

Penerapan Data Mining Untuk Analisis Sentimen Masyarakat Terhadap Ibu Kota Nusantara Pada Media Sosial X

Rival Valentino Rayean, M Afdal*, Inggih Permana, Nesdi Evrilyan Rozanda

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia

E-Mail: ¹12050312583@students.uin-suska.ac.id, ^{2,*}m.afdal@uin-suska.ac.id, ³inggihermana@uin-suska.ac.id,

⁴nesdi.er@uin-suska.ac.id

Email Penulis Korespondensi: m.afdal@uin-suska.ac.id

Submitted: 09/05/2025; Accepted: 13/06/2025; Published: 13/06/2025

Abstrak– Kebijakan pemindahan Ibu Kota Negara ke Nusantara (IKN) telah menjadi isu viral yang hangat diperbincangkan di Indonesia, memicu beragam reaksi publik mulai dari dukungan hingga penolakan. Penelitian ini bertujuan untuk menganalisis perubahan sentimen publik di media sosial X sebelum dan sesudah upacara kemerdekaan 2024, sekaligus membandingkan efektivitas algoritma Naive Bayes dan Random Forest dalam mengklasifikasikan sentimen tersebut. Untuk mencapai tujuan ini, sebanyak 1000 data tweet (500 sebelum dan 500 sesudah upacara) dikumpulkan, dilabeli secara manual, dan dianalisis menggunakan kedua algoritma dengan teknik balancing data SMOTE untuk mengatasi ketidakseimbangan kelas. Hasil analisis menunjukkan adanya pergeseran sentimen yang signifikan pasca-upacara kemerdekaan. Sentimen negatif tercatat meningkat dari 44% menjadi 50%, sementara sentimen positif mengalami kenaikan tipis dari 30% menjadi 33%, dan sentimen netral menurun dari 26% menjadi 17%. Dalam evaluasi performa model, algoritma Random Forest menunjukkan akurasi klasifikasi yang sedikit lebih tinggi dibandingkan Naive Bayes. Meskipun demikian, perbedaan akurasi yang tergolong kecil mengindikasikan bahwa keduanya sama-sama cukup efektif untuk melakukan analisis sentimen pada dataset penelitian ini. Studi ini berhasil menyajikan gambaran komprehensif mengenai dinamika dan polaritas opini publik terhadap kebijakan Ibu Kota Nusantara yang tengah berlangsung.

Kata Kunci: Ibu Kota Nusantara; Naive Bayes; Random Forest; Text Classification; X

Abstract–The policy of relocating the National Capital City to Nusantara (IKN) has become a viral and hotly debated issue in Indonesia, sparking diverse public reactions ranging from support to opposition. This research aims to analyze the shift in public sentiment on social media X before and after the 2024 Independence Day ceremony, while also comparing the effectiveness of the Naive Bayes and Random Forest algorithms in classifying these sentiments. To achieve these objectives, a total of 1000 tweet data (500 before and 500 after the ceremony) were collected, manually labeled, and analyzed using both algorithms with the SMOTE data balancing technique to address class imbalance. The analysis reveals a significant shift in sentiment post-ceremony. Negative sentiment increased from 44% to 50%, while positive sentiment saw a slight rise from 30% to 33%, and neutral sentiment decreased from 26% to 17%. In the model performance evaluation, the Random Forest algorithm showed slightly higher classification accuracy than Naive Bayes. Nevertheless, the small difference in accuracy indicates that both algorithms were quite effective for the sentiment analysis task on this dataset. This study successfully presents a comprehensive overview of the dynamics and polarity of public opinion regarding the ongoing the National Capital City to Nusantara policy.

Keywords: Capital City Of Nusantara; Naive Bayes; Random Forest; Text Classification; X

1. PENDAHULUAN

Indonesia, dengan luas wilayah mencapai 1.905.000 kilometer persegi, adalah salah satu negara terluas di Asia Tenggara dan menempati peringkat ke-14 sebagai negara terbesar di dunia. Pada tahun 2022, populasi Indonesia tercatat sebanyak 275,77 juta jiwa, menjadikannya negara dengan jumlah penduduk terbanyak keempat di dunia [1]. Ibu kotanya, Jakarta, ditetapkan secara resmi sebagai ibu kota Indonesia melalui Undang-Undang Nomor 10 Tahun 1964, yang disahkan oleh Presiden Republik Indonesia saat itu, Ir. Soekarno, pada tanggal 31 Agustus 1964 [2].

Namun, seiring berjalannya waktu, isu konsentrasi penduduk menjadi pertimbangan penting. Berdasarkan data dari Bappenas, mayoritas besar penduduk Indonesia (56,56%) menetap di Pulau Jawa, tepatnya di Kota Jakarta. Sebaliknya, Pulau Kalimantan hanya menjadi tempat tinggal bagi 7,33% penduduk. Tingginya perbedaan persentase ini menunjukkan bahwa Pulau Jawa sangat padat penduduk, yang kemudian mendorong pemerintah untuk mempertimbangkan pemindahan ibu kota negara ke Pulau Kalimantan guna pemerataan. Setelah melalui serangkaian konsultasi publik dan berbagai kajian mendalam, akhirnya dalam pertemuan yang dilaksanakan pada tanggal 29 April 2019, Presiden Joko Widodo secara resmi memutuskan dan mengesahkan rencana pemindahan ibu kota negara ke Pulau Kalimantan, dengan lokasi yang ditetapkan berada di Kalimantan Timur [3].

Keputusan pemindahan ini kemudian dikukuhkan dan diatur lebih lanjut melalui landasan hukum. Pada tanggal 15 Februari 2022, disahkan Undang-Undang Republik Indonesia Nomor 3 Tahun 2022 tentang Ibu Kota Negara (IKN), yang menjadi dasar hukum untuk mempermudah keseluruhan proses pemindahan. Meskipun langkah pemindahan ke Ibu Kota Nusantara (IKN) ini dirancang dengan berbagai potensi keuntungan dan manfaat strategis, prosesnya juga tidak lepas dari sejumlah kontroversi. Beberapa isu yang mengemuka meliputi aspek pemilihan lokasi pembangunan, persepsi publik mengenai pengesahan Undang-Undang yang dianggap sebagian pihak terlalu cepat, hingga wacana pelibatan masyarakat dalam skema pendanaan melalui galang dana (*crowdfunding*) [4]. Dengan alokasi anggaran yang direncanakan sekitar Rp.466 triliun, tahap awal pembangunan Ibu Kota Nusantara (IKN) telah dimulai dan dijadwalkan berlangsung antara tahun 2022 hingga 2024. Pemerintah menekankan bahwa dalam keseluruhan prosesnya, pemindahan ibu kota negara ini akan sangat memperhatikan dampak terhadap lingkungan sekitar. Meskipun demikian, pembangunan infrastruktur skala besar dan proses penyesuaian lahan yang ekstensif tak

bisa dipungkiri berpotensi menimbulkan berbagai tantangan lingkungan, seperti risiko kerugian ekologis, kehilangan habitat bagi keanekaragaman hayati, serta potensi kontribusi terhadap perubahan iklim. Lebih lanjut, dampak sosial seperti timbulnya ketidaknyamanan bagi komunitas lokal, perubahan lanskap sosial-ekonomi, dan kemungkinan terjadinya konflik selama proses relokasi juga menjadi aspek krusial yang memerlukan perhatian serius. Oleh karena itu, pemerintah memikul tanggung jawab besar untuk memastikan bahwa setiap individu dan komunitas yang terdampak oleh proyek ini mendapatkan perlindungan hukum, kompensasi yang adil dan layak, serta dukungan komprehensif selama menjalani periode perubahan dan adaptasi [5].

Menurut Survei APJII 2024, Pengguna media sosial X di Indonesia adalah 1,06% dari pengguna media sosial yang ada di Indonesia. Ibukota baru Republik Indonesia, Ibu kota Nusantara (IKN), adalah topik diskusi yang hangat dari media sosial dan debat yang menarik. Dengan melakukan kebijakan pemindahan Ibu Kota akan menimbulkan pro dan kontra dan membuat sebagian masyarakat Indonesia setuju dan sebagian lain tidak setuju. Selain itu, nama "Nusantara" yang digunakan sebagai nama baru Ibu Kota menuai pro dan kontra, karena nama "Nusantara" merujuk pada wilayah yang luas yang sebanding dengan Indonesia[6] Untuk mengetahui apakah penduduk Indonesia merespons secara positif, negatif, dan netral terhadap masalah ini, analisis pada sentimen harus dilakukan berdasarkan jumlah tanggapan dari pengguna media sosial ke salah satu topik hangat saat ini. Maka dari itu, penelitian ini memberikan visualisasi mengenai opini publik tentang kebijakan pemerintah yang mengatur pemindahan ibukota negara Indonesia [7].

Beberapa *tweet* yang berisi pendapat tentang berbagai topik seperti ekonomi, politik, sosial, budaya, dan hukum bisa dianalisis melalui analisis sentimen untuk mengetahui isi atau perasaan yang terkandung didalamnya. Dengan demikian, para peneliti dapat memanfaatkan media sosial X sebagai alat untuk mengamati suasana hati dan emosi seseorang dengan menganalisis sentimen yang terkandung dalam *tweet* yang dibuatnya [8].

Terdapat dua pendekatan yang berguna dalam melakukan evaluasi sentiment dalam dokumen: pendekatan *lexicon* atau *machine learning*. Pendekatan *lexicon* mengklasifikasikan data berdasarkan kata positif atau negatif yang ada di dalamnya, dan metode *machine learning* menggunakan pembelajaran dari data latih [9]. Dalam teknik pengolahan dan klasifikasi *data mining* yang menggunakan bayes teorama, metode klasifikasi NBC adalah algoritma yang paling efektif dalam klasifikasi teks dan analisis data berdimensi tinggi jika dibandingkan dengan metode lainnya, yang menunjukkan bahwa NBC lebih unggul dalam akurasi, *presisi*, *recall*, dan *f-measure* [10]. *Random Forest* adalah metode algoritma pembelajaran kelompok dimana banyak menggunakan pohon keputusan (*decision tree*) algoritma ini unggul dalam menganalisis data yang berukuran besar dan mengurangi resiko *overfitting*, Semua pohon Keputusan (*decision tree*) belajar memprediksi emosi berdasarkan karakteristik data. Setelah melatih model algoritma *Random Forest*, *testing* set akan digunakan untuk memprediksi sentimen populasi yang dipilih secara acak [11]

Berbagai publikasi jurnal ilmiah menunjukkan kemajuan dalam penelitian *text mining* yang menggunakan teknik analisis sentimen. Berbagai metode atau algoritma yang telah digunakan oleh peneliti sebelumnya digunakan untuk melakukan penelitian ini [12]. Penelitian terdahulu yang dilakukan oleh Dedi Pramana pada tahun 2023 bertujuan untuk mengetahui analisis sentimen dari opini publik terkait pemindahan ibu kota negara. Hasil penelitian ini menunjukkan bahwa hasil dari eksperimen perbandingan algoritma NBC sebesar 63,09% dan KNN sebesar 69,25%. Sementara itu, pandangan masyarakat terhadap sentiment terkait pemindahan ibu kota Negara memiliki polaritas, yaitu persentase sentimen positif 28% dengan jumlah 649 komentar, persentase sentimen negatif 30% dengan jumlah 730 komentar, dan persentase sentimen netral 42% dengan jumlah 1025 komentar[1]. Lalu penelitian terkait yang memiliki topik yang sama yang dilakukan oleh Syahril Dwi pada tahun 2023 mengenai relokasi ibu kota negara Dimana penelitian ini menggunakan algoritma *Naive Bayes* dan KNN penelitian ini menghasilkan tingkat akurasi analisis sentimen sebesar 82,27%, tingkat *precision* sebesar 86,36%, dan nilai *recall* sebesar 76,93% diperoleh dari data *tweet* yang dikumpulkan menggunakan metode NB. Metode KNN mencapai tingkat akurasi sebesar 88,12%, dengan *precision* sebesar 93,98% dan *recall* sebesar 81,53%. Berdasarkan angka-angka tersebut, dapat disimpulkan bahwa metode KNN memiliki akurasi yang lebih baik dibandingkan metode NB[13].

Didasarkan pada penjelasan sebelumnya dan didukung oleh penelitian sebelumnya, penelitian ini akan melakukan evaluasi dan klasifikasi serta membandingkan sentimen masyarakat terhadap ibu kota nusantara dimedia sosial X pada data pra upacara kemerdekaan 2024 dan pasca upacara kemerdekaan 2024.

2. METODOLOGI PENELITIAN

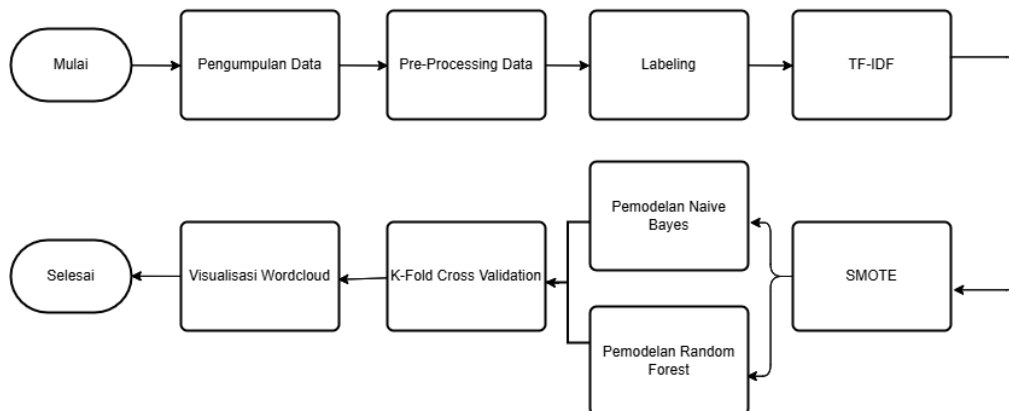
2.1 Tahapan Penelitian

Didalam penelitian ini beberapa tahapan yang dilakukan dimulai dari tahapan perencanaan penelitian lalu, tahap pengumpulan data, setelah itu data yang dikumpulkan dilakukan proses *pre-processing* untuk membersihkan data, data yang sudah dibersihkan dan diberikan label lalu dimodelkan menggunakan algoritma klasifikasi *Naive Bayes* dan algoritma klasifikasi *Random Forest*, lalu kemudian model algoritma dilakukan proses validasi menggunakan *K-Fold Cross Validation*.

Penelitian ini menggunakan data yang diambil melalui proses *crawling*[14]. Metode ini menggunakan *Tweet harvest*, sebuah *tool* yang menggunakan *library* dari bahasa pemrograman *Python*, dan *auth* token pengguna akun

sosial media X. Data yang diambil merupakan data tahun 2024 pada pra upacara kemerdekaan 2024 dan pasca upacara kemerdekaan 2024 dengan kata kunci “IKN”.

Setelah itu, proses preprocessing data, pembersihan data dan seleksi data dilakukan. Selanjutnya, data diberi tanda sentimen positif, negatif, atau netral secara manual oleh ahli bahasa Indonesia. Setelah proses pelabelan selesai, kata-kata dalam data tersebut diberi bobot menggunakan metode *TF-IDF*, *TF-IDF* digunakan untuk pembobotan kata. Setelah tahap persiapan data selesai, data dapat diklasifikasikan dan dimodelkan menggunakan dua metode algoritma *Naive Bayes* dan metode algoritma *Random Forest*. Klasifikasi dilakukan menggunakan bahasa pemrograman *Python*, dan kemudian validasi dilakukan menggunakan *K-Fold Cross Validation*.



Gambar 1. Metodologi Penelitian

2.2 Text Classification

Text Classification adalah bidang dari *data mining* yaitu data dari teks yang terstruktur atau tidak terstruktur, dilakukan proses analisis untuk menghasilkan informasi yang berguna, informasi yang telah didapat kemudian diproses [15]. *Text Classification* adalah proses mengambil dan mengelola informasi dari sejumlah besar data berbentuk teks. Proses ini biasanya digunakan untuk mengolah dan mengorganisir data dengan cara menganalisis informasi yang penting dan relevan. Dengan demikian, Proses ini bertujuan untuk mengolah data berbentuk teks yang tersedia dan mengambil informasi yang dibutuhkan [8].

2.3 Analisis Sentimen

Analisis sentimen adalah proses pemahaman secara otomatis, mengekstraksi dan memproses teks, dan mendapatkan informasi emosional implisit tentang pendapat, berbasis pada pendapat, emosi, dan perasaan. Untuk pengumpulan data dalam penelitian ini, media sosial X digunakan untuk menyimpan pendapat publik [16].

2.4 Text PreProcessing

Teknik yang digunakan untuk membersihkan, mengubah, dan menstandarisasi data teks mentah menjadi format yang lebih terstruktur dan mudah dipahami oleh algoritma komputer, khususnya dalam tugas-tugas *Natural Language Processing (NLP)* [17].

2.5 Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF)

Pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* merupakan salah satu proses metode ekstraksi fitur dengan memberikan nilai pada masing-masing kata yang terdapat pada *tweet* yang dilatih (*training*), sehingga pembobotan atau perhitungan dilakukan untuk menentukan seberapa berharga sebuah kata dalam mewakili sebuah kalimat. Hasil dari skor *TF-IDF* didapatkan itu dari berapa kali kata muncul dalam dokumen [18].

$$W_{t,d} = W_{tf,t,d} \times idf_t \tag{1}$$

2.6 Naïve Bayes

Untuk mendapatkan probabilitas tertinggi dalam kategori yang paling tepat, algoritma tunggal yang paling umum digunakan untuk mengklasifikasikan data tidak terstruktur. Ini adalah metode yang tepat dalam melakukan klasifikasi teks untuk digunakan dalam sentimen analisis dan memiliki kemampuan untuk mengklasifikasikan data [1].

$$P(C|X) = \frac{P(X|C)P(C)}{P(x)} \tag{2}$$

2.7 Random Forest

Model algoritma klasifikasi *Random Forest* adalah evolusi dari model Algoritma pohon keputusan (*Decision Tree*), yang menggunakan sampel individu untuk melatih setiap pohon untuk berpikir. Model ini menghasilkan banyak pohon dengan cara yang sama. Pohon berkembang seiring banyaknya data. Algoritma yang dikenal sebagai *Random Forest*



Classifier bekerja untuk membuat pohon regresi dan klasifikasi, dengan *node* dipisahkan oleh algoritma yang disempurnakan untuk meminimalkan *squared-error loss*[19].

2.8 K-Fold Cross Validation

K-Fold Cross Validation mengurutkan set data yaitu terdiri dari *data training* dan *data testing*. Metode ini menggunakan partisi data untuk menjadikannya salah satu metode pemindaian data yang paling umum dengan melakukan partisi data dengan jumlah yang sama menjadi k-fold atau bagian data. Selama pelatihan dan pengujian berulang, proses ini dilakukan sebanyak k-n [18].

2.9 SMOTE

Untuk mengatasi kelas dalam data yang tidak seimbang, *SMOTE (Synthetic Minority Over-sampling Technique)* digunakan. Kelas yang tidak seimbang terjadi ketika data memiliki beberapa kelas. Sehingga kelas yang tidak seimbang ini dapat menimbulkan model memiliki gap yang signifikan terhadap kelas mayoritas, yang dapat membuat model kurang sensitif terhadap hal-hal yang berkaitan dengan kelas minoritas yang mungkin memiliki nilai prediktif yang cukup besar. Karena kelas mayoritas berkuasa, akurasi model dapat sangat tinggi, tetapi hasil prediksi untuk kelas minoritas akan sangat kecil[20].

3. HASIL DAN PEMBAHASAN

Penelitian ini mengambil data yang berasal dari *platform X* terkait berbagai opini masyarakat mengenai fenomena Ibu Kota Nusantara (IKN), data diambil dengan proses *crawling* menggunakan *library tweet harvest* dari pemograman *python*. Data terdiri dari dua variabel yaitu data Pra upacara kemerdekaan tahun 2024 dan Pasca upacara kemerdekaan tahun 2024.

3.1 Data Awal

Penelitian ini menggunakan data sebanyak 500 data diambil pada saat Pra upacara kemerdekaan Indonesia tahun 2024 dan 500 data diambil pada saat Pasca upacara kemerdekaan Indonesia tahun 2024.

Tabel 1. Data Awal

Data Set	Data
Data Awal Pra Upacara Kemerdekaan 2024	<ol style="list-style-type: none"> 1. Aneh banget ada orang mau milih 02 supaya pendukung 01 nangis-nangis dan menderita. Wkwkwk dude Gue pendukung 01 tapi sumber duit gua ada di IKN 02 menang gua pribadi malah cuan secara materi. Tapi bagaimana dengan negara gua? 2. Joe Biden Menyoroti Kepindahan IKN: Terdapat Potensi Banjir dan Korupsi Miliaran Dolar! 3. IKN 2 bulan lagi
Data Awal Pasca Upacara Kemerdekaan 2024	<ol style="list-style-type: none"> 1. AHY Resmi Serahkan Sertifikat Istana IKN kepada Pemerintah 2. Pendapat TPUA adalah bahwa IKN wajib dibatalkan karena UU IKN melanggar konstitusi. IKN itu pemerintah daerah tapi kok disebut badan Otorita. Menurut UU pemerintah daerah itu ada kabupaten provinsi dan kota. Tidak bisa berbentuk Otoritas. 3. Alhamdulillah hari ini Istana Negara sudah di resmikan dan training centee buat Timnas di resmikan oleh Presiden @jokowi di masa akhir jabatannya lanjut Istana Garuda kelak akan di resmikan oleh Presiden Terpilih 2024-2029 @prabowo#IKN#NusantaraBaru #TrimakasihPakJokowi

Dilihat Pada Tabel 1 bahwa data yang diambil merupakan data mentah yang belum diolah, masih menampilkan noise dan huruf yang beragam seperti simbol serta huruf besar kecil.

3.2 PreProcessing

Data yang sudah didapatkan sebelumnya kemudian diolah dalam tahap *preprocessing*, tahapan ini meliputi proses pembersihan huruf besar serta karakter, simbol, kalimat dan kata yang tidak formal/baku, kata yang tidak bermakna, angka, *hashtag* (#), *link url*, *username*, dan data kalimat yang duplikat, data dari data sebelumnya. Berikut adalah hasil dari tahapan *preprocessing* yang dapat dilihat pada Tabel 2.

Tabel 2. Pembersihan Data

Dataset	Data
Data Pra Upacara Kemerdekaan 2024	1. aneh banget ada orang mau milih supaya pendukung menangis menangis dan menderita wkwkwk dude saya pendukung tapi sumber uang saya ada di ikn menang saya pribadi bahkan cuan secara materi tapi bagaimana dengan negara saya



Dataset	Data
Data Pasca Upacara Kemerdekaan 2024	2. joe Biden menyoroiti kepindahan ikn terdapat potensi banjir dan korupsi miliaran dolar 3. ikn bulan lagi
	1. ahy resmi serahkan sertifikat istana ikn kepada pemerintah 2. pendapat tpua adalah bahwa ikn wajib dibatalkan karena uu ikn melanggar konstitusi ikn itu pemerintah daerah tapi kok disebut badan otorita menurut uu pemerintah daerah itu ada kabupaten provinsi dan kota tidak bisa berbentuk otoritas 3. alhamdulillah hari ini istana negara sudah di resmikan dan training centee buat timnas di resmikan oleh presiden di masa akhir jabatan nya lanjut istana garuda kelak akan di resmikan oleh presiden terpilih

Pada Tabel 3 dilakukan proses tokenisasi. Tokenisasi adalah proses memecah teks menjadi bagian-bagian kecil yang disebut token, biasanya berupa kata-kata atau tanda baca. Saat bekerja dengan data teks, tahap *preprocessing* umum yang disebut tokenisasi membuat pemrosesan dan analisis teks lebih mudah. Di sini, *library word_tokenize python* digunakan untuk memotong proses data yang terdiri dari kalimat dan kemudian dibagi menjadi kata-kata [13].

Tabel 3. Tokenize

Dataset	Data
Data Pra Upacara Kemerdekaan 2024	1. ['aneh', 'banget', 'ada', 'orang', 'mau', 'milih', 'supaya', 'pendukung', 'menangis', 'menangis', 'dan', 'menderita', 'wkwwkw', 'dude', 'saya', 'pendukung', 'tapi', 'sumber', 'uang', 'saya', 'ada', 'di', 'ikn', 'menang', 'saya', 'pribadi', 'bahkan', 'cuan', 'secara', 'materi', 'tapi', 'bagaimana', 'dengan', 'negara', 'saya'] 2. ['joe', 'Biden', 'menyoroti', 'kepindahan', 'ikn', 'terdapat', 'potensi', 'banjir', 'dan', 'korupsi', 'miliaran', 'dolar'] 3. ['ikn', 'bulan', 'lagi']
Data Pasca Upacara Kemerdekaan 2024	2. ['pendapat', 'tpua', 'adalah', 'bahwa', 'ikn', 'wajib', 'dibatalkan', 'karena', 'uu', 'ikn', 'melanggar', 'konstitusi', 'ikn', 'itu', 'pemerintah', 'daerah', 'tapi', 'kok', 'disebut', 'badan', 'otorita', 'menurut', 'uu', 'pemerintah', 'daerah', 'itu', 'ada', 'kabupaten', 'provinsi', 'dan', 'kota', 'tidak', 'bisa', 'berbentuk', 'otoritas'] 3. ['alhamdulillah', 'hari', 'ini', 'istana', 'negara', 'sudah', 'di', 'resmikan', 'dan', 'training', 'centee', 'buat', 'timnas', 'di', 'resmikan', 'oleh', 'presiden', 'di', 'masa', 'akhir', 'jabatan', 'nya', 'lanjut', 'istana', 'garuda', 'kelak', 'akan', 'di', 'resmikan', 'oleh', 'presiden', 'terpilih']

Data yang tersedia di X yaitu 1 tentang pemindahan Ibu Kota Negara adalah data tanpa label. Oleh karena itu, data harus dilabelkan. Pelabelan dilakukan secara manual dengan bantuan pakar. Selama proses pelabelan data *tweet*, *tweet* akan diberi nilai. Nilai positif menunjukkan bahwa *tweet* tersebut setuju dengan Ibu Kota Nusantara, nilai negatif menunjukkan bahwa *tweet* tersebut tidak setuju Ibu Kota Nusantara, dan nilai netral menunjukkan bahwa *tweet* tersebut tidak memiliki tujuan mendukung atau menolak Ibu Kota Nusantara proses pelabelan ditunjukkan dalam Tabel 4.

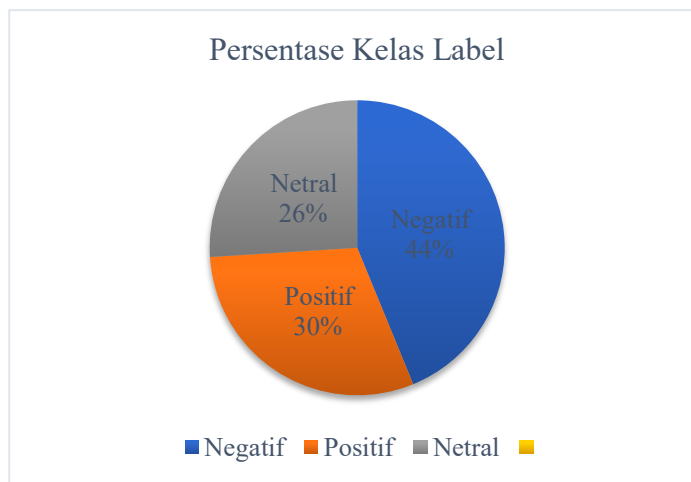
Tabel 4. Hasil Labelling

Dataset	Data	Label
Data Pra Upacara Kemerdekaan 2024	1. aneh banget ada orang mau milih supaya pendukung menangis menangis dan menderita wkwwkw dude saya pendukung tapi sumber uang saya ada di ikn menang saya pribadi bahkan cuan secara materi tapi bagaimana dengan negara saya	Negatif
	2. joe Biden menyoroiti kepindahan ikn terdapat potensi banjir dan korupsi miliaran dolar	Negatif
	3. ikn bulan lagi	Netral

	500. komnas ham terima adu duga langgar ham proyek ikn	Negatif
Data Pasca Upacara Kemerdekaan 2024	1. ahy resmi serahkan sertifikat istana ikn kepada pemerintah	Netral
	2. pendapat tpua adalah bahwa ikn wajib dibatalkan karena uu ikn melanggar konstitusi ikn itu pemerintah daerah tapi kok disebut badan otorita menurut uu pemerintah daerah itu ada kabupaten provinsi dan kota tidak bisa berbentuk otoritas	Negatif

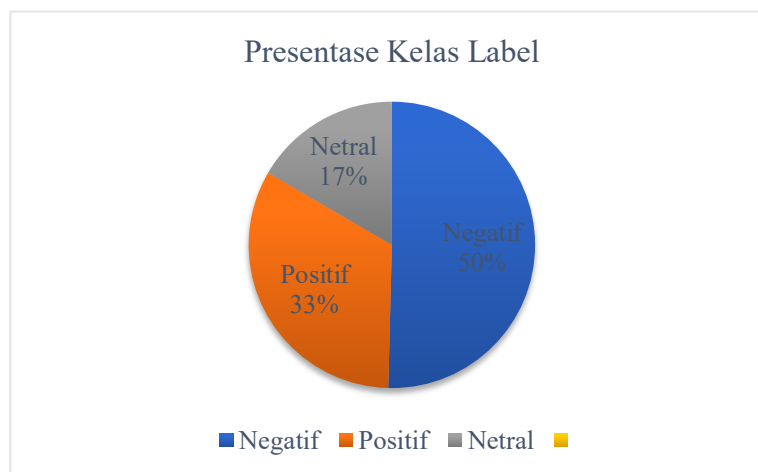
Dataset	Data	Label
	3. alhamdulillah hari ini istana negara sudah di resmikan dan training center buat timnas di resmikan oleh presiden di masa akhir jabatannya lanjut istana garuda kelak akan di resmikan oleh presiden terpilih	Positif
	...	
	500. lupa mayoritas rakyat kelas tengah jatuh miskin perintah foya foya gelar upacara ikn	Negatif

Dari 500 data Pra Upacara kemerdekaan 2024 yang diberikan label dapat dilihat persentase hasil dari pelabelan, yang mana menunjukkan *tweet* negatif memiliki persentase 44% dengan 219 *tweet*, persentase komentar positif 30% dengan 151 *tweet* dan persentase *tweet* netral 26% dengan 130 *tweet* yang dapat dilihat dalam gambar 2.



Gambar 2. Hasil Pelabelan Data Pra Upacara Kemerdekaan 2024

Dan begitu juga data dari *tweet* pasca upacara kemerdekaan 2024, pada gambar 3 terlihat dari 500 data menunjukkan 50% (252 Label) memiliki hasil negatif dan kontra terhadap IKN, 33% (165 Label) memiliki hasil positif dan setuju terhadap Ibu Kota Nusantara (IKN), sedangkan sisanya yaitu 17% (83 Label) memilih netral dan tidak berpihak.



Gambar 3. Hasil Pelabelan Data Pasca Upacara Kemerdekaan 2024

Dalam tahap selanjutnya, *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan dalam penambangan teks dan pemrosesan bahasa alami untuk menilai seberapa penting sebuah kata dalam satu dokumen jika dibandingkan dengan kata-kata dalam kumpulan dokumen yang sama [14].

Tabel 5. TF-IDF

Dataset	TF-IDF
Data Pra Upacara Kemerdekaan 2024	menang 0.611902 dukung 0.271088 aneh 0.230038
Data Pasca Upacara Kemerdekaan 2024	perintah 0.324738



Dataset	TF-IDF
istana	0.324738
ikn	0.093363

Dari Tabel 5 menunjukkan Tabel ini menggunakan metode TF-IDF untuk menemukan kata-kata kunci yang paling unik dan paling khas yang mewakili percakapan publik tentang IKN pada dua periode waktu yang berbeda: sebelum dan sesudah Upacara Kemerdekaan 2024. Sederhananya, TF-IDF memberikan skor tinggi pada kata yang sering muncul di satu periode tetapi jarang di periode lainnya, sehingga ia bisa menyoroti fokus utama atau "jiwa zaman" dari percakapan pada waktu tersebut.

Dari tabel ini, terlihat jelas adanya pergeseran narasi yang sangat signifikan. Pada periode sebelum upacara, percakapan sangat diwarnai oleh nuansa politik pasca-pemilu. Hal ini dibuktikan dengan tingginya skor kata-kata seperti 'menang', 'dukung', dan 'aneh'. Kata-kata ini menunjukkan bahwa diskusi IKN saat itu sangat terkait erat dengan kontestasi politik, antara pihak yang mendukungnya sebagai bagian dari kemenangan dan pihak yang merasa kebijakan ini janggal atau aneh.

Namun, setelah upacara dilaksanakan, fokus percakapan publik bergeser total. Kata-kata yang menjadi paling unik adalah 'perintah' dan 'istana'. Pergeseran ini menandakan bahwa diskusi tidak lagi soal pro-kontra politik, melainkan telah beralih ke pembahasan mengenai eksekusi, otoritas, dan simbol-simbol konkret pemerintahan seperti pembangunan Istana dan instruksi dari presiden.

Beberapa kata yang memiliki bobot berdasarkan seberapa penting sebuah kata dalam satu dokumen yang kemudian dikonversikan menjadi angka, yang kemudian digunakan dalam proses analisis.

3.3 Analisa dan Hasil Klasifikasi Model

Dalam penelitian ini analisis dalam klasifikasi data menggunakan model algoritma *Naïve Bayes* dan model algoritma *Random Forest* serta Teknik *SMOTE* untuk penyeimbangan data yang timpang dari data positif, negatif dan netral . Dengan memanfaatkan *K-Fold Cross Validation* untuk menilai kinerja penelitian ini, yang menunjukkan berapa besar akurasi prediksi yang dibuat oleh model algoritma *Naïve Bayes* dan model aldiritma *Random forest*. Penelitian ini menggunakan data yaitu dari data *tweet* Pra upacara kemerdekaan Indonesia 2024 dan Pasca upacara kemerdekaan Indonesia 2024.

Setiap pengujian data dilakukan dengan melakukan *split* data *training* dan *testing* menggunakan *K-Fold Cross Validation* dengan nilai $k = 10$. Sehingga didapatkan dimana nilai k tertinggi dapat dilihat pada rentang $k-1$ sampai dengan $k-10$ di setiap dataset. Hasil dari pengujian dapat dilihat pada Tabel 6.

Tabel 6. Hasil Klasifikasi Model

Dataset	K-Fold Cross Validation	Algoritma	
		Naïve Bayes	Random Forest
Pra Upacara Kemerdekaan 2024	k- 1	75%	73%
	k- 2	74%	74%
	k- 3	75%	66%
	k- 4	63%	68%
	k- 5	72%	75%
	k- 6	72%	78%
	k- 7	78%	78%
	k- 8	75%	75%
	k- 9	75%	75%
	k- 10	73%	73%
	Rata -Rata Akurasi	73%	73%
Pasca Upacara Kemerdekaan 2024	k- 1	85%	82%
	k- 2	78%	86%
	k- 3	71%	89%
	k- 4	85%	84%
	k- 5	69%	68%
	k- 6	77%	77%
	k- 7	80%	84%
	k- 8	78%	82%
	k- 9	80%	81%
	k- 10	80%	78%
	Rata -Rata Akurasi	78%	81%

Dari Tabel 6 menunjukkan akurasi dari pengujian setiap 1000 data tweet yang dilakukan split data menjadi 500 data tweet Pra Upacara Kemerdekaan 2024 dan 500 tweet Pasca Upacara Kemerdekaan 2024 data dengan memakai pengujian K-Fold Cross Validation. Melalui uji coba yang sudah dilakukan dari dataset pertama yaitu data Pra Upacara Kemerdekaan 2024 menghasilkan akurasi tertinggi untuk model algoritma Naive Bayes pada split k-7 akurasi 78%

dan rata rata akurasi 73%, kemudian untuk hasil dari model algoritma Random Forest sendiri yaitu memiliki akurasi tertinggi pada split dan k- 6 dan k-7 yaitu dengan akurasi 78% dan rata rata akurasi 73%. Sedangkan untuk dataset Pasca Upacara Kemerdekaan 2024 untuk model algoritma Naive Bayes mendapatkan akurasi tertinggi pada split k- 1 dan k-4 yaitu 85% dan rata rata akurasi 78% dan untuk model algoritma Random Forest mendapatkan akurasi tertinggi pada k- 3 dengan akurasi 89% dan rata rata akurasi yaitu 81%.

Dari kedua algoritma yang digunakan menunjukkan model algoritma *Random Forest* memiliki performa yang lebih baik dari model algoritma *Naive Bayes* dalam melakukan analisis klasifikasi *dataset* penelitian ini, dengan rata rata akurasi yang dimiliki algoritma *Random Forest* yaitu 73% dan 81% sedangkan rata rata akurasi dari algoritma *Naive Bayes* yaitu 73% dan 78%, walaupun demikian dengan perbedaan akurasi yang sangat kecil antara kedua algoritma, dapat dikatakan bahwa kedua algoritma bekerja dengan baik dalam dataset yang menggunakan kedua algoritma tersebut.

3.4 Visualisasi

Berdasarkan hasil dari langkah – langkah proses yang telah selesai dilakukan, hasil penelitian dapat disajikan dalam bentuk visualisasi. Untuk melihat kata apa yang paling sering dibicarakan pada Pra Upacara Kemerdekaan 2024, menggunakan visualisasi *word cloud*. Hasilnya, pada sentimen positif, kata-kata seperti 'ikn', 'untuk', 'dukung', 'Ibukota', dan 'siap' paling sering muncul. Ini artinya, sentimen positif di sini sebenarnya adalah dukungan masyarakat pada pembangunan Ibu Kota Nusantara. Berbeda dengan sebelumnya, pada sentimen negatif, *word cloud* menyoroti penolakan terhadap gerakan boikot karena dampak buruknya. Lima kata yang paling sering ditemukan adalah 'ikn', 'jadi', 'jokowi', 'rakyat', dan 'proyek' sedangkan untuk visualisasi *wordcloud* netral kata yang sering muncul itu adalah 'ikn', 'nusantara', 'presiden', 'baru', dan 'jadi' yang dapat dilihat pada Gambar 4.



Gambar 4. Wordcloud Pra Upacara Kemerdekaan 2024

Sedangkan untuk gambar 5 visualisasi sentimen *Wordcloud* Pasca Upacara Kemerdekaan 2024 yaitu dari sudut pandang positif, kata-kata dominan seperti "dukung", "lanjut", dan "investasi" menggambarkan narasi yang penuh optimisme. Kelompok ini fokus pada dukungan terhadap pemerintah dan melihat Ibu Kota Nusantara sebagai sebuah proyek pembangunan yang harus terus berjalan demi kemajuan dan peluang ekonomi.

Narasi ini berbanding terbalik pada sentimen negatif. Di sini, kata-kata yang sama seperti "bangun" dan "Jokowi" diapit oleh kekhawatiran yang kuat, yang ditandai dengan besarnya kata "gagal", "uang", dan "rakyat". Ini menunjukkan bahwa kelompok kontra tidak hanya menolak, tetapi juga secara aktif mempertanyakan kelayakan proyek, transparansi biaya, dan dampaknya bagi masyarakat luas.

Di tengah polarisasi ini, ada kelompok netral. Percakapan mereka lebih didominasi oleh kata-kata informatif seperti "pemerintah", "jakarta", "rencana", dan "tanya". Ini menggambarkan adanya segmen masyarakat yang tidak memihak, namun aktif mencari fakta, memahami proses, dan mengajukan pertanyaan mengenai rencana pemindahan dari Jakarta.



Gambar 5. Wordcloud Pasca Upacara Kemerdekaan 2024

4. KESIMPULAN

Berdasarkan hasil analisis klasifikasi sentimen serta evaluasi model algoritma *Naive Bayes* dan model algoritma *Random Forest* terhadap data sebelum dan sesudah upacara kemerdekaan 2024, ditemukan bahwa sentimen negatif mendominasi dengan persentase yang hampir sama tinggi, yaitu sebesar 44% untuk data pra dan 50% untuk data pasca. Sementara itu, sentimen positif hanya mencapai 30% data pra dan 33% untuk data pasca dan netral sebesar 26%.



untuk data pra dan 17% untuk data pasca. Hal ini menunjukkan bahwa sebagian masyarakat Indonesia masih meragukan pembangunan Ibu Kota Negara (IKN). Keraguan ini disebabkan oleh banyaknya isu negatif yang beredar di masyarakat terkait proyek tersebut. Dalam proses analisis, digunakan dua model algoritma, yaitu model algoritma *Naive Bayes* dan model algoritma *Random Forest*, serta diterapkan teknik *balancing SMOTE* untuk mengatasi ketidakseimbangan data. Hasilnya yang didapat menunjukkan algoritma *Random Forest* menunjukkan performa yang lebih baik berdasarkan metode validasi *K-Fold Cross Validation*. Untuk data pra upacara kemerdekaan 2024, *Random Forest* mencatat akurasi rata-rata 95%, sedangkan pasca upacara kemerdekaan Indonesia 2024 mencapai 98%. Sementara itu, algoritma *Naive Bayes* memiliki akurasi sedikit lebih rendah, yaitu 94% untuk data pra upacara kemerdekaan Indonesia 2024 dan 97% untuk data pasca upacara kemerdekaan Indonesia 2024. Dengan demikian, *Random Forest* terbukti lebih efektif dalam mengklasifikasikan sentimen masyarakat terhadap isu pembangunan Ibu Kota Negara (IKN). Meskipun demikian, penelitian ini memiliki keterbatasan yang perlu dipertimbangkan seperti penggunaan algoritma *Random Forest*, walaupun *Random Forest* menunjukkan akurasi tinggi, keterbatasan inheren dari model klasifikasi sentimen mungkin belum sepenuhnya mampu menangkap nuansa bahasa yang kompleks seperti sarkasme, ironi, atau konteks budaya spesifik. Sehingga penelitian selanjutnya memungkinkan untuk mengeksplorasi penggunaan model algoritma yang lebih canggih, termasuk algoritma *deep learning* (misalnya, *LSTM*, *BERT*) yang mungkin lebih baik dalam menangkap konteks dan nuansa bahasa, serta mempertimbangkan fitur linguistik yang lebih kompleks.

REFERENCES

- [1] D. Pramana, M. Afdal, M. Mustakim, and I. Permana, "Analisis Sentimen Terhadap Pemandangan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, pp. 1306–1314, Jul. 2023, doi: 10.30865/mib.v7i3.6523.
- [2] T. C. Herdiyani and A. U. Zailani, "Sentiment Analysis Terkait Pemandangan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia," *Jurnal Teknologi Sistem Informasi*, vol. 3, no. 2, pp. 154–165, Sep. 2022, doi: 10.35957/JTSI.V3I2.2920.
- [3] R. Y. Liemawan, "Pengembangan Aplikasi Analisis Sentimen Terhadap Rencana Pemandangan Ibu Kota Negara Republik Indonesia," *KALBISIANA Jurnal Sains, Bisnis dan Teknologi*, vol. 9, no. 2, pp. 455–461, Jul. 2023, Accessed: Feb. 18, 2024. [Online]. Available: <http://ojs.kalbis.ac.id/index.php/kalbisiana/article/view/755>
- [4] R. K. Septiani, S. Anggraeni, and S. D. Saraswati, "Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *TEKNIKA: Jurnal Ilmiah Bidang Ilmu Rekayasa*, vol. 16, no. 2, pp. 245–254, Sep. 2022, doi: 10.5281/ZENODO.7535887.
- [5] A. Wahyu and A. Frinaldi, "Analisa Dampak Lingkungan Ibu Kota Nusantara," *Jurnal Kajian Hukum Dan Kebijakan Publik*, vol. 1, no. 2, pp. 29–33, Jan. 2024, doi: 10.62379/JKHKP.V1I2.97.
- [6] S. Aripriyanto, T. Tukino, A. Sufyan, and R. Nandaputra, "Sentimen Analisis Twitter Ibu Kota Negara Nusantara Menggunakan Long Short-Term Memory dan Lexicon Based," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 12, no. 2, pp. 119–125, Dec. 2022, doi: 10.36448/expert.v12i2.2821.
- [7] K. lubis, T. Ari Bangsa, and A. Yudertha, "Analisis Sentimen Opini Masyarakat Terhadap Pindahannya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naive Bayes," *Jurnal Teknoinfo*, vol. 18, no. 1, pp. 226–238, Jan. 2024, Accessed: Feb. 18, 2024. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/article/view/3708>
- [8] Y. Cahyana and A. Mutoi Siregar, "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *Faktor Exacta*, vol. 16, no. 3, pp. 1979–276, Oct. 2023, doi: 10.30998/faktorexacta.v16i3.16703.
- [9] E. R. Indriyani, Paradise, and M. Wibowo, "Perbandingan Metode Naive Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, pp. 1545–1553, Jul. 2022, doi: 10.30865/mib.v6i3.4220.
- [10] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naive Bayes Classifier," *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 524–531, Jul. 2022, doi: 10.47065/JOSH.V3I4.1944.
- [11] A. Susanto and I. Agung Dzulkarnain, "Analisis Sentimen Data Twitter Topik Ekonomi Dan Industri Dengan Metode Naive Bayes Dan Random Forest," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 20, pp. 59–65, Oct. 2023, doi: 10.5281/ZENODO.8398895.
- [12] D. J. Putnarubun, "Penerapan Metode Machine Learning-Naive Bayes Pada Analisis Sentimen Pemandangan Ibu Kota Negara Baru," *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 01, pp. 87–95, 2023.
- [13] S. Dwi, Prasetyo, S. Shofiah Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/KOMTEKINFO.V10I1.330.
- [14] D. Pramudita, Y. Akbar, and T. Wahyudi, "Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah pada Media Sosial X Menggunakan Algoritma Naive Bayes," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1420–1430, Aug. 2024, doi: 10.57152/MALCOM.V4I4.1565.
- [15] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *Jurnal Transformatika*, vol. 18, no. 1, pp. 71–80, Jul. 2020, doi: 10.26623/TRANSFORMATIKA.V18I1.2317.
- [16] H. Dhery, A. Assyam, and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 341–349, Aug. 2023, doi: 10.30865/KLIK.V4I1.957.



- [17] R. Sabian, “TA: Analisis Sentimen Publik terhadap Bjorka dalam Insiden Kebocoran Data KOMINFO menggunakan Algoritma Support Vector Machine,” Skripsi, Universitas Dinamika, Surabaya, 2023. Accessed: Jun. 27, 2024. [Online]. Available: <http://repository.dinamika.ac.id/id/eprint/6945>
- [18] Z. Cindy Dwinnie and R. Novita, “Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum Dan Sesudah Debat Calon Presiden Dan Wakil Presiden Tahun 2024 Tugas Akhir,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, pp. 758–767, 2024, doi: <https://doi.org/10.30865/mib.v8i2.7504>.
- [19] M. Y. Aldean, P. Paradise, and N. A. S. Nugraha, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac),” *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 4, no. 2, pp. 64–72, 2022.
- [20] M. Persada Pulungan, A. Purnomo, and A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 5, pp. 1033–1042, Oct. 2024, doi: 10.25126/JTIK.1077989.