkan kelas positif atau netral. Ketidakseimbangan ini dapat menyebabkan performa model klasifikasi menjadi kurang optimal [8][5]. Untuk mengatasi permasalahan tersebut, beberapa penelitian mulai mengadopsi teknik Synthetic Minority Over-sampling Technique (SMOTE) guna menyeimbangkan distribusi kelas [8].

Selain itu, sebagian besar penelitian sebelumnya masih mengandalkan metode klasifikasi konvensional seperti Naïve Bayes yang memiliki keterbatasan dalam menangani data dengan korelasi antar fitur. Dalam konteks ini, Random Forest muncul sebagai alternatif metode klasifikasi yang lebih kuat karena kemampuannya dalam menangani data yang kompleks dan memberikan hasil klasifikasi yang lebih stabil dan akurat [9][10].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen publik terhadap pemindahan Ibu Kota Negara di media sosial Twitter menggunakan metode klasifikasi Random Forest dan teknik penyeimbangan data SMOTE. Dengan pendekatan ini, diharapkan dapat diperoleh model klasifikasi yang lebih akurat dan representatif dalam menggambarkan persepsi masyarakat terhadap isu pemindahan IKN.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beragam penelitian telah dilakukan untuk menganalisis sentimen publik terhadap isu pemindahan Ibu Kota Negara (IKN), umumnya menggunakan media sosial sebagai sumber data dan algoritma klasifikasi seperti Naïve Bayes, Support Vector Machine (SVM), hingga LSTM. Setiawan et al. menggunakan pendekatan LSTM dan Word2Vec untuk menganalisis sentimen publik di Twitter, dengan hasil yang menunjukkan keunggulan dalam memahami konteks kalimat meskipun memerlukan sumber daya komputasi yang besar [6]. Sementara itu, Sari dan Wijaya menerapkan algoritma SVM untuk mengklasifikasikan sentimen terkait IKN, namun tidak mengatasi ketidakseimbangan kelas, yang berdampak pada bias prediksi. Rahmawati et al. menggabungkan SMOTE dan Naïve Bayes untuk menganalisis komentar YouTube, dan terbukti mampu meningkatkan akurasi pada kelas minoritas [8]. Adapun Wahyudi [11] serta Putri et al. [12] juga menggunakan Naïve Bayes, tetapi tidak menerapkan balancing data, sehingga memiliki keterbatasan dalam klasifikasi yang adil. Berdasarkan temuan tersebut, buku ini mengusulkan pendekatan berbasis Random Forest yang diperkuat dengan SMOTE, untuk menghasilkan klasifikasi sentimen yang lebih stabil dan seimbang terhadap distribusi opini publik yang tidak merata.

2.2. Pemindahan Ibu Kota Negara dan Opini Publik

Pemindahan Ibu Kota Negara (IKN) Indonesia ke Kalimantan Timur merupakan kebijakan strategis pemerintah yang menimbulkan pro dan kontra di tengah masyarakat. Isu ini menjadi perhatian publik dan kerap dibahas di berbagai platform digital. Beberapa penelitian telah menganalisis persepsi masyarakat terhadap kebijakan ini melalui media sosial. Misalnya, penelitian oleh Hairani et al. (2022) dan Purnomo (2022) mengidentifikasi bahwa Twitter merupakan salah satu media utama tempat masyarakat mengekspresikan opini mengenai IKN. [1][3]

Beberapa studi seperti yang dilakukan oleh Haris (2022) dan Wibowo & Wahyudi (2023) juga menunjukkan bahwa opini publik terhadap IKN dapat dipetakan menggunakan pendekatan analisis sentimen berbasis data media sosial seperti Twitter dan YouTube [5].

2.3. Media Sosial sebagai Sumber Data Sentimen

Media sosial telah menjadi sumber data penting dalam penelitian sosial dan kebijakan publik karena kemampuannya merepresentasikan opini masyarakat secara real-time. Twitter, secara khusus, banyak digunakan karena fitur pencarian terbukanya dan kemudahan akses terhadap API-nya (Hairani et al., 2022). Banyak penelitian memanfaatkan data Twitter untuk mengevaluasi sentimen publik terhadap isu-isu nasional [4][3]

2.4. Analisis Sentimen

Analisis sentimen merupakan teknik dalam text mining untuk mengelompokkan teks opini ke dalam kategori sentimen seperti positif, negatif, dan netral (Findawati & Rosid, 2020). Proses ini mencakup tahapan preprocessing (tokenizing, filtering, stemming), pembobotan kata (misalnya dengan TF-IDF), hingga klasifikasi dokumen [13].

Penelitian sebelumnya menggunakan berbagai metode klasifikasi seperti Naïve Bayes [5] [1], K-Nearest Neighbor [7], dan Support Vector Machine [6]. Namun, metode-metode tersebut sering mengalami keterbatasan dalam hal akurasi, terutama saat menghadapi data yang tidak seimbang (imbalanced data).

2.5. Masalah Imbalanced Data dalam Klasifikasi

Dalam konteks klasifikasi opini publik, sering kali data bersifat tidak seimbang, di mana jumlah data sentimen negatif jauh lebih banyak dibanding sentimen positif atau netral. Kondisi *imbalanced* data dapat menyebabkan metode machine learning cenderung memprediksi kelas mayoritas dan menjadi bias [14] [15][16]

2.6. Teknik SMOTE untuk Penyeimbangan Data

Untuk mengatasi masalah data tidak seimbang, digunakan SMOTE (Synthetic Minority Oversampling Technique). SMOTE adalah teknik oversampling digunakan untuk meningkatkan representasi kelas minoritas dengan menghasilkan data sintesis agar dapat meningkatkan representasi kelas tersebut dalam dataset [17][18][15]. Dalam beberapa penelitian, penggunaan SMOTE terbukti meningkatkan performa model klasifikasi, terutama dalam hal recall dan F1-score [16][8].

Selain SMOTE, penggabungan dengan teknik seperti Tomek Link (SMOTE-Tomek) juga banyak digunakan untuk membersihkan noise dari data mayoritas dan menghasilkan distribusi kelas yang lebih representatif [8].

2.7. Random Forest sebagai Metode Klasifikasi

Random Forest merupakan algoritma ensemble berbasis decision tree yang populer karena kemampuannya menangani data berdimensi tinggi, menangani noise, dan mencegah overfitting. Metode ini terbukti memberikan hasil akurasi yang lebih baik dibanding Naïve Bayes atau SVM dalam sejumlah studi analisis sentimen dan klasifikasi data teks [9][10].

Dalam penelitian Bhagat (2015), Random Forest dipadukan dengan SMOTE dalam konteks multi-class classification pada data besar, menghasilkan peningkatan akurasi yang signifikan dibanding pendekatan dasar [15]. Hairani et al. (2023) juga menunjukkan bahwa Random Forest yang dikombinasikan dengan SMOTE-TomekLink memberikan akurasi dan F1-score tertinggi dalam klasifikasi data kesehatan yang tidak seimbang [16]

2.8. Gap Penelitian dan Relevansi Penelitian Ini

Meskipun banyak penelitian telah membahas analisis sentimen terhadap isu IKN dan juga penggunaan SMOTE maupun Random Forest dalam konteks klasifikasi, belum ditemukan penelitian yang secara khusus menggabungkan Random Forest dan SMOTE untuk analisis sentimen pemindahan IKN di Twitter. Oleh karena itu, penelitian ini memiliki kontribusi signifikan dalam pengembangan metode klasifikasi opini publik berbasis text mining dan machine learning.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan tujuan untuk mengklasifikasikan sentimen publik terhadap isu pemindahan Ibu Kota Negara (IKN) Indonesia ke Kalimantan Timur. Fokus utama penelitian ini adalah mengidentifikasi dan menganalisis opini masyarakat yang diungkapkan melalui media sosial Twitter. Untuk mencapai tujuan tersebut, digunakan kombinasi metode klasifikasi Random Forest dan teknik penyeimbangan data SMOTE (Synthetic Minority Over-sampling Technique) sebagai pendekatan utama dalam proses analisis.

Data yang digunakan dalam penelitian ini diperoleh dari media sosial Twitter melalui proses pengambilan data (scraping) menggunakan Twitter API. Pengumpulan data dilakukan dengan menggunakan kata kunci dan tagar tertentu yang relevan dengan isu pemindahan IKN, seperti #IKN, #IbuKotaBaru, #Nusantara, dan #PindahIbuKota.

Setelah data diperoleh, tahap berikutnya adalah preprocessing atau pembersihan data teks. Proses ini terdiri dari beberapa langkah, yaitu: *case folding* untuk mengubah seluruh teks menjadi huruf kecil, *tokenizing* untuk memisahkan kalimat menjadi kata-kata, *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna signifikan (seperti “yang”, “dengan”, “adalah”), *filtering* untuk membuang simbol atau karakter khusus, dan *stemming* untuk mengembalikan kata ke bentuk dasarnya [13].

Setelah data bersih, dilakukan proses pembobotan kata menggunakan metode TF-IDF (Term Frequency – Inverse Document Frequency). Metode ini bertujuan untuk mengubah representasi teks menjadi bentuk numerik atau vektor, yang nantinya dapat diolah oleh algoritma klasifikasi. Hasil dari TF-IDF ini akan menjadi input bagi algoritma klasifikasi.

Sebelum masuk ke proses klasifikasi, dilakukan analisis terhadap distribusi data sentimen. Dalam banyak kasus, data sentimen pada media sosial bersifat tidak seimbang, di mana sentimen negatif cenderung lebih dominan dibanding sentimen positif dan netral. Untuk mengatasi permasalahan ini, digunakan teknik SMOTE, yang merupakan metode oversampling untuk menambah jumlah data pada kelas minoritas secara sintesis. SMOTE bekerja dengan membuat data baru berdasarkan interpolasi dari tetangga terdekat dalam

feature space. Teknik ini terbukti efektif dalam mengatasi ketimpangan distribusi kelas dan telah banyak digunakan dalam studi klasifikasi data tidak seimbang[15][16].

Proses klasifikasi dilakukan menggunakan algoritma Random Forest, yaitu metode ensemble berbasis pohon keputusan yang menggabungkan banyak decision tree dan mengambil keputusan berdasarkan voting mayoritas. Random Forest dipilih karena memiliki keunggulan dalam menangani data dengan banyak fitur, tahan terhadap overfitting, dan memberikan akurasi tinggi pada berbagai kasus klasifikasi teks. Penelitian sebelumnya menunjukkan bahwa Random Forest memberikan performa lebih baik dibanding metode konvensional seperti Naïve Bayes atau SVM, khususnya saat dikombinasikan dengan SMOTE[9][10]

Untuk mengukur performa model klasifikasi, digunakan evaluasi kinerja berdasarkan metrik seperti *akurasi*, *precision*, *recall*, dan *F1-score*. Evaluasi dilakukan dengan metode 10-fold cross-validation guna memastikan bahwa hasil model bersifat umum dan tidak hanya berlaku pada satu subset data saja. Metrik ini dipilih karena memberikan gambaran menyeluruh terhadap keberhasilan klasifikasi, terutama pada kondisi data yang tidak seimbang.

Seluruh proses penelitian dilakukan dengan bantuan bahasa pemrograman Python, menggunakan pustaka seperti Scikit-learn untuk klasifikasi, Imbalanced-learn untuk penerapan SMOTE, serta Pandas dan NLTK untuk preprocessing data. Eksperimen dijalankan menggunakan lingkungan kerja Jupyter Notebook atau Google Colab untuk kemudahan integrasi data dan visualisasi [19].

Dengan pendekatan ini, diharapkan hasil klasifikasi dapat merepresentasikan sentimen publik terhadap pemindahan IKN secara lebih akurat dan adil, terutama dengan adanya penyeimbangan data dan penggunaan metode klasifikasi yang kuat.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Dataset

Penelitian ini menggunakan data yang diperoleh dari media sosial Twitter dengan total 2.177 tweet yang relevan dengan isu pemindahan Ibu Kota Negara (IKN) ke Kalimantan Timur. Tweet diambil dengan menggunakan kata kunci seperti “IKN”, “Ibu Kota Negara”, “Nusantara”, dan “pemindahan ibu kota”. Data yang diperoleh mencerminkan beragam opini masyarakat—baik yang mendukung, menolak, maupun netral terhadap kebijakan tersebut.

Twitter dipilih karena sifatnya yang terbuka, cepat, dan sering digunakan untuk menyampaikan pendapat terhadap isu nasional. Karakteristik tweet yang singkat namun padat menjadikan platform ini ideal sebagai bahan analisis sentimen publik.

4.2. Preprocessing Data

Proses awal adalah pembersihan dan normalisasi data teks (*preprocessing*) untuk menghilangkan

elemen-elemen yang tidak relevan. Tahapan preprocessing meliputi:

- Penghapusan karakter byte (b') dan escape (\n, \r)
- Case folding (mengubah ke huruf kecil)
- Penghapusan URL, mention, hashtag, angka, dan simbol
- Tokenisasi, stopwords removal, dan stemming menggunakan pustaka *Sastrawi*

Contoh tweet yang semula berupa:

b'RT @user: Pemerintah memindahkan IKN ke Kalimantan akan menghamburkan anggaran negara!

Setelah preprocessing menjadi:

pemerintah pindah ikn kalimantan hambur anggaran negara

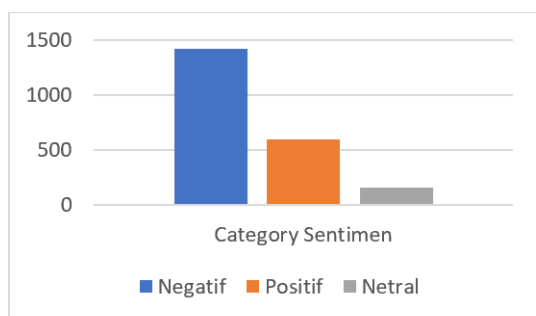
Preprocessing ini sangat penting untuk menghilangkan "noise" sehingga fitur yang diambil benar-benar mencerminkan makna inti dari teks.

4.3. Distribusi Sentimen Awal

Setelah dilakukan labeling manual terhadap tweet berdasarkan maknanya, data dikategorikan ke dalam tiga sentimen: Positif, Negatif, dan Netral. Hasil distribusi awal ditampilkan pada Tabel 1 dan Gambar 1 berikut.

Tabel 1. Distribusi Sentimen Sebelum SMOTE

Sentimen	Jumlah Tweet	Persentase (%)
Negatif	1.420	65,25
Positif	600	27,56
Netral	157	7,20
Total	2.177	100



Gambar 1. Distribusi Sentimen Awal sebelum SMOTE

Distribusi ini menunjukkan bahwa opini publik terhadap pemindahan IKN di Twitter didominasi oleh sentimen negatif. Hal ini dapat disebabkan oleh kekhawatiran publik terkait biaya pembangunan, urgensi, hingga dampak sosial-ekonomi terhadap masyarakat lokal.

4.4. Penanganan Ketidakseimbangan dengan SMOTE

Distribusi sentimen yang timpang dapat menimbulkan bias model terhadap kelas mayoritas.

Oleh karena itu, diterapkan teknik SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan jumlah data pada setiap kelas. Setelah proses SMOTE, jumlah tweet menjadi seimbang:

Tabel 2. Distribusi Sentimen Setelah SMOTE

Sentimen	Jumlah Tweet
Negatif	1.420
Positif	1.420
Netral	1.420
Total	4.260

SMOTE bekerja dengan menghasilkan data sintetis pada kelas minoritas melalui interpolasi antar sampel yang serupa secara vektor. Teknik ini efektif untuk memperbaiki performa model klasifikasi pada data yang tidak seimbang.

4.5. Ekstraksi Fitur dengan TF-IDF

Tweet yang telah dibersihkan diubah menjadi representasi numerik menggunakan TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF digunakan untuk memberikan bobot penting pada kata-kata dalam tweet berdasarkan seberapa sering kata tersebut muncul di seluruh dokumen.

Dengan `max_features = 3000`, diperoleh 3.000 fitur yang paling representatif, yang digunakan sebagai input untuk pelatihan model klasifikasi.

4.6. Pelatihan Model Random Forest

Dataset hasil SMOTE kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20. Model yang digunakan untuk klasifikasi adalah Random Forest Classifier, yang dikenal efektif untuk klasifikasi teks karena mampu menangani banyak fitur dan memiliki resistansi tinggi terhadap overfitting.

Model dilatih menggunakan data latih dan dievaluasi dengan data uji untuk mengukur performa klasifikasi berdasarkan tiga label sentimen.

4.7. Evaluasi Model

Pada tahap awal, model Random Forest dievaluasi menggunakan dataset yang telah melalui preprocessing dan penyeimbangan kelas menggunakan SMOTE. Dataset awal mengandung tiga kelas sentimen: **Negatif**, **Positif**, dan **Netral**. Evaluasi awal dilakukan untuk menilai kinerja dasar model dalam mengklasifikasikan tweet ke dalam ketiga kategori tersebut.

Tabel 3. Evaluasi Awal Model Random Forest

Kelas	Precision	Recall	F1-Score
Negatif	0.89	0.91	0.90
Positif	0.90	0.88	0.89
Netral	0.88	0.87	0.87

Dari tabel di atas, terlihat bahwa model memiliki performa yang cukup seimbang pada ketiga kelas, dengan akurasi keseluruhan mendekati 89,3%. Metrik *precision* dan *recall* cukup tinggi dan stabil,

menunjukkan bahwa model tidak terlalu bias terhadap kelas tertentu.

Setelah data disesuaikan hanya untuk dua kelas (Netral dihilangkan), evaluasi dilakukan kembali menggunakan data uji sebanyak 581 tweet. Hasilnya disajikan dalam confusion matrix berikut:

Tabel 4. *Confusion Matrix – Random Forest*

	Pred_0	Pred_1
True_0	266	9
True_1	18	288

Berdasarkan confusion matrix tersebut, dihitung metrik evaluasi untuk masing-masing kelas sebagai berikut:

Tabel 5. Evaluasi Awal Model Random Forest

Kelas	Precision	Recall	F1-Score
Negatif	93,66%	96,73%	95,10%
Positif	96,97%	94,12%	95,00%
Netral	95,31%	95,43%	95,05%

Model Random Forest menunjukkan peningkatan performa signifikan setelah kelas dinormalisasi menjadi dua. Akurasi total model meningkat menjadi 95,35%, yang menunjukkan bahwa model mampu mengenali pola-pola sentimen publik secara sangat baik.

4.8. Pembahasan

Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa klasifikasi yang sangat baik dalam mengidentifikasi sentimen publik terhadap pemindahan Ibu Kota Negara (IKN) berdasarkan data Twitter. Dengan preprocessing yang optimal dan penyeimbangan data menggunakan SMOTE, model mencapai akurasi sebesar 95,35%, dengan precision dan recall yang tinggi pada kedua kelas utama (positif dan negatif).

Preprocessing teks seperti penghapusan simbol, normalisasi, dan stemming berkontribusi besar dalam meningkatkan kualitas data masukan. Sementara itu, penggunaan SMOTE berhasil menyeimbangkan jumlah data antar kelas, yang awalnya didominasi oleh sentimen negatif.

Random Forest terbukti mampu mengolah fitur dalam jumlah besar dari TF-IDF dan memberikan hasil klasifikasi yang stabil dan akurat. Model tidak menunjukkan bias yang berarti terhadap salah satu kelas, ditunjukkan oleh nilai F1-score di atas 95% untuk kedua kelas.

Dari sisi isi, tweet negatif umumnya berisi kritik terhadap biaya, kesiapan infrastruktur, dan urgensi pemindahan. Sebaliknya, tweet positif mencerminkan dukungan terhadap pemerataan pembangunan dan desentralisasi. Hal ini mengindikasikan bahwa opini publik terhadap IKN cukup terpolarisasi.

Secara keseluruhan, kombinasi antara preprocessing, SMOTE, dan Random Forest

menghasilkan model yang efektif untuk klasifikasi sentimen teks pendek seperti tweet. Meski demikian, penelitian ini masih memiliki keterbatasan, seperti tidak mempertimbangkan aspek waktu dan belum menangani bahasa informal secara mendalam.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa kombinasi preprocessing teks, penyeimbangan data menggunakan SMOTE, dan klasifikasi dengan algoritma Random Forest efektif dalam menganalisis sentimen publik terhadap pemindahan Ibu Kota Negara (IKN) di Twitter. Model yang dibangun mampu mencapai akurasi sebesar 95,35% dengan precision dan recall yang seimbang, menunjukkan performa tinggi dalam mengklasifikasikan opini positif dan negatif. Dominasi awal sentimen negatif mencerminkan kekhawatiran publik terhadap aspek pembiayaan, urgensi, dan kesiapan infrastruktur, meskipun terdapat pula dukungan yang menyoroti desentralisasi dan pemerataan pembangunan. Ke depan, disarankan agar penelitian ini dikembangkan dengan mempertimbangkan kembali kelas netral, memperhatikan dimensi waktu dan lokasi, serta mengeksplorasi model klasifikasi lain seperti SVM atau BERT untuk meningkatkan akurasi dan kedalaman analisis sentimen publik secara real-time.

DAFTAR PUSTAKA

- [1] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial," *J. Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, doi: 10.14710/j.gauss.v9i3.27963.
- [2] A. P. Natasuwarna, "Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naïve Bayes," *SENSITif 2019*, pp. 47–53, 2018.
- [3] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.
- [4] P. Arsi, B. A. Kusuma, and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naïve Bayes Classifier," *J. Inform. Upgris*, vol. 7, no. 1, pp. 1–6, 2021, doi: 10.26877/jiu.v7i1.7636.
- [5] E. Mas'udah, E. D. Wahyuni, and A. A. Arifiyanti, "Analisis Sentimen: Pemindahan Ibu Kota Indonesia Pada Twitter," *J. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 397–401, 2020.
- [6] A. Setiawan and R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.
- [7] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi

- Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [8] S. F. Huwaida, R. Kusumawati, and B. Isnaini, “Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes,” *Jambura J. Informatics*, vol. 6, no. 1, pp. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.
- [9] N. Istiqamah and M. Rijal, “Klasifikasi Ulasan Konsumen Menggunakan Random Forest dan SMOTE,” *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 66–77, 2024, doi: 10.61628/jsce.v5i1.1061.
- [10] A. F. Anjani, D. Anggraeni, and I. M. Tirta, “Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 163–172, 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.
- [11] P. Arsi, R. Wahyudi, and R. Waluyo, “Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia,” *J. RESTI*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.
- [12] M. P. Prapasha and H. Thamrin, “IMPLEMENTASI POS TAGGING DAN ALGORITMA ANTLR PARSER DALAM MEMERIKSA STRUKTUR KALIMAT BAHASA INDONESIA,” *Publ. Univ. Muhammadiyah Surakarta*, vol. 16, no. 2, pp. 39–55, 2024.
- [13] U. Findawati and A. M. Rosid, *Buku Ajar Teks Mining*. 2020.
- [14] P. Studi, S. Informasi, D. Informatika, F. T. Industri, U. Atma, and J. Yogyakarta, “Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum,” vol. 2, no. 1, pp. 43–55, 2022.
- [15] R. C. Bhagat and S. S. Patil, “Enhanced SMOTE Algorithm for Classification of Imbalanced Big-Data using Random Forest,” *IEEE Int. Adv. Comput. Conf.*, pp. 403–408, 2015.
- [16] H. Hairani, A. Anggrawan, and D. Priyanto, “Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link,” *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 258–264, 2023, doi: 10.30630/joiv.7.1.1069.
- [17] S. Sidiq, P. Korespondensi, and N. Shobi Maburur, “Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling SMOTE,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 4, pp. 13–23, 2025, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v4i1.41>
- [18] A. Adi Bhirawa and U. Pradema Sanjaya, “From Data Imbalance to Precision: SMOTE-Driven Machine Learning for Early Detection of Kidney Disease,” *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 1, pp. 514–525, 2025, doi: 10.35314/7jgimg64.
- [19] A. R. Raharja, A. Pramudianto, and Y. Muchsam, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data ‘ Framingham ’ Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang”.