BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

Analisis Sentimen Terhadap Kontroversi Pembangunan IKN di Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes*

Muhammad Dimas Rifqi Chasis Priyanto, Drs. Azahari , Muhammad Ibnu Sa'ad,
Program Studi Sistem Informasi, Stmik Widya Cipta Dharma, Samarinda, Indonesia
Email: [1]dimaschasis13@gmail.com, [2,*]azahari@wicida.ac.id [3,*] saad@wicida.ac.id
(* : coressponding author: dimaschasis13@gmail.com)

Abstrak -Pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur merupakan kebijakan strategis nasional yang memunculkan respons beragam dari masyarakat. Di satu sisi, kebijakan ini dinilai sebagai upaya pemerataan pembangunan, namun di sisi lain menuai kritik terkait dampak lingkungan, sosial, dan pembiayaan. Media sosial, khususnya *Twitter*, menjadi ruang yang ramai digunakan untuk menyampaikan opini publik mengenai isu ini. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pembangunan IKN yang diungkapkan melalui unggahan di media sosial *Twitter*. Dengan memahami kecenderungan sentimen publik, diharapkan dapat memberikan gambaran persepsi masyarakat yang bermanfaat sebagai bahan evaluasi dan pertimbangan bagi pemerintah dan pemangku kebijakan. Metode yang digunakan dalam penelitian ini adalah pendekatan kuantitatif berbasis data mining. Data dikumpulkan melalui teknik web crawling menggunakan library *snscrape*, kemudian diproses dengan tahapan *pre-processing* seperti *cleansing, case folding, tokenisasi, stopword removal, dan stemming*. Analisis sentimen dilakukan dengan pendekatan *lexicon-based* dan klasifikasi menggunakan algoritma *Naïve Bayes* yang didukung pembobotan TF-IDF. Dari 2.178 tweet yang dianalisis, diperoleh hasil bahwa sentimen positif mendominasi dengan persentase 52,4%, diikuti sentimen negatif 28,4%, dan netral 19,3%. Model klasifikasi menunjukkan tingkat akurasi sebesar 75,69%. Hasil ini menunjukkan adanya kecenderungan dukungan publik terhadap pembangunan IKN, sekaligus menegaskan pentingnya analisis sentimen sebagai alat bantu dalam membaca opini publik secara digital..

Kata Kunci: Ibu Kota Negara (IKN), Analisis Sentimen, Media Sosial, *Twitter*, *Naïve Bayes*, *Lexicon-Based*, TF-IDF

Abstract - The relocation of Indonesia's capital city (IKN) from Jakarta to East Kalimantan is a national strategic policy that has generated diverse public responses. On one hand, it is seen as an effort to promote equitable development, but on the other hand, it has drawn criticism related to its environmental, social, and financial impacts. Social media, particularly Twitter, has become a popular platform for expressing public opinion on this issue. This study aims to analyze public sentiment toward the IKN development as expressed through Twitter posts. By understanding public sentiment trends, this research seeks to provide insights into public perception that may serve as valuable input for government evaluation and policymaking. The research employed a quantitative approach using data mining techniques. Data were collected through web crawling using the *snscrape* library and underwent several pre-processing stages, including cleansing, case folding, tokenization, stopword removal, and stemming. Sentiment analysis was conducted using a lexicon-based approach, combined with a Naïve Bayes classification algorithm supported by TF-IDF weighting. Based on 2,178 analyzed tweets, the results showed that positive sentiment dominated at 52.4%, followed by negative sentiment at 28.4%, and neutral sentiment at 19.3%. The classification model achieved an accuracy rate of 75.69%. These findings indicate a general tendency of public support for the IKN development and highlight the importance of sentiment analysis as a strategic tool for interpreting public opinion in the digital era

Keywords: *Capital Relocation* (IKN), *Sentiment Analysis*, *Social Media*, *Twitter*, *Naïve Bayes*, *Lexicon-Based*, TF-IDF

1. PENDAHULUAN

Pemerintah Indonesia telah menetapkan rencana strategis nasional berupa pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur. Proyek ini tidak hanya berdampak pada aspek administratif pemerintahan, tetapi juga membawa konsekuensi besar terhadap lingkungan, sosial, ekonomi, dan teknologi. Meski pemerintah mengusung alasan pemindahan ini sebagai solusi atas permasalahan klasik Jakarta seperti kepadatan penduduk, kemacetan, serta krisis lingkungan, namun rencana tersebut memunculkan pro dan kontra di tengah Masyarakat [1]. Media sosial, khususnya Twitter, menjadi ruang publik digital di mana masyarakat dapat secara terbuka menyampaikan opini, kritik, maupun dukungan terhadap berbagai kebijakan pemerintah, termasuk isu pembangunan IKN [2].

Banyaknya interaksi dan volume data yang tersedia di *Twitter* membuka peluang besar untuk dilakukan analisis sentimen guna memahami persepsi publik secara luas dan real-time. Media sosial seperti *Twitter* memungkinkan masyarakat untuk mengekspresikan berbagai pendapat, mulai dari dukungan, kritik, hingga kekhawatiran terkait isu-isu penting, termasuk pembangunan Ibu Kota Nusantara (IKN). Dengan demikian, analisis sentimen menjadi alat yang sangat berguna untuk mengolah data teks dalam jumlah besar dan tidak terstruktur ini, sehingga dapat mengidentifikasi serta mengkategorikan opini publik ke dalam sentimen positif, negatif, atau netral. Dalam penelitian ini, digunakan metode *Naïve Bayes*, yaitu salah satu algoritma klasifikasi dalam pembelajaran mesin yang berbasis pada prinsip probabilitas. Metode ini dikenal memiliki keunggulan dalam hal efisiensi dan akurasi, khususnya dalam menangani data teks yang kompleks dan bervariasi. Selain itu, *Naïve Bayes* relatif mudah diimplementasikan dan dapat bekerja dengan baik meskipun dataset yang digunakan memiliki ukuran yang besar, sehingga sangat cocok untuk analisis sentimen di media sosial seperti *Twitter* [3].

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

Harapan dari penelitian analisis sentimen terhadap kontroversi pembangunan Ibu Kota Nusantara (IKN) di media sosial *Twitter* menggunakan metode *Naïve Bayes* adalah untuk memberikan gambaran yang lebih akurat mengenai persepsi dan opini publik terhadap kebijakan tersebut. Dengan penerapan metode *Naïve Bayes*, diharapkan klasifikasi sentimen masyarakat dapat dilakukan secara efektif, sehingga dapat menjadi dasar dalam penyusunan strategi komunikasi pemerintah yang lebih tepat sasaran. Selain itu, hasil analisis ini juga diharapkan mampu memberikan masukan bagi pengambil kebijakan agar lebih responsif terhadap kekhawatiran masyarakat, khususnya terkait isu lingkungan, anggaran, dan transparansi proses pembangunan. Penelitian-penelitian sebelumnya seperti yang dilakukan oleh [4], menunjukkan bahwa pendekatan ini efektif dalam menggali opini publik serta berperan penting dalam mendukung pengambilan keputusan yang berbasis data. Dengan demikian, penelitian ini tidak hanya berkontribusi secara akademik tetapi juga memiliki nilai praktis dalam konteks kebijakan publik.

Penelitian Sebelumnya yang dilakukan oleh [5], [6], serta [7] telah mengaplikasikan metode Naive Bayes untuk melakukan klasifikasi sentimen terhadap isu pemindahan Ibu Kota Negara (IKN) menggunakan data dari media sosial Twitter. Ketiga penelitian tersebut menunjukan efektivitas metode ini dalam membedakan opini publik ke dalam 3 kategori umum: positif,negatif, dan netral. Namun, *GAP* dari penelitian – penelitian sebelumnya terletak padaterbatasnya eksplorasi terhadap isu-isu tematik seperti kekhawatiran terhadap dampak lingkungan, pembiayaan proyek, serta respon sosial dari masyarakat terdampak. Selain itu, belum banyak studi yang mempertimbangkan aspek temporal dalam pengumpulan data, padahal opini publik bersifat dinamis dan dapat berubah seiring perkembangan isu.

Sebagian besar studi juga belum membandingkan performa Naive Bayes dengan algoritma lain secara sistematis,sehingga keunggulan relatif metode ini masih belum terverifikasi secara komprehensif. Beberapa studi seperti [8] juga belum mengintegrasikan pendekatan visualisasi atau teknik *topic modeling* untuk memperdalam konteks opini publik. Oleh karena itu, penelitian ini bertujuan untuk menjawab kekurangan tersebut dengan fokus pada analisis yang lebih tematik, memperluas cakupan waktu pengumpulan data, serta menyertakan visualisasi hasil dan pendekatan *Latent Dirichlet Allocation* (LDA) guna memahami konteks sentimen secara lebih mendalam dan aplikatif bagi pengambilan keputusan kebijakan publik.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kontroversi pembangunan Ibu Kota Nusantara (IKN) di media sosial *Twitter* menggunakan metode *Naïve Bayes*, dengan fokus pada penggalian opini publik berdasarkan isu-isu tematik seperti dampak lingkungan, sosial, dan ekonomi. Selain itu, penelitian ini juga berupaya menjawab keterbatasan studi sebelumnya dengan memperluas cakupan waktu pengambilan data, menerapkan visualisasi sentimen, dan mengintegrasikan pendekatan *topic modeling* guna memahami konteks percakapan secara lebih mendalam. Harapan yang ingin dicapai adalah memberikan gambaran yang komprehensif mengenai persepsi publik yang dapat dimanfaatkan sebagai masukan dalam perumusan kebijakan dan strategi komunikasi pemerintah. Penelitian ini juga diharapkan berkontribusi dalam pengembangan metode analisis sentimen berbasis data media sosial yang lebih akurat, kontekstual, dan aplikatif, serta menjadi pijakan bagi studi lanjutan dalam bidang kebijakan publik digital dan teknologi informasi.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif yang mengadopsi pendekatan data mining untuk mengeksplorasi dan menganalisis opini publik terkait pembangunan Ibu Kota Negara (IKN) melalui media sosial, khususnya platform Twitter. Media sosial dipilih sebagai sumber data karena merupakan wadah yang sangat aktif digunakan masyarakat untuk menyampaikan pendapat, kritik, maupun dukungan terhadap isu-isu aktual, termasuk kebijakan pemerintah seperti pemindahan Ibu Kota [8], [9],. Dalam konteks ini, data mining digunakan sebagai metode untuk menggali pola-pola tersembunyi dan memperoleh wawasan dari kumpulan data teks dalam jumlah besar yang diambil dari unggahan pengguna Twitter [10]. Metode analisis sentimen diterapkan untuk mengidentifikasi dan mengelompokkan opini masyarakat ke dalam tiga kategori utama, yaitu sentimen positif, sentimen negatif, dan sentimen netral [11]. Untuk melakukan klasifikasi tersebut, penelitian ini menggunakan algoritma Naïve Bayes, yang merupakan salah satu algoritma klasifikasi dalam machine learning yang sering digunakan dalam pengolahan teks karena kesederhanaannya, efisiensinya, dan performanya yang cukup baik untuk tugas-tugas klasifikasi berbasis teks [12]. Naïve Bayes bekerja dengan prinsip probabilistik dan mengasumsikan bahwa setiap fitur (dalam hal ini kata dalam tweet) bersifat independen satu sama lain. Dengan pendekatan ini, algoritma menghitung kemungkinan suatu teks termasuk dalam masing-masing kategori sentimen berdasarkan frekuensi kata-kata tertentu yang muncul dalam tweet [13]. Hasil dari klasifikasi ini diharapkan dapat memberikan gambaran umum mengenai persepsi masyarakat terhadap kebijakan pembangunan IKN, serta menjadi bahan pertimbangan dalam evaluasi kebijakan publik yang berbasis pada suara masyarakat secara daring [9].

2.2 Teknik Pengumpulan Data

Sumber data dalam penelitian ini adalah data sekunder berupa tweet yang diambil dari Twitter menggunakan teknik *web scraping*. Data dikumpulkan dengan bantuan library Python seperti snscrape, dengan kata kunci “IKN”, “pembangunan IKN”, dan sejenisnya. Serta periode pengambilan data disesuaikan dengan rentang waktu penelitian, yaitu

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

dari April hingga Mei 2025. Studi Pustaka, dilakukan untuk mencari, mempelajari, dan menggunakan berbagai literatur seperti, buku, jurnal, paper, e- book, atau literatur lain yang berhubungan dengan tema penelitian ini.

2.3 Tahapan Penelitian

Tahapan penelitian merupakan serangkaian prosedur yang disusun secara sistematis dan terstruktur untuk menggambarkan alur proses yang akan ditempuh selama penelitian. Proses ini sangat penting untuk memastikan bahwa setiap langkah dalam penelitian dapat dilaksanakan secara logis, runtut, dan sesuai dengan tujuan yang ingin dicapai. Dalam pelaksanaannya, penelitian ini menggunakan Google Colab sebagai alat bantu utama yang mendukung berbagai proses pengolahan data, mulai dari tahap awal hingga akhir. Google Colab dipilih karena menyediakan lingkungan pemrograman berbasis cloud yang fleksibel, mudah digunakan, serta mendukung berbagai pustaka Python yang dibutuhkan untuk analisis data dan pembangunan model. Penjabaran tahapan dilakukan secara menyeluruh dan detail, dimulai dari pengumpulan data mentah yang menjadi dasar utama penelitian, kemudian dilanjutkan dengan proses pra-pemrosesan (pre-processing) data yang mencakup pembersihan data, penghapusan simbol atau karakter yang tidak relevan, normalisasi teks, tokenisasi, hingga penghapusan stopwords untuk memastikan bahwa data dalam kondisi siap dianalisis. Setelah data siap, dilakukan pelabelan sentimen menggunakan pendekatan berbasis leksikon (lexicon-based) yang mengandalkan daftar kata dengan nilai sentimen tertentu untuk menentukan kategori sentimen dari masing-masing tweet. Selanjutnya, dilakukan pembangunan model klasifikasi menggunakan algoritma Naïve Bayes Classifier yang dikenal sederhana namun efektif dalam mengklasifikasikan teks berdasarkan probabilitas. Untuk mengukur kinerja model, digunakan metode evaluasi dengan confusion matrix yang mencakup metrik-metrik penting seperti akurasi, presisi, recall, dan f1-score. Akhirnya, hasil dari analisis ditampilkan dalam bentuk visualisasi guna memudahkan interpretasi dan pemahaman terhadap pola sentimen yang ditemukan. Rangkaian tahapan tersebut secara keseluruhan dapat dilihat pada Gambar 1 yang menggambarkan alur proses penelitian ini secara komprehensif dari awal hingga akhir.

Gambar 1.Alur Teknik Pengumpulan Data

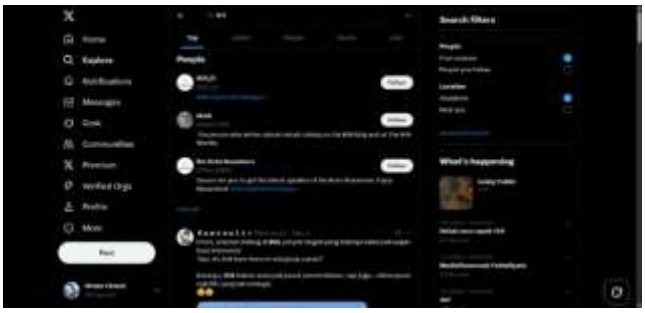
Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis, yaitu: pengumpulan data, pra-pemrosesan data, pelabelan sentimen, pembobotan fitur, pembangunan model klasifikasi, evaluasi model, dan visualisasi hasil. Data dikumpulkan dari media sosial Twitter, kemudian melalui tahapan pra-pemrosesan seperti cleansing, case folding, tokenisasi, stopword removal, dan stemming. Pelabelan sentimen dilakukan dengan pendekatan lexicon-based, sedangkan pembobotan fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Model klasifikasi dibangun menggunakan algoritma Naïve Bayes dan dievaluasi menggunakan confusion matrix dengan metrik seperti akurasi, presisi, dan recall. Seluruh hasil analisis kemudian divisualisasikan dalam bentuk diagram dan word cloud untuk memudahkan interpretasi.

Beberapa penelitian terdahulu telah membuktikan efektivitas pendekatan ini. Fachreza dan Yaqin [9] berhasil menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen masyarakat terhadap pemindahan IKN dengan tingkat akurasi yang baik. Sementara itu, Zamzami et al. [8] menunjukkan bahwa integrasi pendekatan lexicon-based dan pembobotan TF-IDF dapat meningkatkan kualitas analisis sentimen secara signifikan. Oleh karena itu, rancangan tahapan dalam penelitian ini didasarkan pada praktik terbaik dari studi-studi sebelumnya, dengan fokus pada efisiensi pemrosesan dan akurasi klasifikasi.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

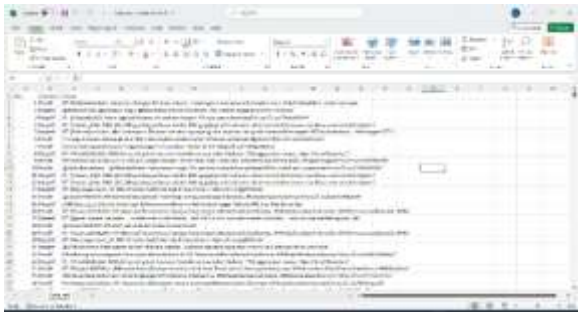


Gambar 2. Media Sosial Twitter

Pada penelitian ini, dataset yang digunakan berasal dari media sosial Twitter, dengan objek penelitian berupa cuitan atau tweet yang membahas kontroversi terkait pembangunan Ibu Kota Negara (IKN). Adapun tahapan proses pengolahan data akan dikembangkan hingga membentuk dataset yang siap digunakan untuk pengujian dalam penelitian ini. Gambar 2 menunjukkan tampilan salah satu akun atau postingan di media sosial Twitter yang relevan dengan topik pembangunan IKN.

Gambar 3. hasil crawling data dari repositori GitHub milik Sheila (2024) yang digunakan dalam penelitian ini.

Tahap awal dalam penelitian ini dimulai dengan pengumpulan data yang bersumber dari media sosial Twitter. Data yang digunakan bukan diperoleh secara langsung melalui proses crawling mandiri, melainkan memanfaatkan dataset yang telah dikumpulkan dan dipublikasikan secara terbuka oleh peneliti lain, yaitu Sheila, melalui repositori GitHub. Dataset tersebut bersifat publik dan telah mengalami proses awal seperti ekstraksi dan penyusunan dalam format terstruktur (.csv), sehingga sangat membantu dalam mempercepat proses penelitian ini. Isi dari dataset tersebut mencakup sejumlah informasi penting, antara lain teks tweet asli yang berkaitan dengan isu pembangunan Ibu Kota Negara (IKN), label sentimen yang menunjukkan kecenderungan opini (positif, negatif, atau netral), serta sejumlah metadata tambahan seperti ID pengguna, tanggal tweet, dan informasi akun Twitter. Dengan menggunakan dataset ini, peneliti tidak perlu lagi melakukan proses web crawling secara langsung, yang biasanya memakan waktu dan memerlukan konfigurasi teknis tambahan, melainkan dapat langsung fokus pada tahapan selanjutnya, yaitu pre-processing data, pelabelan sentimen, pembobotan, serta klasifikasi menggunakan metode yang telah ditentukan. Penggunaan dataset dari sumber terpercaya ini juga memberikan efisiensi dalam hal validitas data, karena telah tersedia dalam bentuk yang siap pakai untuk keperluan analisis. Adapun Gambar 2.3 berikut memperlihatkan tampilan awal dari data mentah yang diunduh melalui repositori GitHub, sebelum memasuki tahap pra-pemrosesan lebih lanjut.



Gambar 4. Tampilan dataset hasil crawling dari repositori GitHub milik Sheila (2024) yang telah disesuaikan dan digunakan dalam penelitian ini.

Dalam penelitian ini, data diperoleh dari repositori GitHub milik Syenira Sheila, yang menyediakan dataset hasil *crawling* dari media sosial Twitter terkait isu pembangunan Ibu Kota Negara (IKN). Dataset ini tersedia dalam format .csv dengan nama file Data.csv, dan berisi total 2.177 entri. Awalnya, file ini memuat beberapa kolom tambahan seperti informasi

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

pengguna dan metadata lainnya, namun karena terdapat kendala pada format encoding yang menyebabkan data tidak terbaca dengan baik, maka proses pembersihan dilakukan secara manual menggunakan Notepad. Setelah proses tersebut, dataset yang digunakan hanya memuat dua atribut utama, yaitu tweet yang berisi isi teks dari unggahan pengguna, dan sentimen yang menunjukkan label klasifikasi opini (positif, negatif, atau netral). Dataset ini kemudian digunakan dalam proses *preprocessing* dan analisis sentimen menggunakan algoritma klasifikasi.

Gambar 5. Tampilan Hasil Pemanggilan dan Visualisasi Data Tweet Menggunakan Google Colab

3.2 Pre-Processing

1. Cleansing

Setelah data berhasil dimuat ke dalam Google Colab, langkah selanjutnya yang dilakukan adalah tahap pembersihan (cleansing) terhadap data, khususnya pada bagian teks yang terdapat dalam kolom komentar atau isi tweet. Proses cleansing ini sangat penting karena teks mentah yang diambil langsung dari Twitter biasanya masih mengandung banyak elemen yang tidak relevan, seperti simbol-simbol khusus, tagar, tanda baca berlebihan, emoji, tautan URL, serta karakter lain yang dapat mengganggu proses analisis. . Langkah pembersihan dilakukan dengan menggunakan kode program Python di dalam Google Colab yang dirancang uchtntuk menghapus elemen-elemen tersebut dan mengubah teks menjadi bentuk yang lebih standar dan mudah diproses. Hasil dari proses ini disimpan dalam kolom baru pada dataset dengan nama cleansing, sedangkan kolom sebelumnya tetap dipertahankan sebagai referensi perbandingan. Dengan pembersihan ini, diharapkan setiap entri komentar memiliki struktur teks yang lebih bersih dan konsisten, sehingga mendukung keakuratan pada tahapan analisis sentimen yang akan dilakukan berikutnya. Contoh hasil dari proses pembersihan ini dapat dilihat pada Gambar 6 berikut.

Gambar 6. Hasil Tahapan Cleansing

Tahap pembersihan data merupakan proses awal yang sangat penting dalam pengolahan data, khususnya pada data teks seperti hasil scraping dari media sosial atau platform digital lainnya. Pada tahap ini, berbagai elemen yang dianggap tidak relevan atau mengganggu dianalisis dan dihapus. Elemen-elemen tersebut meliputi URL (tautan), tanda pagar (hashtag), sebutan pengguna (mention), emotikon, serta karakter khusus yang tidak memberikan kontribusi signifikan terhadap analisis. Tujuannya adalah untuk menyederhanakan dan menstandarkan data sehingga lebih mudah diproses, dianalisis, dan diinterpretasikan. Dengan data yang lebih bersih dan terstruktur, hasil analisis yang diperoleh pun akan menjadi lebih akurat, konsisten, dan sesuai dengan tujuan penelitian yang ingin dicapai.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

2. *Case Folding*

Gambar 7. Hasil Tahapan Case Folding

Setelah proses pembersihan data selesai, langkah berikutnya adalah melakukan pre-processing yang lebih rinci, salah satunya adalah case folding. Pada tahap case folding, semua huruf kapital dalam teks komentar diubah menjadi huruf kecil secara konsisten. Tujuan dari tahapan ini adalah untuk menyamakan format teks sehingga memudahkan analisis lebih lanjut. Gambar 7 menunjukkan perbandingan kondisi data sebelum dan sesudah proses case folding dilakukan.

3. Tokenization

Setelah huruf kapital pada teks tweet berhasil diubah menjadi huruf kecil, langkah berikutnya adalah melakukan tahap tokenisasi. Tokenisasi merupakan proses pemisahan satu kalimat dalam tweet menjadi beberapa kata atau token yang lebih kecil. Tahapan ini penting untuk memudahkan analisis teks secara lebih detail. Gambar 8 menunjukkan contoh kondisi data sebelum dan sesudah proses tokenisasi dilakukan.

Gambar 8 Hasil Tahapan Tokenization

4. Stopword/Removal

Setelah kalimat pada tweet berhasil dipisahkan menjadi beberapa kata melalui proses tokenisasi, langkah berikutnya adalah melakukan tahap penghilangan stopword. Tahap ini bertujuan untuk menghapus kata-kata yang dianggap kurang penting atau tidak memberikan makna signifikan dalam analisis, sehingga fokus hanya pada kata-kata yang relevan saja. Contoh hasil dari proses penghilangan stopword dapat dilihat pada Gambar 9.

Gambar 9 Hasil Tahapan Stopword/Removal

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

5. Stemming

Setelah proses penghilangan kata-kata yang tidak penting (stopword removal) selesai diterapkan pada teks tweet, langkah selanjutnya dalam tahapan pra-pemrosesan adalah **stemming**. Stemming merupakan salah satu komponen krusial dalam pemrosesan teks karena berfungsi untuk mengembalikan setiap kata ke bentuk dasarnya (root word atau stem) dengan cara menghapus imbuhan (afiks) yang melekat pada kata tersebut. Imbuhan ini bisa berupa awalan (prefiks), sisipan (infiks), maupun akhiran (sufiks) yang sering kali menyebabkan satu kata memiliki banyak bentuk turunan yang berbeda, meskipun maknanya masih berkaitan. Dengan melakukan stemming, sistem akan mampu memperlakukan kata-kata seperti "membangun", "pembangunan", dan "dibangun" sebagai satu entitas yang sama, yakni "bangun". Hal ini tidak hanya menyederhanakan representasi data, tetapi juga mengurangi kompleksitas dalam proses analisis karena menghindari redundansi istilah yang berlebihan. Sebagai hasilnya, analisis sentimen dan proses klasifikasi yang dilakukan pada data teks akan menjadi lebih efisien dan akurat, karena model dapat fokus pada makna dasar dari kata-kata yang dianalisis. Proses stemming dalam penelitian ini dilakukan menggunakan library atau pustaka Python yang kompatibel dengan bahasa Indonesia, sehingga mampu mengidentifikasi dan mengekstrak akar kata dengan baik berdasarkan aturan morfologi bahasa Indonesia. Gambar 3.9 memperlihatkan hasil perbandingan data sebelum dan sesudah dilakukan proses stemming. Dari gambar tersebut terlihat jelas perubahan kata-kata dalam tweet menjadi bentuk dasarnya, yang menandai keberhasilan tahap ini dalam mereduksi variasi kata dan menyederhanakan data untuk keperluan analisis lanjutan

Gambar 10. Hasil Tahapan Stemming

3.3 Pelabelan Lexicon Based

Setelah proses pre-processing menghasilkan data dalam variabel stemming_data, dilakukan tahapan pelabelan sentimen terhadap data tersebut. Proses pelabelan ini menggunakan pendekatan lexicon-based, yaitu dengan membandingkan kata-kata dalam setiap tweet dengan daftar kata positif dan negatif yang diperoleh dari repositori GitHub. Daftar kata tersebut merupakan hasil kurasi dari proyek analisis sentimen berbahasa Indonesia dan telah tersedia dalam format .tsv atau .txt. Setiap tweet yang telah melalui proses pre-processing diubah menjadi daftar token, lalu dianalisis berdasarkan jumlah kemunculan kata positif dan negatif sesuai dengan isi lexicon. Jika jumlah kata positif lebih banyak, maka tweet dikategorikan sebagai positive. Jika jumlah kata negatif lebih banyak, maka dikategorikan sebagai negative. Apabila jumlahnya sama atau tidak ditemukan kata yang cocok, maka tweet diklasifikasikan sebagai neutral. Berikut adalah potongan kode program yang digunakan untuk melakukan pelabelan sentimen secara otomatis berdasarkan pendekatan tersebut.

Gambar 11. Hasil coding fungsi analisis sentimen yang menghitung skor berdasarkan jumlah kata positif dan negatif dalam teks menggunakan pendekatan lexicon-based

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

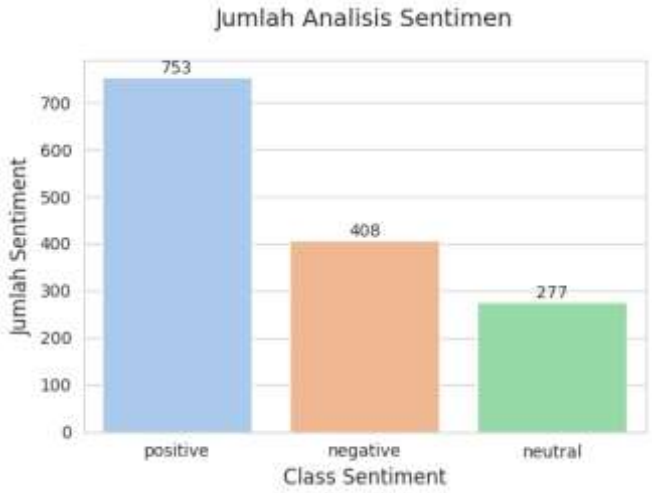
	word	weight
0	hai	3
1	merekam	2
2	ekstensif	3
3	paripurna	1
4	detail	2

Gambar 12 Tabel ini berisi kumpulan kata-kata positif yang diambil dari (Fajri Koto) kamus lexicon dan digunakan sebagai acuan untuk menentukan nilai sentimen positif dalam data tweet. Kata-kata ini berfungsi sebagai indikator untuk mengidentifikasi opini atau perasaan positif yang terkandung dalam teks, sehingga membantu dalam proses analisis sentimen secara lebih akurat dan terstruktur

	word	weight
0	putus tali gantung	-2
1	gelebah	-2
2	gobar hati	-2
3	tersentuh (perasaan)	-1
4	isak	-5

Gambar 13. Tabel kata-kata negatif dari kamus lexicon yang digunakan untuk mendeteksi sentimen negatif dalam tweet

Setelah proses pra-pemrosesan selesai, analisis sentimen dilakukan dengan menggunakan fungsi sentiment_analysis_lexicon_indonesia, yang menghitung skor sentimen secara sederhana berdasarkan kemunculan kata-kata dalam daftar kata positif dan negatif. Daftar kata tersebut diperoleh dari repositori GitHub yang menyediakan lexicon sentimen berbahasa Indonesia. Fungsi ini membandingkan setiap token dalam tweet dengan kata-kata yang ada di dalam lexicon. Setiap kemunculan kata positif akan menambah skor, sedangkan kata negatif akan menguranginya. Hasil akhir dari perhitungan skor ini digunakan untuk mengklasifikasikan setiap tweet ke dalam kategori *positif*, *negatif*, atau *netral*.



Gambar 14 Hasil Jumlah Analisis Sentimen

Hasil analisis sentimen terhadap data tweet yang telah melalui proses pra-pemrosesan menunjukkan bahwa setiap tweet dapat diklasifikasikan ke dalam tiga kategori utama, yaitu sentimen negatif, netral, dan positif. Proses klasifikasi ini dilakukan dengan metode perhitungan sederhana, yakni menghitung skor sentimen berdasarkan jumlah kemunculan kata-kata yang terdapat dalam daftar kata positif dan negatif yang diperoleh dari repositori GitHub. Jika skor akhir dari sebuah tweet menunjukkan nilai positif, maka tweet tersebut diklasifikasikan sebagai sentimen positif; jika skor bernilai negatif maka dikategorikan sebagai sentimen negatif; dan apabila skor berada pada posisi netral atau nol, maka dikategorikan sebagai sentimen netral. Skor Sentimen dari setiap tweet kemudian disimpan dalam struktur berupa DataFrame untuk mempermudah proses analisis lanjutan dan visualisasi hasil. Setelah seluruh data dianalisis dan diklasifikasikan, hasilnya divisualisasikan dalam bentuk diagram untuk menggambarkan distribusi sentimen secara keseluruhan. Pada Gambar 14, terlihat bahwa komentar dengan sentimen positif mendominasi jumlah data, yaitu sebanyak 753 tweet. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan tanggapan yang mendukung atau bersifat optimis terhadap isu pembangunan Ibu Kota

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

Negara (IKN). Komentar dengan sentimen negatif tercatat sebanyak 408 tweet, mencerminkan adanya kelompok pengguna yang memberikan respons kritis atau kurang setuju. Sementara itu, jumlah komentar dengan sentimen netral mencapai 277 tweet, yang mengindikasikan bahwa sebagian pengguna menyampaikan opini secara objektif atau tidak secara eksplisit mengekspresikan emosi terhadap topik tersebut.

Gambar 15 Hasil Presentase Klasifikasi Sentimen

Jika dilihat dari hasil presentasi klasifikasi sentimen pada Gambar 2.15, terlihat bahwa sentimen positif mendominasi dengan persentase sebesar **52.4%**, diikuti oleh sentimen negatif sebesar **28.4%**, dan sentimen netral sebesar **19.3%**. Hasil ini menunjukkan bahwa sebagian besar tweet yang dianalisis cenderung bersifat positif, mengekspresikan opini secara eksplisit terhadap kontroversi pembangunan Ibu Kota Negara (IKN). Persentase sentimen negatif dan netral kecil mengindikasikan bahwa hanya sedikit pengguna Twitter yang memberikan tanggapan secara langsung dalam bentuk dukungan atau kritik terhadap kebijakan tersebut. Temuan ini memperlihatkan bahwa dalam wacana publik di media sosial, mayoritas pengguna lebih memilih untuk menyampaikan informasi atau pandangan secara objektif menunjukkan sikap emosional yang kuat, baik dalam bentuk dukungan maupun penolakan.

3.4 Pembobotan TF-IDF

Setelah proses pelabelan sentimen selesai dilakukan, tahap selanjutnya adalah pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dengan memanfaatkan modul TfidfVectorizer dari library *scikit-learn*. . Pada tahap ini, digunakan sebanyak 1438 data tweet yang telah melewati tahapan pra-pemrosesan. Proses Proses pembobotan menghasilkan total 1006 kata unik yang mewakili berbagai istilah yang muncul dalam kumpulan data tersebut. Tahapan TF-IDF ini bertujuan untuk mengubah kumpulan dokumen teks menjadi representasi numerik dalam bentuk matriks istilah-dokumen. Proses ini memungkinkan setiap kata yang muncul dalam tweet diberi bobot berdasarkan seberapa sering kata tersebut muncul di satu dokumen dibandingkan kemunculannya di seluruh dokumen lainnya. Gambar 2.16 menjelaskan bagaimana kamus kata yang dihasilkan dari data digunakan dalam proses CountVectorizer atau TfidfVectorizer. Dalam proses ini, setiap kata diberikan indeks unik untuk mengidentifikasinya dalam vektor atau matriks. Misalnya, kata "indonesia" memiliki indeks 631, sedangkan kata "jokowi" memiliki indeks 722, dan seterusnya. Tahapan ini penting untuk mendukung proses analisis sentimen menggunakan metode Naïve Bayes karena memungkinkan algoritma untuk mengenali dan membedakan bobot pengaruh dari masing-masing kata dalam menentukan sentimen dari sebuah tweet.

Gambar 16 Daftar kata unik hasil tokenisasi yang dihitung berdasarkan frekuensi kemunculan, digunakan untuk analisis sentimen berdasarkan lexicon positif dan negatif.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

3.5. Permodelan Naïve Bayes

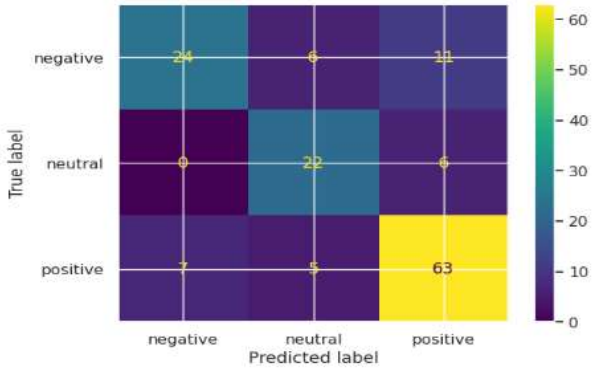
Setelah melalui tahapan pra-pemrosesan teks dan pembentukan daftar kata positif serta negatif, proses selanjutnya adalah melakukan perhitungan skor sentimen secara sederhana. Dalam pendekatan ini, setiap tweet dianalisis berdasarkan jumlah kemunculan kata-kata positif dan negatif. Skor dihitung dengan cara mengurangkan jumlah kata negatif dari jumlah kata positif, kemudian hasilnya digunakan untukmengklasifikasikan tweet ke dalam tiga kategori sentimen: **positif**, **negatif**, dan **netral**. Berdasarkan hasil klasifikasi terhadap 1438 tweet, diperoleh bahwa mayoritas tweet tergolong positif sebanyak 753 tweet (52,4%), disusul oleh sentimen negatif sebanyak 408 tweet (28,4%), dan sentimen netral sebanyak 277 tweet (19,3%). Gambar 17 memperlihatkan visualisasi hasil klasifikasi sentimen menggunakan pendekatan lexicon sederhana ini.

	precision	recall	f1-score	support
negative	0.77	0.59	0.67	41
neutral	0.67	0.79	0.72	28
positive	0.79	0.84	0.81	75
accuracy			0.76	144
macro avg	0.74	0.74	0.73	144
weighted avg	0.76	0.76	0.75	144

Gambar 17 Hasil permodelan Naïve Bayes

3.6. Confusion Matrix

Setelah melakukan klasifikasi sentimen terhadap kontroversi pembangunan IKN menggunakan metode Naïve Bayes, tahap berikutnya adalah evaluasi performa model dengan menggunakan confusion matrix pada data testing sebanyak 144 data. Distribusi sentimen pada data testing ini terdiri dari 63 data berlabel positif, 22 data netral, dan 24 data negatif. Confusion matrix digunakan untuk membandingkan hasil prediksi model dengan data sebenarnya sehingga dapat dihitung metrik evaluasi seperti **akurasi, presisi**, dan **recall**. Perhitungan manual dari metrik tersebut dapat dijelaskan sebagai berikut



Gambar 18 Pengujian Confusion Matrix

Perhitungan Precision dan Recall

Sentimen Positif

Diketahui:

TP =63
FP =11+6=17
FN = 7 + 5 = 12

Precision:

Precision = TP / (TP + FP)
= 63 / (63 + 17)
= 63 / 80
= 0.7875
=> 78.75%

Recall:

Recall = TP / (TP + FN)
= 63 / (63 + 12)
= 63 / 75
= 0.84
=> 84%

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

Sentimen Negatif
Diketahui:
TP = 48
FP = 14
FN = 34
Precision:
Precision = TP / (TP + FP)
= 48 / (48 + 14)
= 48 / 62
= 0.774
=> 77.4%
Recall:
Recall = TP / (TP + FN)
= 48 / (48 + 34)
= 48 / 82
= 0.585
=> 58.5%
Sentimen Netral
Diketahui
TP = 51
FP = 25.6
FN = 14
Precision:
Precision = TP / (TP + FP)
= 51 / (51 + 25.6)
= 51 / 76.6
= 0.666
= 66.6%
Recall:
Recall = TP / (TP + FN)
= 51 / (51 + 14)
= 51 / 65
= 0.785
= 78.5%

Tabel 3 Ringkasan Nilai TP, FP, FN

Sentimen	TP	FP	FN	Precision(%)	Recall(%)
Negatif	48	14	34	77.4%	58.5%
Netral	51	25.6	14	66.6%	78.5%
Positif	63	17	12	78.75%	84%

Visualisasi

Gambar 19 Word Cloud Sentimen negative

Setelah memperoleh hasil penghitungan jumlah dan persentase klasifikasi sentimen, langkah berikutnya adalah membuat visualisasi menggunakan word cloud yang dapat dilihat pada gambar 3.18 Visualisasi tersebut memperlihatkan bahwa pada ulasan dengan sentimen negatif, frekuensi kemunculan kata-kata tertentu meningkat, yang ditunjukkan dengan ukuran kata yang lebih besar dalam word cloud.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

Gambar 20 Word Cloud Sentimen netral

Sementara itu gambar 20 menampilkan hasil visualisasi netral, Dimana terdapat kata “ikn” , “ bangun ikn”, dll

Gambar 21 Word Cloud Sentimen positif

Sementara itu gambar 21 menampilkan hasil visualisasi ulasan positif contoh nya seperti kata, “ dukung”, “ manfaat” , dll.

3. KESIMPULAN

Penelitian ini menganalisis sentimen masyarakat terhadap kontroversi pembangunan Ibu Kota Negara (IKN) berdasarkan data dari media sosial Twitter. Metode yang digunakan menggabungkan pendekatan lexicon-based untuk pelabelan sentimen dan algoritma Naïve Bayes dengan pembobotan TF-IDF untuk klasifikasi, yang menghasilkan akurasi sebesar 75,17%. Dari 2.178 tweet yang dianalisis, sentimen positif mendominasi sebesar 52,7%, diikuti negatif 27,6% dan netral 19,7%. Hasil ini menunjukkan kecenderungan dukungan publik terhadap pembangunan IKN. Temuan tersebut memperlihatkan bahwa analisis sentimen berbasis media sosial dapat menjadi alat strategis dalam menangkap opini publik secara cepat dan luas. Meski demikian, penelitian ini memiliki keterbatasan, seperti fokus hanya pada data Twitter, keterbatasan metode dalam menangkap makna kontekstual seperti sarkasme, serta cakupan data yang bersifat snapshot. Selain itu, belum adanya klasifikasi berbasis isu spesifik mengurangi kedalaman analisis. Untuk itu, penelitian lanjutan disarankan menggunakan data lintas platform, pendekatan NLP yang lebih kompleks, dan analisis tematik guna memperoleh pemahaman yang lebih komprehensif terhadap persepsi publik.

REFERENCES

[1] H. Dhery, A. Assyam, and F. N. Hasan, “Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining,” *Media Online)*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.

[2] S. Lestari and M. Mupaat, “Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter,” *Sistem Informasi*, vol. 8, no. 1, pp. 13–22, Jun. 2022.

[3] N. Hadi and D. Sugiarto, “Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, Jan. 2025, doi: 10.30591/jpit.v10i1.7106.

[4] M. Reinaldy Destra Fachreza and M. Ainul Yaqin, “Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes,” 2023.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 97 - 108
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1993
https://journal.fkpt.org/index.php/BIT

[5] I. Dwi, N. #1, M. Hafid, and D. Irmayanti, "Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara (IKN) Pada Platform Twitter Menggunakan Metode Naïve Bayes," 2023. [Online]. Available: https://katadata.co.id

[6] M. Reinaldy Destra Fachreza and M. Ainul Yaqin, "Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," 2023.

[7] A. Munawaroh and R. Ridhoi, "SENTIMENT ANALYSIS DENGAN NAÏVE BAYES BERBASIS ORANGE TERHADAP RESIKO PEMBANGUNAN IKN," 2024.

[8] F. Zamzami, R. Hidayat, and R. Fathonah, "PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN KOMENTAR TWITTER PROYEK PEMBAGUNAN IKN," *Faktor Exacta*, vol. 17, no. 1, May 2024, doi: 10.30998/faktorexacta.v17i1.22265.

[9] M. Reinaldy Destra Fachreza and M. Ainul Yaqin, "Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," 2023.

[10] I. Dwi, N. #1, M. Hafid, and D. Irmayanti, "Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara (IKN) Pada Platform Twitter Menggunakan Metode Naïve Bayes," 2023. [Online]. Available: https://katadata.co.id

[11] A. Kurniawan and S. Waluyo, "Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter," 2022. [Online]. Available: https://senafti.budiluhur.ac.id/index.php

[12] N. S. Wardani, A. Prahutama, and P. Kartikasari, "ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA DENGAN KLASIFIKASI NAÏVE BAYES UNTUK MODEL BERNOULLI DAN MULTINOMIAL", [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[13] C. Huda and M. Betty Yel, "Analisa Sentimen Tentang Ibu Kota Nusantara (IKN) Dengan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI V*, vol. 7, no. 1, pp. 126–130, 2024.

[14] Fajri Koto, "Kamus Positif Dan Negatif," https://github.com/fajri91/InSet/blob/master/positive.tsv, https://github.com/fajri91/InSet/blob/master/negative.tsv.