
Analisis Sentimen dan Ujaran Kebencian Pemberitaan Online Tentang IKN Menggunakan Algoritma K-NN

Tirsa R. O. Tumimomor¹, Irene R.H.T. Tangkawarow², Audy A. Kenap³

19210080@unima.ac.id¹, irene.tangkawarow@unima.ac.id², audykenap@unima.ac.id³

^{1,2,3} Universitas Negeri Manado

Informasi Artikel

Diterima : 24 Mar 2025
Direvisi : 2 Apr 2025
Disetujui : 30 Apr 2025

Kata Kunci

Analisis Sentimen, Berita Online, Data Mining, K-Nearest Neighbor

Abstrak

Pemberitaan online tentang Ibu Kota Nusantara (IKN) menjadi topik perbincangan yang memicu berbagai reaksi masyarakat. Isu pemindahan ibu kota tergolong sensitif dan banyak dibahas di berbagai platform. Informasi yang beredar turut memengaruhi persepsi publik, terutama terkait ujaran kebencian yang dapat merusak citra positif IKN. Penelitian ini bertujuan menganalisis sentimen dalam pemberitaan online tentang IKN dengan metode K-Nearest Neighbor (KNN). Data diambil dari Google News (595 berita) dan YouTube (398 berita), lalu diklasifikasikan menjadi empat kelas: positif, negatif, netral, hate speech. Hasil penelitian menunjukkan akurasi tertinggi pada Google News mencapai 100%, sedangkan YouTube memperoleh 88,19% pada K=3. Kesimpulannya, data dari Google News lebih mudah diklasifikasikan menggunakan model KNN dibandingkan data dari YouTube.

Keywords

Data Mining, Ibu Kota Nusantara, K-Nearest Neighbor, Online News, Sentiment Analysis

Abstract

Online news about Ibu Kota Nusantara (IKN) has sparked diverse public reactions, particularly regarding the capital relocation, a highly sensitive topic. The spread of information shapes public perception, especially when news contains hate speech, which can damage IKN's reputation. This study applies sentiment analysis to online news about IKN using the K-Nearest Neighbor (KNN) algorithm. Data were gathered from Google News (595 articles) and YouTube (398 videos) and classified into four categories: positive, negative, neutral, and hate speech. The results show that Google News achieved 100% accuracy, while YouTube data reached 88.19% at K=3. These findings suggest that Google News articles are easier to classify with KNN compared to YouTube content, highlighting differences in text structure and characteristics between platforms.

A. Pendahuluan

Pemindahan Ibu Kota Negara (IKN) Indonesia dinilai sebagai langkah efektif dalam mewujudkan pemerataan ekonomi di Tanah Air[1]. Pembangunan Ibu Kota Nusantara dimulai pada pertengahan Maret 2022 dan akan dipindahkan secara bertahap dari tahun 2024 hingga tahun 2045. Pada Senin, 26 Agustus 2019, melalui siaran pers, diumumkan bahwa lokasi baru IKN terletak di Kabupaten Penajam Paser Utara, Provinsi Kalimantan Timur, mencakup sebagian wilayah Kabupaten Kutai Kartanegara. Selanjutnya, Rancangan Undang-Undang IKN disahkan menjadi Undang-Undang pada 18 Januari 2022 [2]. Pemindahan IKN bertujuan guna mempercepat laju pertumbuhan ekonomi serta mengurangi ketimpangan pembangunan di luar Pulau Jawa, khususnya di wilayah timur Indonesia. Pemerintah meyakini bahwa relokasi IKN ke pusat ekonomi nasional akan memberikan berbagai manfaat, seperti peningkatan Produk Domestik Bruto (PDB), pemerataan pembangunan, perluasan kesempatan kerja, pengurangan angka kemiskinan, serta peningkatan kesejahteraan masyarakat secara keseluruhan[4].

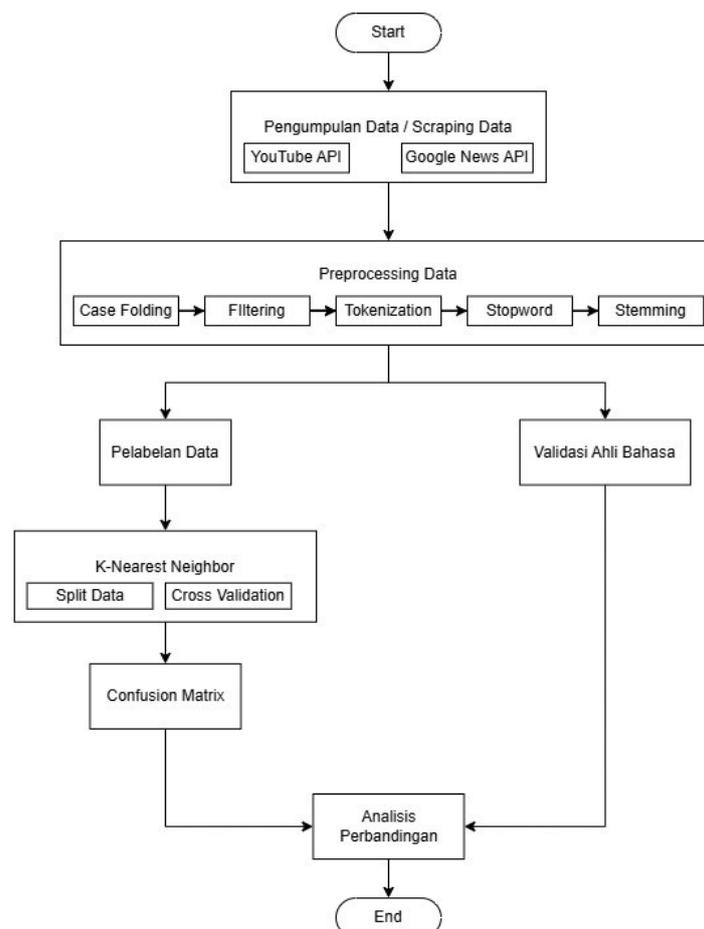
IKN sebagai ibu kota baru Indonesia menjadi topik yang banyak dibahas dalam berbagai media online. Pemberitaan mengenai IKN mencakup berbagai perspektif, baik yang bersifat positif, netral, maupun negatif. Namun, Penyebaran informasi melalui internet saat ini berlangsung dengan sangat cepat dan mudah, tanpa terhalang oleh waktu maupun jarak[5]. Di era digital saat ini, penyebaran informasi tidak selalu terkendali, sehingga berpotensi memunculkan ujaran kebencian yang dapat memengaruhi opini publik secara negatif. Ujaran kebencian dalam konteks permasalahan ini diartikan sebagai pernyataan yang mengandung diskriminasi, ancaman, permusuhan, atau prasangka terkait dengan isu-isu seperti gender, suku, agama kepercayaan, suku, ras, kebangsaan, keterbatasan fisik, atau preferensi seksual[6]. Dari perspektif hukum, ujaran kebencian mengacu pada pernyataan, tindakan, tulisan, atau pertunjukan yang tidak diizinkan karena dapat memicu kekerasan serta mendorong diskriminasi, baik oleh pihak yang menyampaikan maupun yang terkena dampaknya[7]. Pada surat Edaran Nomor SE/06/X/2015 tentang Penanganan Ujaran Kebencian yang diterbitkan oleh Kapolri menjelaskan bahwa ujaran kebencian merupakan tindakan yang dapat dikategorikan sebagai tindak pidana dalam KUHP. Dalam konteks hukum, ujaran kebencian merujuk pada ungkapan, perilaku, tulisan, atau pertunjukan yang dilarang karena berpotensi menimbulkan kekerasan dan mendorong tindakan diskriminatif, baik oleh pelaku maupun pihak yang terdampak[8]. Berita online juga menjadi salah satu inovasi teknologi yang memungkinkan banyak orang memperoleh informasi melalui internet, sehingga berita dapat tersebar dengan lebih cepat [9].

Klasifikasi teks merupakan bagian dari text mining dan berfungsi sebagai teknik untuk memprediksi kategori kelas dari suatu data. Beberapa contoh metode klasifikasi yang digunakan untuk melakukan klasifikasi teks adalah metode Naive Bayes Classifier, Support Vector Machine, Logistic Regression, K-Nearest Neighbor, dan Neural Network[10]. Analisis sentimen bertujuan untuk mengklasifikasikan teks dalam sebuah kalimat berdasarkan opini yang terkandung di dalamnya. Aspek atau data dokumen digunakan untuk menentukan apakah suatu kalimat atau dokumen memiliki sentimen positif, negatif, maupun netral [11]. Proses ini dapat dilakukan dengan menerapkan model analisis text mining [12]. Beberapa penelitian

yang telah dilakukan sebelumnya tentang klasifikasi sentiment diantaranya menunjukkan kinerja analisis dengan akurasi mencapai 88,12%, precision sebesar 93,98%, serta recall sebesar 81,53% [12]. Penelitian lainnya yang menganalisis ujaran kebencian menggunakan knn memperoleh nilai pengujian dengan akurasi 35% presisi 20% dan recall 100% pada nilai K=3 [13]. Berdasarkan uraian sebelumnya, penelitian ini akan menganalisis sentimen pemberitaan online terhadap IKN. K-NN diterapkan untuk mengklasifikasikan berita online mengenai IKN ke dalam empat kategori: positif, negatif, netral, dan *hate speech*. Dua sumber utama digunakan sebagai data penelitian, yaitu Google News dan YouTube, yang memiliki karakteristik konten berbeda. Penelitian ini bertujuan untuk mengevaluasi efektivitas model K-NN dalam mendeteksi hate speech, membandingkan kinerja deteksi pada Google News dan YouTube, serta menganalisis sejauh mana hasil klasifikasi model mendekati analisis ahli bahasa

B. Metode Penelitian

penelitian ini menerapkan metode klasifikasi K-NN. Namun untuk memperoleh hasil terbaik, pada penelitian ini akan dilakukan berbagai tahapan.



Gambar 1. Metode Penelitian

Tahapan Penelitian dapat diuraikan sebagai berikut :

1. Pengumpulan Data (Data Crawling)

Penelitian ini mengimplementasikan teknik pengumpulan data melalui pengumpulan berita online dari berbagai sumber yang secara khusus membahas mengenai Ibu Kota Nusantara (IKN). Proses ini memanfaatkan aplikasi *web scraping* pada *platform* Google News untuk mengakses dan mengumpulkan berita secara sistematis. Selain itu, data juga diperoleh dari video YouTube yang selaras dengan fokus penelitian. Data dihimpun dengan memanfaatkan bahasa pemrograman Python serta memanfaatkan Google Colaboratory sebagai platform pengembangannya. [14].

2. Preprocessing Data

Preprocessing merupakan proses menyiapkan data mentah untuk dianalisis dalam model pembelajaran pada mesin. Tahap pemrosesan dilakukan untuk mengolah data mentah menjadi kumpulan data yang siap digunakan, sehingga memungkinkan pemilihan data yang relevan untuk diproses dalam dokumen [12]. Tahapan ini mencakup, *case folding, tokenization, stopword removal, stemming*).

3. Pelabelan Data

Pelabelan data merupakan proses menetapkan kategori berdasarkan karakteristik yang terdapat pada sebuah dokumen [13]. Proses pelabelan dataset akan lebih mudah jika jumlah data relatif sedikit, namun bisa menjadi sangat memakan waktu dan sulit dilakukan secara mandiri jika dataset yang digunakan sangat besar. Dalam penelitian ini, kamus InSet Lexicon akan digunakan untuk melakukan pelabelan otomatis pada setiap berita. Kamus InSet Lexicon terdiri dari kumpulan kata-kata yang dikelompokkan berdasarkan nilai positif, negatif, dan ujaran kebencian.

4. K-Nearest Neighbor

K-NN merupakan salah satu metode dalam pembelajaran mesin yang berfungsi untuk proses klasifikasi. Metode ini bekerja dengan mengklasifikasikan data berdasarkan tingkat kedekatan atau jarak suatu data dengan data lainnya [15]. Perhitungan Euclidean distance digunakan untuk menentukan jarak antar data [13]. Perhitungan jarak euclidean menggunakan persamaan [16]:

$$d_i = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Pada tahap ini, metode K-NN dihitung dan diproses dengan menggunakan dua metode yaitu *Cross Validation* dan *Split Data* untuk mencari akurasi terbaik.

5. Confusion Matrix

Confusion matrix adalah tabel yang menampilkan jumlah dari data uji yang diklasifikasikan sebagai benar atau salah oleh model yang digunakan. Tabel confusion matrix diperlukan sebagai acuan dalam menentukan performa terbaik dari model klasifikasi [17]. Metode ini sering digunakan untuk mengukur tingkat akurasi suatu model [18]. Tabel confusion matrix dalam penelitian ini disesuaikan dengan menggunakan empat kelas yaitu kelas positif (P), kelas negatif (N), kelas netral (Ne), dan kelas hate speech (HS). Tabel confusion matrix dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Nilai Prediksi	Nilai Aktual			
	Positif (P)	Negatif (N)	Hate speech (HS)	Netral (NE)
Positif	TP	FP	FP	FP
Negatif	FN	TN	FN	FN
Hate speech	FHS	FHS	THS	FHS
Netral	FNE	FNE	FNE	TNE

6. Validasi Ahli Bahasa

Pengujian ini melibatkan ahli bahasa dalam menilai tingkat akurasi dan keandalan model K-Nearest Neighbor (KNN) dalam mengklasifikasikan berita online terkait Ibu Kota Negara Nusantara (IKN) berdasarkan sentimen. Evaluasi ini bertujuan untuk memastikan bahwa model mampu mengidentifikasi sentimen dengan tepat, sehingga hasil klasifikasi lebih valid dan sesuai dengan konteks linguistik yang digunakan dalam berita.

7. Analisis Perbandingan

Tahapan validasi oleh ahli bahasa dilakukan dengan membandingkan hasil klasifikasi sentimen yang dihasilkan oleh model K-Nearest Neighbor (KNN) dengan hasil analisis berbasis InSet lexicon serta evaluasi manual dari ahli bahasa sebagai *ground truth*. Dalam proses ini, setiap berita yang diklasifikasikan oleh model akan dianalisis kembali menggunakan pendekatan leksikal dari InSet lexicon untuk mengukur konsistensi sentimen yang dihasilkan. Selanjutnya, hasil klasifikasi model dan InSet lexicon dibandingkan dengan penilaian ahli bahasa guna menilai sejauh mana akurasi dan keandalan model dalam merefleksikan sentimen yang sesuai dengan standar linguistik.

C. Hasil dan Pembahasan

1. Pengumpulan Data

Proses data crawling pada Google News dan YouTube difilter Berita yang diambil adalah berita yang berbahasa Indonesia dengan *keyword* "ikn ibu kota nusantara". Pengambilan data pada YouTube dilakukan secara otomatis menggunakan skrip yang telah dirancang untuk mengonversi konten video menjadi teks. Video yang dipilih merupakan video yang memiliki subtitle guna mempermudah proses ekstraksi dan analisis teks. Teknik pengumpulan data ini menggunakan API (*Application Programming Interface*) dari Google News dan YouTube untuk mengakses data secara langsung dan terstruktur. Total berita yang dikumpulkan terdiri dari 398 berita dari YouTube dan 599 berita dari Google News, yang kemudian disimpan dalam format .xlsx.

Tabel 2. Crawling Data YouTube

No.	URL	Text
1.	https://www.youtube.com/watch?v=qtdwTcloC4w	ini adalah istana terindah di dunia katanya art sence and technolog adalah bagian multidisiplin untuk pembuatan mahakarya ini setelah9 tahun sebuah istana yang merepresentasikan identitas negeri kita..
2.	https://www.youtube.com/watch?v=a3Fur7CyRIY	Hai inilah suku-suku yang mendiami wilayah IX Nusantara Hai seperti diketahui karena sudah banyak infonya Total luas wilayah ibukota negara Nusantara di Kalimantan Timur adalah 320 4331...
3.	https://www.youtube.com/watch?v=iE8KbrBNcb4	kalau memang presiden gagal dan saya sebut tolol ya kebijakannya yang tolol Nah itu misalnya soal ikn itu konteksnya ikn saya itu kan kebijakan tolol udah ditolak di mana-mana masih didagangin juga tuh pengamat politik Rocky Gerung...
4.	https://www.youtube.com/watch?v=yMZYS6Z4ejk	[tribuners beredar video di media sosial yang memperlihatkan tiktok tiktoker yang mengaku bule Eropa dengan akun tiktok @ombule WNI lovers yakni yakni ia melontarkan kalimat yang bernada penghinaan terhadap pemerintah negara Indonesia dan pembangunan Ibu Kota Nusantara atau IKN yang sedang berjalan di Kalimantan Timu

Tabel 3. Crawling Data Google News

No.	URL	Text
1.	https://prokopim.bengkaliskab.go.id/web/detailberita/15643/bupati-bengkalis-ikuti-rakornas-otorita-ibu-kota-nusantara-(ikn)-	Usai mengikuti Rakornas. Wakil Bupati Bengkalis H Bagus Santoso menilai, pentingnya sinergi antara Pemda dan IKN dalam upaya membangun kota yang lebih maju,...
2.	https://www.voaindonesia.com/a/ibu-kota-nusantara-merusak-hutan-atau-memperbaiki-lingkungan-/7106732.html	Ibu Kota Nusantara (IKN) dibangun di atas lahan yang berstatus hutan. Pakar menilai pembangunan ini sama saja dengan merusak kawasan meski pemerintah...

No.	URL	Text
3.	https://www.cnnindonesia.com/nasional/20230803160423-12-981593/persoalan-proyek-ikn-dan-umpatan-rocky-gerung	Pernyataan "bajingan tolol" yang disampaikan Rocky Gerung kepada Jokowi dinilai umpatan khas oposisi seperti Rocky untuk menyikapi kebijakan pemerintah...
4.	https://kompaspedia.kompas.id/baca/paparan-topik/percepatan-pembangunan-ikn-nusantara	Pemindahan Ibu Kota Negara dari Jakarta ke Nusantara direncanakan dilakukan pada tahun 2024. Upacara Kemerdekaan ke-79 Republik Indonesia pada 17 Agustus...

2. Preprocessing Data

Tahapan selanjutnya setelah data selesai dikumpulkan yaitu melakukan tahap *text preprocessing*. Tahap ini dilakukan untuk mengolah data teks dari pemberitaan online agar lebih mudah dipahami dan diproses dalam analisis lebih lanjut. Text berita sering kali mengandung berbagai elemen yang perlu dibersihkan atau disesuaikan agar sesuai dengan format yang lebih standar. Tahapan dalam *preprocessing* terdiri dari, *case folding*, *tokenization*, *filtering*, *stopwords*, dan *stemming* yang selanjutnya disimpan dalam format .xlsx. Berikut merupakan penjelasan secara rinci mengenai tahap *text preprocessing*.

2.1. Case Folding

Proses ini mengonversi seluruh teks ke dalam huruf kecil (lowercase) guna menjaga konsistensi dalam pemrosesan kata [19].

Tabel 4. Case Folding

Text	Case Folding
Pemindahan Ibu Kota Negara dari Jakarta ke Nusantara direncanakan dilakukan pada tahun 2024. Upacara Kemerdekaan ke-79 Republik Indonesia pada 17 Agustus...	pemindahan ibu kota negara dari jakarta ke nusantara direncanakan dilakukan pada tahun 2024. upacara kemerdekaan ke-79 republik indonesia pada 17 agustus...

2.2. Filtering

Data yang telah diperoleh setelah melalui proses case folding selanjutnya akan difilter. Proses filtering bertujuan untuk menghilangkan komponen yang tidak diperlukan, seperti tanda baca, angka, dan karakter simbol, agar teks lebih bersih dan terstruktur [20].

Tabel 5. Filtering

Text	Filtering
pemindahan ibu kota negara dari jakarta ke nusantara direncanakan dilakukan pada tahun 2024. upacara kemerdekaan ke-79 republik indonesia pada 17 agustus...	pemindahan ibu kota negara dari jakarta ke nusantara direncanakan dilakukan pada tahun upacara kemerdekaan ke republik indonesia pada agustus

2.3. Tokenization

Selanjutnya data yang telah melewati proses filtering akan diproses lebih lanjut dengan membagi setiap kalimat dalam korpus menjadi unit yang lebih kecil, yang disebut token [21]. Tahapan ini bertujuan untuk membagi teks ke dalam kata-kata atau frasa yang lebih sederhana untuk dianalisis.

Tabel 6. Tokenization

<i>Text</i>	<i>Tokenization</i>
pemindahan ibu kota negara dari jakarta ke nusantara direncanakan dilakukan pada tahun upacara kemerdekaan ke republik indonesia pada agustus	['pemindahan', 'ibu', 'kota', 'negara', 'dari', 'jakarta', 'ke', 'nusantara', 'direncanakan', 'dilakukan', 'pada', 'tahun', 'upacara', 'kemerdekaan', 'ke', 'republik', 'indonesia', 'pada', 'agustus']

2.4. Stopword

Stopword adalah kata yang kerap muncul dalam suatu teks tetapi tidak mempunyai kontribusi yang signifikan dalam analisis, contohnya "dan", "yang", "di", "ke", dan sejenisnya. Penghapusan stopwords dilakukan untuk menyaring kata-kata yang tidak penting [22] supaya analisis lebih terarah pada kata yang mempunyai arti lebih signifikan dalam konteks.

Tabel 7. Stopword

<i>Text</i>	<i>Stopword</i>
['pemindahan', 'ibu', 'kota', 'negara', 'dari', 'jakarta', 'ke', 'nusantara', 'direncanakan', 'dilakukan', 'pada', 'tahun', 'upacara', 'kemerdekaan', 'ke', 'republik', 'indonesia', 'pada', 'agustus']	['pemindahan', 'ibu', 'kota', 'negara', 'jakarta', 'nusantara', 'direncanakan', 'dilakukan', 'tahun', 'upacara', 'kemerdekaan', 'republik', 'indonesia', 'agustus']

2.5. Stemming

Stemming adalah proses untuk mengubah kata menjadi bentuk dasar dengan menghapus imbuhan, sehingga teks lebih seragam dan mudah dianalisis [23]. Contohnya, kata "bermain" diubah menjadi "main" dan "terbaik" menjadi "baik." Proses ini penting dalam pemrosesan bahasa alami (NLP) untuk meningkatkan akurasi pencarian informasi, klasifikasi teks, dan analisis data.

Tabel 8. Stemming

<i>Text</i>	<i>Stemming</i>
['pemindahan', 'ibu', 'kota', 'negara', 'jakarta', 'nusantara', 'direncanakan', 'dilakukan', 'tahun', 'upacara', 'kemerdekaan', 'republik', 'indonesia', 'agustus']	['pindah', 'ibu', 'kota', 'negara', 'jakarta', 'nusantara', 'rencana', 'laku', 'tahun', 'upacara', 'merdeka', 'republik', 'indonesia', 'agustus']

3. Pelabelan Data

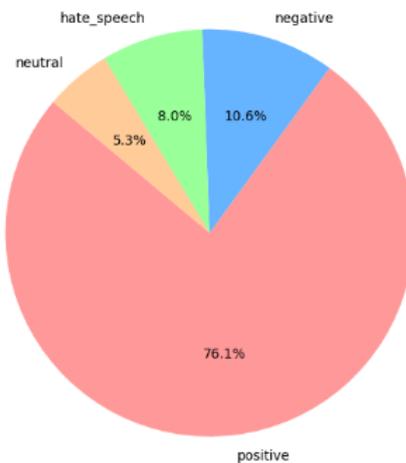
Pada tahap ini, setiap berita dalam dataset diberikan label sesuai dengan sentimen: Positif, Negatif, *Hate Speech*, dan Netral. Penelitian ini menggunakan kamus kata sentimen dengan mencocokkan kata dalam teks dengan daftar kata dalam kamus. Karena tidak semua sentimen Negatif mengandung ujaran kebencian, tetapi ujaran kebencian selalu bersentimen Negatif, kata-kata yang sama di kedua

kamus akan dihapus untuk memastikan sistem memilih kelas yang tepat. Setiap kata dalam berita dianalisis berdasarkan skor sentimen yang terdapat dalam Indonesian Sentiment Lexicon (InSet) untuk menentukan kategori sentimennya.

Tabel 9. Hasil Pelabelan YouTube

No.	Text.	Label
1	istana indah dunia kata art sence and technolog bagi multidisiplin buat mahakarya tahun buah istana representasi identitas negeri	positif
2	hai ini sukusuku diam wilayah ix nusantara hai tahu banyak info total luas wilayah ibukota negara nusantara kalimantan timur	positif
3	kalau memang presiden gagal sebut tolol bijak tolol nah misal soal ikn konteks ikn kan bijak tolol udah tolak manamana didagangin tuh amat politik rocky gerung	hate_speech
4.	tribuners edar video media sosial lihat tiktok tiktoker aku bule eropa akun tiktok ombule wni lovers lontar kalimat nada hina perintah negara indonesia bangun ibu kota nusantara ikn sedang jalan kalimantan timur	hate_speech

Persentase dari hasil pelabelan YouTube divisualisasikan dengan diagram pie chart.



Gambar 1. Pie Chart Persentase Pelabelan YouTube

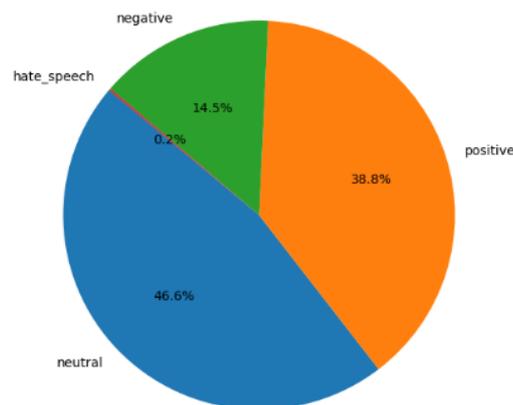
Dari 398 berita YouTube yang dilabeli, 303 berita atau 76,1% sentimen Positif, 42 berita atau 10,6% sentimen Negatif, 21 berita atau 5,3% sentimen Netral, dan 32 berita atau 8,0% merupakan sentimen *Hate Speech*.

Setelah memperoleh hasil pelabelan YouTube, selanjutnya data dilakukan pelabelan pada Google News.

Tabel 10. Hasil Pelabelan Google News

No.	Text.	Label
1	usai ikut rakornas wakil bupati bengkal h bagus santoso nilai penting sinergi pemda ikn upaya bangun kota lebih maju	positif
2	ibu kota nusantara ikn bangun atas lahan status hutan pakar nilai bangun sama dengan rusak kawasan meski perintah	negative
3	nyata bajing tolol sampai rocky gerung jokowi nilai umpat khas oposisi rocky sikap bijak pemerintah	hate_speech
4.	pindah ibu kota negara jakarta nusantara rencan laku tahun upacara merdeka republik indonesia agustus	neutral

Selanjutnya persentase dari hasil pelabelan Google News divisualisasikan dengan pie chart dapat dilihat pada Gambar 2.

**Gambar 2.** Pie Chart Persentase Pelabelan Google News

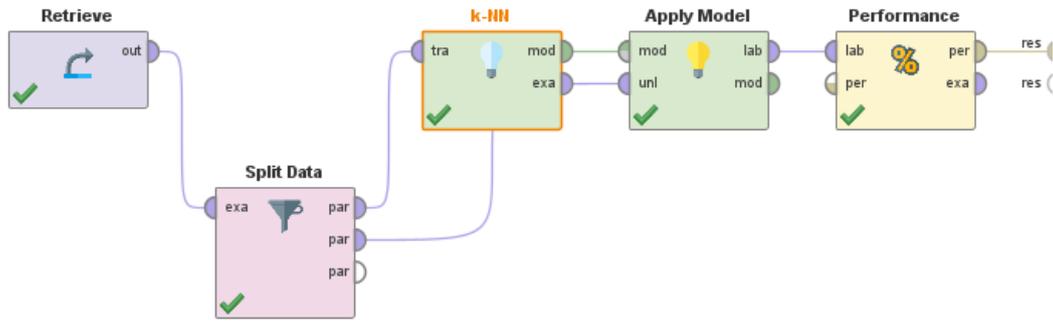
Dari 595 berita Google News yang dilabeli, 277 berita atau 46,6% bersentimen Netral, 231 berita atau 38,8% sentiment Positif, 86 berita atau 14,5% sentimen Negatif, dan 1 berita atau 0,2% mengandung sentimen *Hate Speech*.

4. K-Nearest Neighbor

Selanjutnya, akan dilakukan klasifikasi sentimen menggunakan algoritma K-NN. Pada tahap ini akan diuji menggunakan performa aplikasi pada dua metode yaitu Split Data dan Cross Validation.

4.1. Split Data

Dalam tahap Split Data, peneliti membuat dataset menjadi dua yaitu data latih serta data uji, dengan rasio perbandingan 80:20 pada $k = 3, 5, 7, 9$ untuk memperoleh akurasi terbaik. Data latih berfungsi untuk membangun model, dan data uji digunakan untuk mengevaluasi akurasi prediksi model.



Gambar 3. Proses Modeling Split Data

Setelah menerapkan model langkah selanjutnya menentukan nilai K untuk memperoleh akurasi tertinggi.

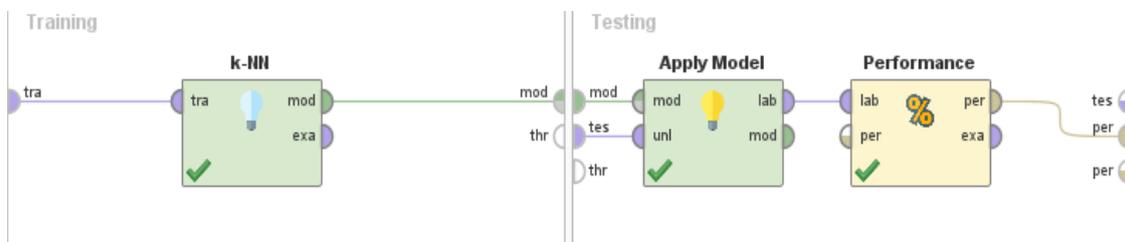
Tabel 11. Perbandingan Akurasi dengan Cross Validation

Nilai K	Akurasi	
	YouTube	Google News
3	86.08%	100%
5	82.28%	100%
7	84.81%	100%
9	84.81%	100%

Berdasarkan pada hasil perbandingan maka dapat disimpulkan bahwa akurasi tertinggi sebesar 86,08% pada K=3 pada pemberitaan YouTube dan 100% pada pemberitaan Google News untuk semua nilai K.

4.2. Cross Validation

Pada tahapan ini, data latih dan data uji displit secara acak menjadi 5 bagian (number of folds = 5) Prosesnya melibatkan pembagian Data dibagi menjadi 'K' bagian dengan ukuran yang setara. Proses pelatihan dan pengujian dilakukan sebanyak 'K' kali[16]. Selanjutnya untuk pemilihan nilai K yang digunakan sama seperti sebelumnya yaitu 3, 5, 7, 9.



Gambar 4. Proses Cross Validation

Selanjutnya setelah menerapkan model langkah selanjutnya menentukan nilai K untuk memperoleh akurasi tertinggi.

Tabel 12. Perbandingan Akurasi dengan Cross Validation

Nilai K	Akurasi	
	YouTube	Google News
3	88.19%	100.00%
5	85.18%	100.00%
7	83.92%	100.00%
9	83.92%	100.00%

Berdasarkan yang terlampir pada Tabel 11 dengan demikian, dapat disimpulkan bahwa tingkat akurasi tertinggi yang diperoleh adalah sebesar 88,19% pada K=3 pada pemberitaan YouTube dan 100% pada pemberitaan Google News untuk semua nilai K.

5. Confusion Matrix

Penelitian ini melakukan evaluasi menggunakan confusion matrix yang dihasilkan dari tahap pemodelan dengan menerapkan algoritma K-NN dimana pada bab sebelumnya akurasi dengan cross validation lebih tinggi dibandingkan dengan split data. Oleh karena itu, confusion matrix yang ditampilkan merupakan hasil evaluasi performa klasifikasi K-NN terhadap dataset pengujian dengan cross validation.

Tingkat akurasi model pada dataset YouTube dihitung berdasarkan proporsi prediksi yang tepat. Dengan nilai akurasi sebesar 88.19%, model mampu menghasilkan prediksi yang benar sekitar 88% dari total data yang digunakan.

accuracy: 88.19% +/- 1.90% (micro average: 88.19%)

	true positive	true negative	true hate_speech	true neutral	class precision
pred. positive	297	2	21	5	91.38%
pred. negative	0	32	3	1	88.89%
pred. hate_speech	3	3	7	0	53.85%
pred. neutral	3	5	1	15	62.50%
class recall	98.02%	76.19%	21.88%	71.43%	

Gambar 5. Confusion Matrix Youtube

Selanjutnya hasil confusion matrix Google News dapat ditinjau pada Gambar 6.

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true hate_speech	true positive	true neutral	true negative	class precision
pred. hate_speech	5	0	0	0	100.00%
pred. positive	0	231	0	0	100.00%
pred. neutral	0	0	277	0	100.00%
pred. negative	0	0	0	86	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	

Gambar 6. Confusion Matrix Google News

Perlu diketahui bahwa pemberitaan Google News yang mengandung ujaran kebencian hanya memiliki satu berita, sehingga kelas hate_speech tidak dapat

memprediksi hasil dari berita ujaran kebencian. Maka dari itu solusi yang dilakukan adalah menambah berita yang mengandung ujaran kebencian untuk menguji sistem apakah dapat bekerja secara optimal. Dari gambar 5 dapat disimpulkan bahwa confusion matrix daripada Google News memberi hasil akurasi, presisi serta recall yang baik yaitu 100%.

6. Validasi Ahli Bahasa

Pengujian ini melibatkan pakar bahasa dalam mengevaluasi akurasi dan reliabilitas model K-NN dalam mengklasifikasikan sentimen berita terkait IKN. Evaluasi dilakukan dengan membandingkan hasil klasifikasi yang dihasilkan dari InSet Lexicon untuk menguji model K-NN dengan penilaian pakar bahasa, yang dijadikan sebagai *ground truth*. Analisis ini bertujuan untuk menilai sejauh mana model mampu merepresentasikan penilaian linguistik yang akurat dan andal dalam klasifikasi sentimen berita, dengan validasi yang dilakukan oleh ahli bahasa

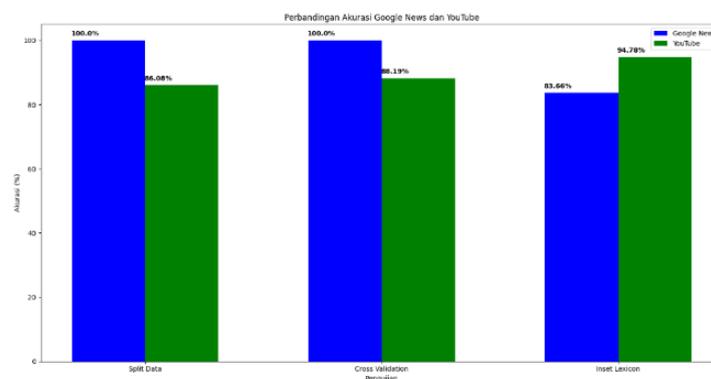
Tabel 13. Validasi Ahli Bahasa

Pengujian	Akurasi Ahli Bahasa
YouTube	83.66%
Google News	94.78%

Pada pemberitaan YouTube tingkat akurasi keseluruhan sebesar 83.66% menunjukkan bahwa model InSet Lexicon mampu mengklasifikasikan 83.66% data dengan benar sesuai kategori sentimen yang seharusnya. Selanjutnya akurasi keseluruhan pada Google News sebesar 94.8% menunjukkan bahwa model InSet Lexicon mampu mengklasifikasikan 94.8% data dengan benar sesuai kategori sentimen yang seharusnya. Validasi yang dilakukan oleh ahli bahasa menguatkan bahwa hasil klasifikasi model selaras dengan analisis linguistik, sehingga menunjukkan keandalan model dalam mendeteksi sentimen secara akurat.

7. Analisis Perbandingan

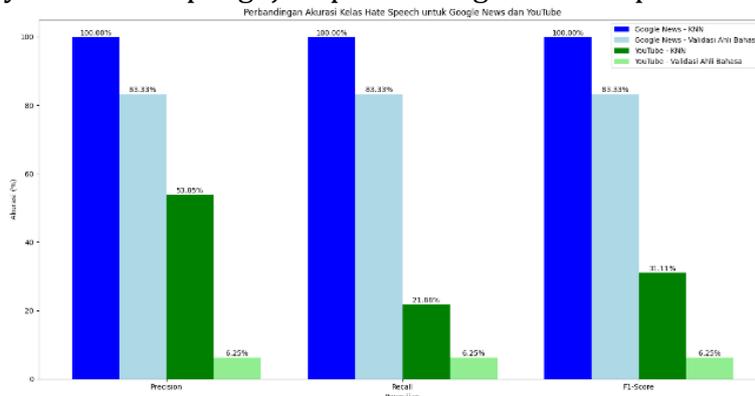
Hasil klasifikasi dari K-NN dibandingkan dengan InSet lexicon dan ahli bahasa untuk menilai seberapa baik model KNN mampu mencerminkan penilaian sentimen yang benar.



Gambar 7. Grafik Perbandingan Akurasi

Pengujian terhadap pemberitaan di Google News menunjukkan bahwa model K-NN mencapai akurasi 100%, sementara hasil validasi oleh ahli bahasa, yang berfungsi sebagai *ground truth*, mencatat akurasi 94.78%. Hal ini mengindikasikan bahwa model K-NN mampu mencerminkan penilaian ahli bahasa dengan sangat baik pada data dari Google News, sehingga dapat memahami dan mengklasifikasikan teks dari sumber ini dengan tingkat akurasi yang tinggi. Sementara itu, pada pengujian pemberitaan dari YouTube, model K-NN mencapai akurasi 88.19%, sedangkan validasi oleh ahli bahasa menunjukkan akurasi 83.66%. Perbandingan ini mengungkapkan bahwa algoritma K-NN memiliki performa yang cukup baik dalam menganalisis sentimen berita dari YouTube, meskipun terdapat sedikit perbedaan dibandingkan dengan evaluasi manual ahli bahasa.

Selanjutnya dilakukan pengujian perbandingan khusus pada kelas *Hate Speech*.



Gambar 8. Grafik Perbandingan Akurasi Kelas Hate Speech

Hasil menunjukkan bahwa algoritma K-NN menunjukkan performa yang optimal pada data Google News, dengan precision, recall, dan F1-score masing-masing mencapai 100%. Hal ini menunjukkan bahwa model mampu mengidentifikasi kelas *Hate Speech* dengan sempurna tanpa kesalahan. Sebaliknya, pada data YouTube, performa model mengalami penurunan yang signifikan, dengan precision sebesar 53.85%, recall 21.88%, dan F1-score 31.11%. Hasil ini menunjukkan bahwa model masih menghadapi kendala dalam mendeteksi Hate Speech dengan baik, sehingga banyak kasus yang tidak teridentifikasi secara akurat.

Sementara itu, validasi ahli bahasa menunjukkan hasil yang lebih rendah dibandingkan model K-NN, terutama pada data YouTube, dengan precision, recall, dan F1-score hanya mencapai 6.25%. Temuan ini dapat mengindikasikan bahwa data YouTube mengandung kompleksitas tinggi atau tingkat noise yang signifikan, sehingga baik model K-NN maupun evaluasi manual oleh ahli bahasa mengalami tantangan dalam mengklasifikasikan kelas Hate Speech secara tepat.

D. Simpulan

Penelitian ini menggunakan Total berita yang dikumpulkan terdiri dari 398 berita dari YouTube dan 599 berita dari Google News. Hasil yang diperoleh dari analisis sentimen pada pemberitaan YouTube adalah 21 berita diberi sentimen Netral, 303 berita diberi sentimen positif, 42 berita diberi sentimen negatif, sementara itu, 32 berita diberi label *hate speech*. Selanjutnya sentimen yang diperoleh pemberitaan online Google News yang dari total dataset yaitu sebanyak

277 berita diberi sentimen netral, 231 berita diberi sentimen positif, 86 berita diberi sentimen negatif, sementara itu, 1 berita diberi label *hate speech*.

Selanjutnya, akan dilakukan klasifikasi sentimen menggunakan algoritma K-NN. Pada tahap ini ini akan diuji dalam dua tahapan yaitu *split data* dan *cross validation*. Hasil akurasi tertinggi diperoleh dari pengujian *cross validation* diperoleh hasil evaluasi tertinggi pada pengujian pemberitaan YouTube dengan nilai K = 3 yaitu 88.19%. Pada hasil pemberitaan Google News menduduki hasil terbaik dimana dalam pengujian split data maupun cross validation memperoleh nilai akurasi 100% pada semua pengujian nilai K. Hal ini menunjukkan bahwa pada pemberitaan Google News memiliki akurasi yang lebih baik dibandingkan YouTube karena berita yang ditemukan di platform ini umumnya memiliki struktur yang jelas, dengan penekanan pada fakta dan laporan langsung. Susunan teks yang terorganisir serta penggunaan bahasa formal mempermudah model dalam mengidentifikasi pola pemberitaan, termasuk ujaran kebencian, dengan tingkat akurasi yang lebih tinggi.

E. Ucapan Terima Kasih

Penulis menyampaikan rasa terima kasih serta apresiasi mendalam kepada seluruh pihak yang sudah banyak memberikan bantuan selama proses penelitian ini. Dukungan, doa, bimbingan, serta kontribusi yang diberikan sangat berharga dan berperan penting dalam kelancaran serta keberhasilan penelitian ini. Tanpa bantuan dan kerja sama dari berbagai pihak, penelitian ini tidak akan dapat terselesaikan dengan baik.

F. Referensi

- [1] Tasya Najwa, "Kebijakan Pemindahan Ibu Kota Nusantara (IKN) Perspektif Pembangunan Berkelanjutan Berwawasan Lingkungan," 2024.
- [2] A. Almas and A. Dwi Hartanto, "ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA (IKN) MENGGUNAKAN METODE OVERSAMPLING SYNTHETIC MINORITY (SMOTE)," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, pp. 324–335, 2024.
- [3] A. Almas and A. Dwi Hartanto, "ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA (IKN) MENGGUNAKAN METODE OVERSAMPLING SYNTHETIC MINORITY (SMOTE)".
- [4] 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.
- [5] P. Kinerja Naïve Bayes, A. Supian, B. Tri Revaldo, N. Marhadi, and L. Efrizoni, "Acuan Supian Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara." [Online]. Available: <https://github.com/syenirashaella/Sentiment-Analysis-IKN->
- [6] S. Kartika and N. Nurhayati, "Ujaran Kebencian (Hate Speech) di Media Sosial dalam Konteks Hukum dan Perubahan Sosial (Studi Kasus pada Masyarakat Kota Medan)," *JURNAL MERCATORIA*, vol. 16, no. 1, pp. 99–106, Jun. 2023, doi: 10.31289/mercatoria.v16i1.7668.

- [7] U. Alamri, D. Ekawaty Ismail, and J. U. Puluhulawa, "Upaya Kepolisian Dalam Menanggulangi Kejatan Ujaran Kebencian (Hate Speech) Berdasarkan Surat Edaran No Se/06/X/2015 Di Gorontalo," vol. 1, no. 4, pp. 49–63, 2023, doi: 10.51903/jaksa.v1i4.1401.
- [8] H. Parulian and R. D. Putranto, "Pidana Ujaran Kebencian Melalui Media Sosial Ditinjau dalam Perspektif Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan Atas Undang Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE)," 2022.
- [9] S. Algifari Rismawan, Y. Syahidin, P. Piksi Ganesha Bandung, J. Gatot Subroto No, and K. Batununggal, "Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python," vol. 10, no. 3, pp. 167–178, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [10] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 17–23, Feb. 2021, doi: 10.29207/resti.v5i1.2700.
- [11] A. Syahril *et al.*, "Perbandingan Metode Decision Tree Dan K-Nearest Neighbor Terhadap Ulasan Pengguna Aplikasi Mypertamina Menggunakan Confusion Matrix," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 1085–1094, 2024, doi: 10.47065/josh.v5i4.5639.
- [12] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [13] J. Kalyzta, M. A. Willdan, S. Halfiani, and I. Indra, "Penerapan Analisis Sentimen Ujaran Kebencian Terhadap Vaksinasi Covid-19 Pada Tweet Berbahasa Indonesia Menggunakan Algoritme K-Nearest Neighbor," *IDEALIS: InDonEsiA journal Information System*, vol. 5, no. 2, pp. 87–97, 2022, doi: 10.36080/idealis.v5i2.2959.
- [14] S. N. S. Muslim, F. Nurdiansyah, and A. Y. Rahman, "PERBANDINGAN ALGORITMA NAIVE BAYES DAN KNN DALAM ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI CAPCUT," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5156.
- [15] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES Journal of Mathematics*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [16] J. Muliawan and E. Dazki, "SENTIMENT ANALYSIS OF INDONESIA'S CAPITAL CITY RELOCATION USING THREE ALGORITHMS: NAÏVE BAYES, KNN, AND RANDOM FOREST," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 5, pp. 1227–1236, 2023, doi: 10.52436/1.jutif.2023.4.5.347.
- [17] M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," in *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIconCIT 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 41–44. doi: 10.1109/EIconCIT50028.2021.9431845.

- [18] A. Dwiki, A. Putra, and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," vol. 8, no. 2, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [19] C. Candra, K. W. Chandra, and H. Irsyad, "Efektifitas SMOTE dalam Mengatasi Imbalanced Class Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 31–42, Jul. 2024, doi: 10.54082/jiki.132.
- [20] S. Abdullah, "Penulis Pertama: Visualisasi Data Analisa Sentimen ... 261 Visualisasi Data Analisa Sentimen RUU Omnibus law Kesehatan Menggunakan KNN dengan Software RapidMiner," vol. 8, no. 3, 2023.
- [21] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.
- [22] S. F. Huwaida, R. Kusumawati, and B. Isnaini, "Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes," *Jambura Journal of Informatics*, vol. 6, no. 1, pp. 26–39, Apr. 2024, doi: 10.37905/jji.v6i1.24718.
- [23] H. Y. Pradana, I. Slamet, and E. Zukhronah, "Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma NBC, KNN, dan SVM." [Online]. Available: <http://jurnal.umt.ac.id/index.php/senamu/index>