

SENTIMENT ANALYSIS OF INDONESIA'S CAPITAL CITY RELOCATION USING THREE ALGORITHMS: NAÏVE BAYES, KNN, AND RANDOM FOREST

Joshua Muliawan^{*1}, Erick Dazki^{*}

^{1,2,3}Business Information System Program, Faculty of Science and Technology, Universitas Pradita, Indonesia
Email: joshua.muliawan@student.pradita.ac.id, erick.dazki@pradita.ac.id

(Article received: May 28, 2022; Revision: June 7, 2022; published: June 18, 2022)

Abstract

The relocation of Indonesia's capital city from Jakarta to the IKN Nusantara on the island of Borneo has become a trending topic that triggers conversations and opinions on various social media. The pros and cons of this policy are very pronounced in various media, especially on Twitter or X platform. The purpose of this research is to conduct a public sentiment analysis of public opinion related to the relocation of Indonesia's capital city. Data is taken from tweets comments collected during a certain period from June to September 2023. This research uses a Natural Language Processing approach with data pre-processing techniques to prepare the data before applying labeling and classification algorithms. This research tests the accuracy of three algorithms used in classification, namely Naïve Bayes Classifier, K-Nearest Neighbor, and Random Forest. The results of the data classification show that positive sentiment has a value of 36.8%, neutral sentiment is at 25%, and negative sentiment related to the relocation of the capital city is 38.1%. Then an accuracy test was carried out on the Naïve Bayes Classifier Algorithm method which found an accuracy value of 65.26%, the K-Nearest Neighbor Algorithm of 58.25%, and the Random Forest Algorithm of 45.05%. This shows that the Naïve Bayes Classifier Algorithm method has better accuracy than other algorithms in predicting classification in sentiment analysis. This research also identifies the frequency of key words that often appear in each sentiment which can be valuable information for monitoring public opinion on social media.

Keywords: IKN Nusantara, K-Nearest Neighbor, Naïve Bayes, Random Forest, Sentiment Analysis, Twitter.

ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA INDONESIA MENGGUNAKAN TIGA ALGORITMA: NAÏVE BAYES, KNN, DAN RANDOM FOREST

Abstrak

Pemindahan Ibu Kota Negara Indonesia dari Jakarta menuju IKN Nusantara yang berada di Pulau Kalimantan menjadi *trending topic* yang memicu perbincangan dan pendapat masyarakat di berbagai media sosial. Pro dan kontra akan kebijakan ini sangat terasa di berbagai media, khususnya pada platform Twitter atau X. Adapun tujuan dari penelitian ini adalah untuk melakukan analisis sentimen publik terhadap opini masyarakat terkait pemindahan ibu kota negara Indonesia. Data diambil dari komentar *tweets* yang dikumpulkan selama periode tertentu selama Bulan Juni hingga September 2023. Penelitian ini menggunakan pendekatan *Natural Language Processing* dengan teknik *pre-processing data* untuk mempersiapkan data sebelum diterapkannya pelabelan dan algoritma klasifikasi. Penelitian ini menguji nilai keakuratan tiga algoritma yang digunakan dalam pengklasifikasian, yaitu *Naïve Bayes Classifier*, *K-Nearest Neighbor*, dan *Random Forest*. Hasil dari klasifikasi data menunjukkan bahwa sentimen positif memiliki nilai sebesar 36.8%, sentimen netral berada pada 25%, dan sentimen negatif terkait pemindahan ibu kota sebesar 38.1%. Kemudian dilakukan pengujian akurasi terhadap metode Algoritma *Naïve Bayes Classifier* yang didapati nilai keakuratan sebesar 65.26%, Algoritma *K-Nearest Neighbor* sebesar 58.25%, serta Algoritma *Random Forest* sebesar 45.05%. Hal ini menunjukkan bahwa metode Algoritma *Naïve Bayes Classifier* memiliki akurasi yang lebih baik daripada algoritma lainnya dalam memprediksi klasifikasi data pada analisis sentimen. Penelitian ini juga mengidentifikasi frekuensi kata-kata kunci yang sering muncul dalam setiap sentimen yang dapat menjadi informasi berharga untuk memantau opini publik pada media sosial.

Kata kunci: Analisis Sentimen, IKN Nusantara, K-Nearest Neighbor, Naïve Bayes, Random Forest, Twitter.

1. PENDAHULUAN

Ibu kota merupakan jantung atau pusat dari sebuah negara dalam menjalankan pemerintahan yang di dalamnya memiliki unsur administratif eksekutif, legislatif, dan yudikatif. Selain itu, Ibu kota juga dapat berfungsi sebagai pusat perekonomian. Ibu kota suatu negara sangat penting dalam membentuk identitas negara tersebut, tidak hanya secara simbolis namun juga secara politis, ekonomi, dan sosial-budaya [1]. Oleh sebab itu, peran ibu kota sangatlah penting pada sebuah negara. Hingga saat ini, DKI Jakarta masih merupakan Ibu Kota Negara Indonesia yang ditetapkan melalui Undang-Undang Nomor 10 Tahun 1964 yang disahkan pada 31 Agustus 1964 oleh Presiden Ir. Soekarno.

Banjir, kemacetan, kerusakan lingkungan, kepadatan populasi, kesenjangan sosial, hingga polusi udara yang baru-baru ini tengah menjadi perbincangan hangat merupakan segelintir permasalahan yang saat ini dialami oleh Jakarta sebagai Ibu Kota Negara Indonesia. Pada masa pemerintahan Presiden Ir. Soekarno, ide untuk memindahkan ibu kota negara sudah muncul. Pada saat itu, pemerintah mengantisipasi bahwa masalah kependudukan akan muncul di DKI Jakarta. Sehingga merencanakan untuk memindahkan ibukota ke Palangkaraya, Provinsi Kalimantan Tengah. Namun, rencana pemindahan ibu kota gagal karena kesulitan mendapatkan bahan bangunan dan desakan dari sejumlah duta besar yang ingin DKI Jakarta tetap sebagai ibu kota negara. [2].

Saat ini pada masa pemerintahan Ir. H. Joko Widodo, pemerintah merasa bahwa pemindahan ibu kota harus segera direalisasikan. Tentu saja, hal ini juga didukung oleh alasan keberlanjutan bidang-bidang lain seperti bidang lain seperti sosial-politik dan ekonomi [3]. Lokasi baru IKN diumumkan melalui siaran pers pada Senin, 26 Agustus 2019. Lokasinya terletak di sebagian Kabupaten Kutai Kartanegara di Kabupaten Penajam Paser Utara, Provinsi Kalimantan Timur [4]. Pembangunan IKN Nusantara dilakukan pada pertengahan Maret 2022 dan kemudian akan direlokasi secara bertahap dari tahun 2024 hingga 2045 [5]. Selain itu, pemerintah memiliki rencana untuk memindahkan PNS pusat ke ibu kota baru mulai tahun 2024. Seluruh PNS pusat yang tinggal di wilayah DKI Jakarta dan sekitarnya dipastikan akan terpengaruh oleh rencana ini. [2].

Pada tanggal 30 Mei 2023, Presiden Joko Widodo mengumumkan pemenang sayembara desain logo Pohon Hayat IKN Nusantara. Desain tersebut secara resmi terpilih menjadi identitas visual resmi dari IKN Nusantara. Periode pemilihan logo dilakukan pada 4 April – 20 Mei 2023 dengan persentase suara sebesar 26,6%. Menurut Presiden Joko Widodo, Pohon hayat ini merupakan pohon kehidupan, yang menginspirasi IKN untuk menciptakan tempat kehidupan baru, menjadi

sumber kehidupan bagi seluruh masyarakat Indonesia.

Dengan diumumkannya kebijakan pemerintah yang memindahkan ibu kota negara dari Kota Jakarta menuju Kalimantan, dapat memicu berbagai reaksi luas pada masyarakat dari reaksi positif hingga reaksi negatif atau netral dari masyarakat. Sebagai salah satu *platform* media sosial yang populer di Indonesia, Twitter atau saat ini telah berganti nama sebagai “X” digunakan oleh berbagai kalangan untuk mengalirkan komunikasi dan opini menjadi lebih beragam, salah satunya adalah kritikan dan komentar mengenai kebijakan yang dijalankan pemerintah. Penggunaannya dapat mengekspresikan apa yang dipikirkan dalam pesan waktu nyata. Pengguna Twitter bebas menyampaikan pendapat atau ekspresi mereka tentang layanan, fasilitas atau isu politik atau hal-hal yang sedang dibahas [6].

Dalam mengetahui sentimen publik terhadap pemindahan ibu kota, dilakukan analisis sentimen yang bertujuan untuk menentukan opini yang terkandung apakah cenderung memiliki sentimen positif, negatif, atau netral [7]. Dalam penelitian ini, dilakukan pendekatan *Natural Language Processing* (NLP) yang dikenal sebagai proses pengolahan data teks dimana mesin tidak bisa memahami Bahasa natural dari manusia maka perlu sebuah proses agar dapat mesin dapat memahami Bahasa manusia dengan proses yang disebut *Natural Language Processing*[8]. Sebelum dilakukan pengujian algoritma, dilakukan *pre-processing data* untuk mengolah data mejadi siap untuk digunakan.

Pre-processing data merupakan tahapan pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data, baik itu berupa karakter non-alfabet maupun kata-kata yang tidak diperlukan. Hal ini bertujuan agar data yang digunakan lebih optimal ketika digunakan pada proses pengklasifikasiannya [9]. Dalam tahap *pre-processing data*, dilakukan beberapa tahapan seperti *case folding*, *stopword removal*, *word normalizer*, *stemming*, *tokenizing*, *weighting*.

Terdapat bermacam metode algoritma pengklasifikasian yang dapat digunakan pada kategori data teks biasanya diadopsi dari *traditional machine learning*, seperti seperti Algoritma *Decision Tree*, *K-Nearest Neighbors*, *Naïve Bayes Classifier*, *Random Forest*, dan *Support Vector Machine* (SVM) [10]. Dalam penelitian ini, terdapat tiga algoritma yang dipilih dalam pengujian, yaitu *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbor* (KNN) dan *Random Forest*. Pengklasifikasi Bayesian adalah Teknik klasifikasi statistik yang digunakan untuk memprediksi probabilitas keanggotaan kelas tertentu, menghitung peluang untuk suatu hipotesis, menghitung peluang dari suatu kelas dari setiap kelompok atribut yang ada,

serta menentukan kelas mana yang paling optimal [11].

Terdapat beberapa penelitian yang merujuk pada penelitian ini. Hasil penelitian analisis sentiment pemindahan ibu kota dengan menerapkan *Naïve Bayes Classifier* melalui *text mining* didapatkan nilai akurasi sebesar 60,66% [12]. Sedangkan penelitian terhadap pengguna Twitter menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan jumlah data *tweets* sebanyak 2000 data mendapatkan hasil akurasi tertinggi sebesar 67,2% [13]. Terdapat pula penelitian mengenai analisis sentimen pemindahan ibu kota negara Indonesia dengan membandingkan tiga algoritma, yaitu *Naïve Bayes*, *Random Forest* dan *Support Vector Machine* (SVM). Dari pengujian ini, algoritma *Naïve Bayes* memiliki skor akurasi 58,52% dan algoritma *Random Forest* mendapatkan skor sebesar 71,37% [14]. Penelitian lainnya juga meneliti tentang pembahasan opini publik Twitter menggunakan metode *Naïve Bayes* dan KNN. Dengan metode *Naïve Bayes*, didapati nilai akurasi sebesar 63,10% yang berarti metode *Naïve Bayes* lebih akurat dibanding metode KNN [15].

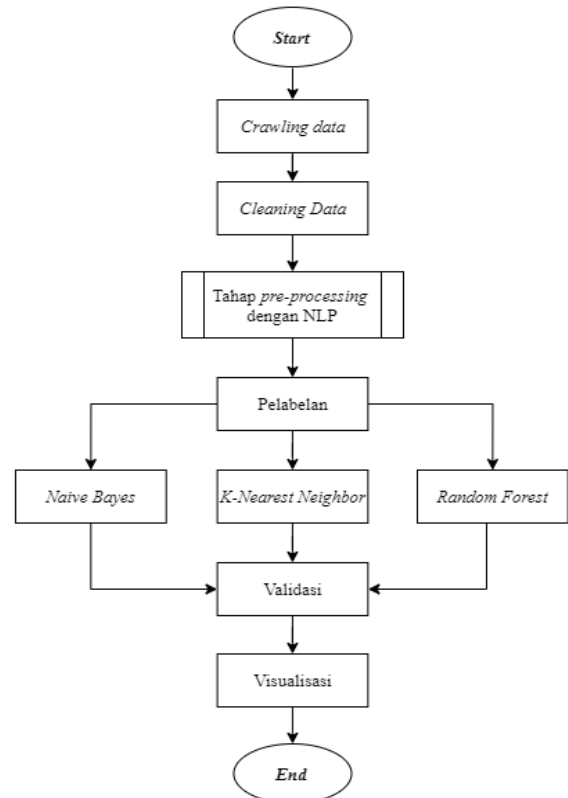
Melalui pelaksanaan penelitian ini, diharapkan dapat mengungkap pemahaman tentang respons masyarakat terhadap pemindahan Ibu Kota Negara Indonesia, apakah cenderung bersifat positif, negatif, atau netral. Selain itu, tujuan dari penelitian ini juga untuk mengidentifikasi kata-kata yang sering muncul dalam setiap sentimen positif, negatif, dan netral. Dengan perbandingan ini, diharapkan pula dapat digunakan untuk mengidentifikasi antara kedua metode algoritma, yaitu *Naïve Bayes Classifier*, *K-Nearest Neighbor*, dan *Random Forest* yang memiliki tingkat akurasi yang lebih tinggi dalam menilai sentimen topik pembicaraan, terutama terkait topik pemindahan atau relokasi ibu kota negara ke IKN.

2. METODE PENELITIAN

Dalam menjalankan penelitian ini, dilakukan tahapan-tahapan yang dijalankan meliputi pengumpulan data atau *crawling data* dari platform media sosial Twitter (X), dilanjutkan dengan *cleaning data*, tahap ketiga dilakukan tahap *pre-processing data* yang meliputi *case folding*, *stopword removal*, *word normalizer*, *stemming*, dan *tokenizing*. Kemudian dilakukan pelabelan data latih yang dilanjutkan dengan pembobotan kata (*weighting*) menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Data yang telah diproses tersebut, akan dilakukan pelabelan dan proses klasifikasi menggunakan Algoritma *Naïve Bayes*, KNN, dan *Random Forest*.

Setelah berhasil menerapkan algoritma tersebut, kemudian dilakukan tahapan validasi untuk memastikan bahwa Algoritma *Naïve Bayes*, KNN, dan *Random Forest* memiliki akurasi serta presisi

yang baik. Validasi model menggunakan metode *K-Fold Cross Validation*.



Gambar 1. Flowchart Metode Penelitian.

2.1. Identifikasi Masalah

Penelitian dilakukan didasarkan oleh permasalahan yang telah dijelaskan sebelumnya yaitu:

1. Untuk mengetahui sentimen publik terhadap pemindahan Ibu Kota Negara Indonesia. Analisis sentimen akan menghasilkan persentase seberapa besar tanggapan Masyarakat yang positif, negatif, atau netral.
2. Untuk mengidentifikasi antara ketiga metode algoritma yang lebih unggul dan akurat dalam menilai sentimen sebuah topik pembicaraan, khususnya pada topik pemindahan ibu kota negara.
3. Mengetahui kata-kata yang sering muncul pada setiap sentiment positif, negatif, dan netral.

2.2. Crawling atau pengumpulan data

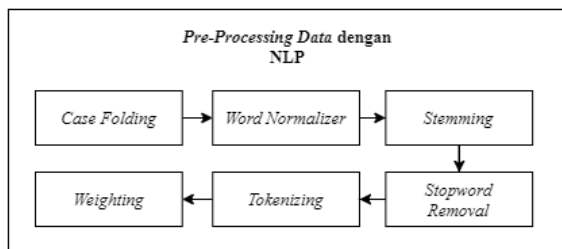
Pengumpulan data merupakan tahapan awal untuk dilakukan analisis sentimen pada sebuah topik. Dalam mengumpulkan data Twitter (X), digunakan *tools* dari Google Colab serta twitter API untuk memperoleh token yang dapat digunakan untuk menarik data *tweet*. Setelah berhasil melakukan penarikan, maka data akan disimpan ke dalam format csv.

2.3. Cleaning Data

Data mentah yang telah didapat melalui *crawling* masih terdapat banyak sekali kotoran seperti data duplikat, *mention*, *hashtag*, *emoji*, alamat *website* URL, hingga simbol-simbol yang tidak relevan untuk digunakan pada penelitian ini. Oleh karena itu, *cleaning data* dilakukan untuk memastikan bahwa data sudah benar-benar bersih dan dapat dibaca dengan baik untuk dapat dilakukan proses lebih lanjut.

2.4. Pre-Processing Data

Tahapan *pre-processing* dilakukan untuk mengolah data menjadi bahasa yang lebih standar, umum, dan dasar. Berikut gambaran tahapan *pre-processing data*:



Gambar 2. Flowchart *pre-processing data*

1. *Case Folding*: Setelah data melewati tahap sebelumnya, maka dilanjutkan dengan *case folding* yang dilakukan untuk menyeragamkan karakter pada data. Penyeragaman kata pada penelitian ini dilakukan dengan mengubah seluruh huruf menjadi huruf kecil.
2. *Word Normalizer*: Untuk memperbaiki kata-kata dalam ulasan sehingga menghasilkan kalimat yang baik dan benar sesuai dengan aturan tata bahasa Indonesia. Contoh *word normalizer* dilakukan seperti pada kata “mksh” diubah menjadi “terima kasih”.
3. *Stemming*: Proses penguraian suatu kata menjadi kata dasarnya tanpa imbuhan di awal maupun di akhir kata.
4. *Stopword Removal*: Kata-kata yang tidak memiliki makna arti yang jelas seperti imbuhan akan dihilangkan dari data yang digunakan pada penelitian.
5. *Tokenizing*: berfungsi untuk membagi teks input menjadi *array token*.
6. *Weighting*: Merupakan proses pemberian bobot pada setiap kata yang ada dalam sebuah data [24]. Pembobotan kata menggunakan metode TF-IDF (*Term Frequency - Inversed Document Frequency*).

2.5. Pelabelan Data

Setelah berhasil melakukan serangkaian proses *pre-processing*, dilakukan pelabelan data secara manual. Data latih yang telah dilakukan pelabelan akan dilatih ke dalam Algoritma *Naïve Bayes*

Classifier maupun Algoritma *Decision Tree* untuk membuat model yang nantinya akan menentukan label sentimen pada data uji. Adapun pelabelan data memiliki tiga klasifikasi, yaitu positif, negatif, atau netral.

2.6. Naïve Bayes Classifier

Naïve Bayes Classifier adalah metode klasifikasi hipotesis Bayes (*Bayes Statistics*) dan menerapkan Teknik statistic serta probabilitas yang dikembangkan oleh seseorang yang bernama Thomas Bayes untuk menentukan arah posisi suatu objek relatif terhadap posisinya di masa lalu [16]. Ini merupakan metode yang dapat dipilih untuk mengklasifikasikan sebuah teks untuk diterapkan dalam analisis sentimen. Berikut merupakan rumus pada *naïve bayes classifier* dengan persamaan [17]:

$$P(C | X) = \frac{P(X | C) P(C)}{P(x)} \quad (1)$$

2.7. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan k-buah data latih yang jaraknya paling dekat dengan objek tersebut. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data. Perhitungan jarak pada KNN dapat menggunakan persamaan [18]:

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

2.8. Random Forest

Random Forest merupakan metode pengklasifikasian yang terdiri dari beberapa pohon keputusan yang dibangun menggunakan vektor acak. Algoritma *random forest* memiliki proses *training* yang lebih cepat dibanding *decision tree* [19].

$$f_{ij} = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} n_{ij}}{\sum_{k \in \text{all nodes}} n_{ik}} \quad (3)$$

2.9. Validasi

Untuk dapat memastikan kinerja dari model Algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* bekerja secara optimal, diperlukan metode validasi yang baik dalam menilai seberapa akurat model tersebut. Salah satu metode yang digunakan adalah metode *K-Fold Cross Validation*. Metode ini digunakan untuk melakukan prediksi model dan memperkirakan keakuratan sebuah model prediktif, metode ini memecah data menjadi ‘K’ bagian set data dengan ukuran yang sama. Pelatihan dan pengujian dilakukan sebanyak ‘K’ kali [20]. Contohnya seperti validasi silang sebanyak 3 kali. Pada iterasi pertama, model dilatih pada dua bagian yang kemudian diuji pada bagian ketiga.

3. HASIL DAN PEMBAHASAN

Penelitian ini melakukan pengumpulan data *tweets* setiap bulannya dalam kurun waktu empat bulan sejak Bulan Juni hingga September 2023. Adapun tanggal pengambilan data pada setiap bulannya dilakukan secara acak sehingga didapati data yang terkumpul sebanyak 2699 data *tweets*. Kata kunci yang dipakai untuk penarikan data adalah "IKN" atau "IKN Nusantara". Untuk dapat mengetahui bahwa yang beropini adalah orang Indonesia, maka dilakukan *filter* penarikan data *tweets* yang diatur hanya yang berbahasa Indonesia.

3.1. Data Mentah

Untuk mendapatkan data mentah, dilakukan Teknik *crawling* menggunakan tools dari Google Colab dengan bahasa pemrograman Python dan *library* pandas sebagai dasar perintah untuk penarikan datanya. Data awal memiliki banyak sekali tulisan yang tidak diperlukan seperti data duplikat, *@mention*, *hashtag*, *emoji*, simbol-simbol, dan URL seperti pada Tabel 1 berikut:

Tabel 1. Data Mentah

Nomor	Teks
1	Sat Jul 29 11:39:32 +0000 2023;1685253722816086016;"Inilah Penampakan Lahan Bandara Naratetama di IKN Nusantara https://t.co/QG6YE5wBt";0;0;0;in;46693036;1685253722816086016;inilahdotcom;https://twitter.com/inilahdotcom/status/1685253722816086016
2	Tue Aug 15 23:41:41 +0000 2023;1691596049906065526;"Dukung pembangunan IKN Nusantara ðŸ™ªðŸ™ª";0;0;0;in;1456189923401224193;1691596049906065526;ZoyaKakak;https://twitter.com/ZoyaKakak/status/1691596049906065526
3	Fri Sep 08 01:38:19 +0000 2023;1699960324659806512;"@jokowi sebegitunya acara buat nanyari investor ikn buang2 duit aja.";0;0;0;in;1653666248616714247;1699070733069197770;labirin37066896;https://twitter.com/labirin37066896/status/1699960324659806512
4	Fri Sep 08 01:33:23 +0000 2023;1699959082646745103;"@AliBej0 IKN Nusantara menjadi simbol kemajuan peradaban bangsa Indonesia";0;0;0;in;2907040202;1699921825306927530;Anggita_lung;https://twitter.com/Anggita_lung/status/1699959082646745103
5	Sat Jul 29 22:19:28 +0000 2023;1685414765936717825;"Perkantoran di IKN Nusantara terapkan konsep Green Building https://t.co/GRtgyeiOAK";0;0;0;in;2943494432;1685414765936717825;Rheny_Prihaty29;https://twitter.com/Rheny_Prihaty29/status/1685414765936717825

3.2. Cleaning Data

Data mentah yang telah diperoleh melalui proses *crawling* masih mengandung banyak elemen yang perlu dibersihkan, seperti data duplikat,

@mention, *hashtag*, *emoji*, URL, dan simbol-simbol yang tidak relevan untuk keperluan penelitian ini. Oleh sebab itu dilakukan *cleaning data* menggunakan tools rapidminer. Dari total 2699 data *tweets* yang berhasil di *cleaning*, tersisa sebanyak 2182 data yang dilanjutkan ke tahap berikutnya. Hasil dari data yang telah melalui proses *cleaning* tertera pada Tabel 2 berikut:

Tabel 2. Cleaning Data

Nomor	Teks
1	Inilah Penampakan Lahan Bandara Naratetama di IKN Nusantara
2	Dukung pembangunan IKN Nusantara
3	sebegitunya acara buat nanyari investor ikn buang2 duit aja
4	IKN Nusantara menjadi simbol kemajuan peradaban bangsa
5	Perkantoran di IKN Nusantara terapkan konsep Green Building

3.3. Pre-Processing Data

Setelah melakukan *cleaning data* dilakukan tahap *pre-processing data*. Tahap ini dilakukan untuk mengolah data *tweets* menjadi bahasa yang lebih standar, umum, dasar, dan bebas dari kata-kata yang tidak memiliki arti yang jelas. Dilakukan beberapa tahapan dalam menjalankan *pre-processing data*, yakni *Case Folding*, *Word Normalizer*, *Stemming*, *Stopword Removal*, *Tokenizing*, dan *Weighting*. Berikut contoh beberapa kalimat *tweets* yang telah mengalami tahapan *word normalizer* dan *stemming* pada Tabel 3 berikut:

Tabel 3. Word Normalizer dan Stemming

Nomor	Word Normalizer	Stemming
1	ini lah penampakan lahan bandara naratetama di ikn nusantara	ini tampak lahan bandara naratetama di ikn nusantara
2	dukung pembangunan ikn nusantara	dukung bangun ikn nusantara
3	sebegitunya acara buat cari investor ikn buang-buang duit saja	sebegitunya acara buat cari investor ikn buang duit saja
4	ikn nusantara menjadi simbol kemajuan peradaban bangsa	ikn nusantara jadi simbol maju adab bangsa
5	perkantoran di ikn nusantara terapkan konsep green building	kantor di ikn nusantara terap konsep green building

3.4. Pelabelan Data

Dataset yang telah dilakukan *pre-processing* belum memiliki label klasifikasi, oleh sebab itu dilakukan pelabelan secara manual. Dataset dibagi ke dalam persentase 70% data latih dan 30% data uji. Data latih merupakan data yang dilakukan pelabelan secara manual, sedangkan data uji merupakan data yang dilakukan pengujian menggunakan Algoritma klasifikasi *naïve bayes* maupun *k-nearest neighbor*. Terdapat 1556 data *tweets* data latih dan 626 data *tweets* data uji yang

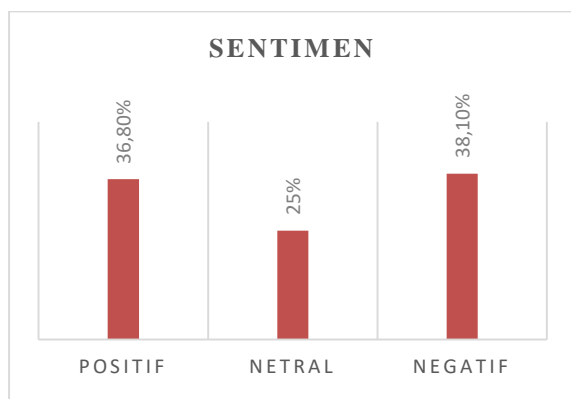
labelnya akan diprediksi oleh algoritma. Total keseluruhan dari data sebanyak 2182 data.

Dalam proses pelabelan, dilakukan tiga klasifikasi apakah komentar *tweets* memiliki makna positif yang berarti setuju atau mendukung penuh pemindahan ibu kota. Netral yang berarti komentar tidak menyatakan dukungan atau ambigu akan topik pemindahan ibu kota. Negatif yang berarti menolak, menyindir, menghina, ataupun tidak setuju akan pemindahan ibu kota Hasil pelabelan tertera pada Tabel 4 berikut:

Tabel 4. Pelabelan Data

Nomor	Data Tweets	Sentimen
1	tampak lahan bandara naratetama ikn nusantara	netral
2	dukung bangun ikn nusantara	positif
3	beginunya acara buat cari investor ikn buang duit	negatif
4	ikn nusantara jadi simbol kemajuan peradaban bangsa	positif
5	kantor ikn nusantara terap konsep green building	netral

Persentase hasil pelabelan sentimen digambarkan pada grafik gambar 3 berikut:



Gambar 3. Grafik Persentase Sentimen

Dari proses pelabelan data *tweets* yang telah dilakukan, sebanyak 804 *tweets* dikategorikan sebagai sentimen positif sebesar 36.80%, 546 *tweets* dikategorikan sebagai sentimen netral sebesar 25%, dan 832 *tweets* dikategorikan sebagai sentimen negatif sebesar 38.10% dari total keseluruhan data penelitian.

3.5. Validasi Naïve Bayes Classifier

Untuk dapat memastikan keakuratan dari model Algoritma *Naïve Bayes*, dilakukan tahap validasi. Hasil dari validasi akan mempengaruhi seberapa akurat model algoritma yang diterapkan pada pengklasifikasian. Metode yang digunakan adalah dengan *K-Fold Cross Validation*. Dilakukan pengulangan sebanyak K-2 hingga K-10 kali. Hasil iterasi ditampilkan pada Tabel 5 berikut:

Tabel 5. Hasil Akurasi Naïve Bayes Classifier

K-Fold	Accuracy
K-2	61.18%
K-3	62.14%
K-4	63.70%
K-5	63.80%
K-6	65.26%
K-7	64.39%
K-8	64.53%
K-9	65.17%
K-10	64.62%

Dengan dijalankannya validasi pada algoritma *Naïve Bayes*, didapati hasil tertinggi pada split K-6 dengan nilai keakuratan mencapai 65.26%.

3.6. Validasi K-Nearest Neighbor (KNN)

Pada penggunaan Algoritma KNN, diterapkan nilai $k=50$. Selanjutnya, model algoritma KNN juga dilakukan validasi dengan *K-Fold Cross Validation* untuk memastikan keakuratan dari model ini. Dilakukan pula pengulangan sebanyak K-2 hingga K-10 kali. Hasil iterasi ditampilkan pada Tabel 6 berikut:

Tabel 6. Hasil Akurasi K-Nearest Neighbor

K-Fold	Accuracy (50-NN)
K-2	58.16%
K-3	58.25%
K-4	57.79%
K-5	57.75%
K-6	57.34%
K-7	56.69%
K-8	57.52%
K-9	57.84%
K-10	57.93%

Dengan pelaksanaan validasi pada algoritma *K-Nearest Neighbor*, tercatat bahwa hasil paling optimal ditemukan pada split K-3 dengan tingkat akurasi sebesar 58.25%.

3.7. Validasi Random Forest

Model algoritma *Random Forest* juga dilakukan validasi dengan *K-Fold Cross Validation* untuk memastikan keakuratan dari model ini. Dilakukan pula pengulangan sebanyak K-2 hingga K-10 kali. Hasil iterasi ditampilkan pada Tabel 7 berikut:

Tabel 7. Hasil Akurasi Random Forest

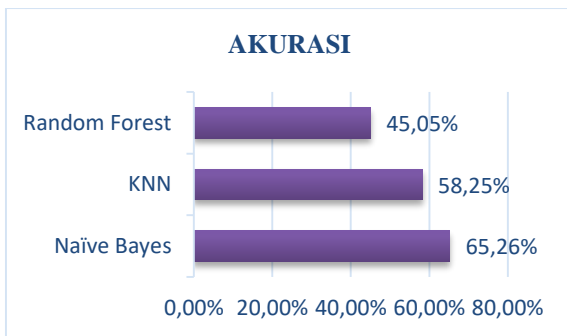
K-Fold	Accuracy
K-2	44.55%
K-3	42.99%
K-4	45.05%
K-5	43.82%
K-6	42.48%
K-7	41.84%
K-8	42.71%
K-9	41.52%
K-10	42.71%

Melalui pelaksanaan validasi pada Algoritma *Random Forest*, ditemukan bahwa hasil terbaik terjadi saat split K-4 dengan nilai keakuratan hanya sebesar 45.05%. Sebelum dilakukan validasi, pada penggunaan Algoritma *Random Forest* digunakan

konsep *prepruning* yaitu teknik yang digunakan untuk menyetop pertumbuhan pohon keputusan sebelum mencapai kedalaman maksimum yang telah ditentukan. Prinsip *prepruning* untuk mencegah pohon menjadi terlalu kompleks atau *overfitting* pada data pelatihan.

3.8. Perbandingan Akurasi

Pada pelaksanaan uji coba penggunaan ketiga metode Algoritma tersebut, didapati hasil keakuratan terbesar dicapai oleh Algoritma *Naïve Bayes Classifier* dengan nilai akurasi tertinggi sebesar 65.26%, sedangkan Algoritma KNN sebesar 58.25%, dan Algoritma *Random Forest* hanya sebesar 45.05%. Dengan hasil penelitian ini, dapat dilihat bahwa penggunaan Algoritma *Naïve Bayes Classifier* lebih baik dibanding kedua algoritma lainnya yang dapat dilihat pada grafik gambar 4:

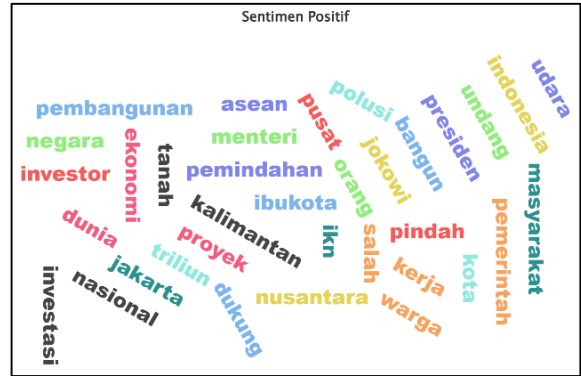


Gambar 4. Perbandingan Akurasi Algoritma

3.9. Visualisasi Data

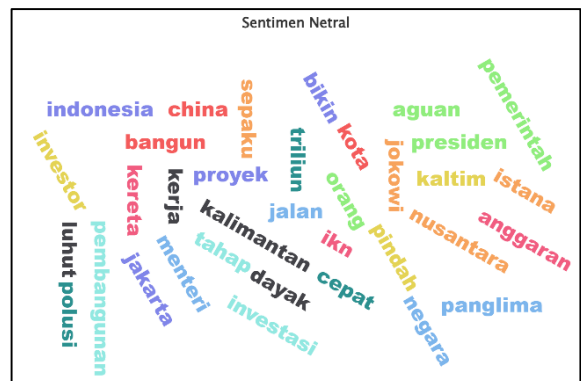
Pada setiap sentimen positif, negatif, dan netral, dilakukan visualisasi kata-kata kunci melalui media *word cloud* untuk mengetahui kata-kata apa saja yang sering muncul pada setiap sentimen tersebut sehingga didapat kesimpulan kata kunci yang sering dibahas oleh Masyarakat terkait topik pemindahan ibu kota negara. Hasil visualisasi ini dibagi menjadi tiga bagian yaitu visualisasi data positif, visualisasi data negatif, dan visualisasi data netral. Setiap visualisasi menggambarkan 35 kata yang paling sering muncul dalam masing-masing sentiment pada sentiment positif, negatif, dan netral.

Berdasarkan visualisasi sentimen positif pada gambar 5, frekuensi kata-kata yang sering muncul pada sentimen positif diantaranya adalah “ikn” sebanyak 745 kali, “pembangunan” sebanyak 153 kali, dan “dukung” sebanyak 54 kali. Penggunaan kata “dukung”, “pembangunan”, “pindah” merujuk pada harapan masyarakat agar program pemindahan ibu kota negara ke IKN dapat dilanjutkan.



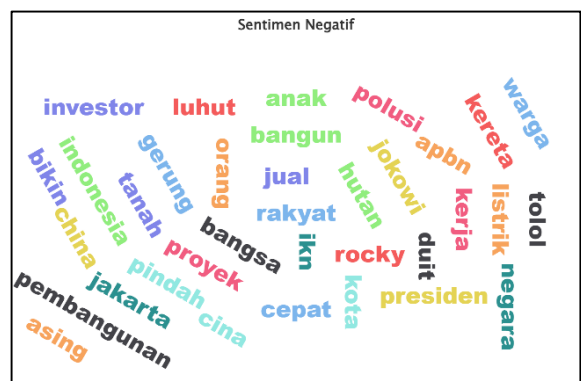
Gambar 5. Word Cloud Sentimen Positif

Pada visualisasi sentimen netral, frekuensi kata-kata yang sering muncul pada diantaranya “nusantara” sebanyak 72 kali, “negara” sebanyak 38 kali, dan “investasi” sebanyak 32 kali. Penggunaan kata-kata tersebut menunjukkan ambiguitas dan ketidakberpihakan pada keberlanjutan atau penolakan terhadap program pemindahan ibu kota negara ke IKN yang sedang dijalankan pemerintah.



Gambar 6. Word Cloud Sentimen Netral

Berdasarkan visualisasi sentimen negatif pada gambar 7, frekuensi kata-kata yang sering muncul pada diantaranya “asing” dan “china” sebanyak 73 kali, “jual” sebanyak 29 kali, dan “duit” sebanyak 27 kali. Penggunaan kata-kata tersebut menunjukkan penolakan, sindiran, ataupun ketidakpercayaan masyarakat terhadap program pemindahan ibu kota negara ke IKN yang sedang dijalankan pemerintah.



Gambar 7. Word Cloud Sentimen Negatif

4. DISKUSI

Penelitian terdahulu telah menghasilkan banyak sumber hasil yang menjadi dasar untuk penelitian ini. Untuk menjalankan penelitian ini, sangat penting untuk memiliki data-data pendukung dari penelitian yang telah dilakukan sebelumnya. Data pendukung tersebut dapat dijadikan referensi atau pertimbangan untuk mengeksplor lebih jauh mengenai topik pemindahan Ibu Kota Negara Indonesia yang berdasarkan analisis media sosial Twitter (X) melalui pendekatan *Natural Language Processing* dengan menggunakan pengujian algoritma *Naïve Bayes Classifier*, *K-Nearest Neighbor*, dan *Random Forest*.

Hasil penelitian terdahulu yang berjudul analisis sentimen terhadap pemindahan ibu kota negara dengan algoritma *Naïve Bayes Classifier*, didapatkan nilai akurasi sebesar 63,09% sedangkan KNN sebesar 69,25% dengan polaritas sentimen positif yang didapat sebesar 28% dengan 649 komentar. Sementara sentimen negatif yang didapat sebesar 30% dengan 730 komentar, sedangkan sentimen netral sebesar 42% dengan 1025 komentar [21].

Pada penelitian lainnya yang berjudul sentimen analisis terkait pemindahan ibu kota negara menggunakan metode *random forest* berdasarkan tweet warga negara Indonesia, didapati sentimen positif sebesar 55%, sentimen netral sebesar 32% dan sentimen negatif sebesar 14%. Pada penelitian tersebut juga didapati hasil pengujian keakuratan dari metode *random forest* mencapai 76% [22].

Selanjutnya, pada penelitian yang dilakukan lainnya, didapati hasil sentimen masyarakat terhadap pemindahan Ibu Kota Negara Indonesia dengan sentimen positif sebanyak 404 data sebesar 32.6% dan sentimen negatif sebanyak 832 data sebesar 67.3% dari total keseluruhan data *tweets* [23].

Selain itu, terdapat pula penelitian yang menyatakan bahwa sentimen positif Masyarakat terhadap pemindahan ibu kota negara adalah sebesar 72.5% sebanyak 396 data sedangkan pada sentimen negatif hanya sebesar 27.4% dengan 150 data [24].

Dari beberapa penelitian terdahulu di atas, menunjukkan bahwa besaran nilai sentimen masyarakat terhadap topik pemindahan ibu kota negara memiliki hasil yang beragam. Terdapat beberapa penelitian yang menemukan bahwa sentimen negatif memiliki nilai yang lebih besar dibanding sentiment positif maupun sebaliknya. Nilai yang beragam ini dapat terjadi disebabkan beberapa faktor, diantaranya perbedaan metode algoritma dan *tools* penelitian yang digunakan, kualitas data yang diambil, hingga periode waktu penelitian yang dijalankan.

5. KESIMPULAN

Penelitian ini pada awalnya mengumpulkan sebanyak 2699 tweets sebagai dataset awal. setelah

dilakukan tahap *cleaning*, didapati bahwa data yang bersih dan siap untuk dijalankan analisis sentimen menjadi sebanyak 2182 jumlah data. Hasil analisis sentimen yang diperoleh terhadap topik pemindahan ibu kota negara dengan sentimen positif masyarakat sebesar 36.8% yang berjumlah 804 data, sentimen netral sebesar 25% yang berjumlah 546 data, dan sentiment negatif dengan nilai 38.1% yang berjumlah 832 data. Kemudian dilakukan pengujian akurasi terhadap metode Algoritma *Naïve Bayes Classifier* yang didapati nilai keakuratan sebesar 65.26%, Algoritma *K-Nearest Neighbor* sebesar 58.25%, serta Algoritma *Random Forest* yang sebesar 45.05%. Hal ini menunjukkan bahwa metode algoritma *Naïve Bayes Classifier* memiliki akurasi yang lebih baik daripada algoritma lainnya dalam memprediksi klasifikasi data pada analisis sentimen. Adapun frekuensi kata-kata kunci yang sering muncul dalam sentimen positif adalah “dukung” dan “pindah”, pada sentimen netral terdapat “Nusantara” dan “investasi”, sedangkan kata “asing” dan “china” merupakan kata yang sering muncul pada sentimen negatif. Hal ini mengindikasikan bahwa kata-kata tersebut merupakan ekspresi yang umum digunakan oleh pengguna Twitter dalam beropini dalam konteks sentimen yang berbeda.

DAFTAR PUSTAKA

- [1] H. Mayer, F. Sager, D. Kaufmann, and M. Warland, *The Political Economy of Capital Cities*, 1st ed. New York : Routledge, 2018: Routledge, 2017. doi: 10.4324/9781315545837.
- [2] J. Ningrum, I. Nafiah, F. Maurist Sitorus, F. Pratita Rari, and I. Siti Maryamah, “DAMPAK PEMINDAHAN IBU KOTA NEGARA TERHADAP PENDUDUK DAN KETENAGAKERJAAN DI PROVINSI JAWA BARAT (THE IMPACT OF CAPITAL CITY DISPLACEMENT TOWARDS POPULATION AND LABOR IN WEST JAVA PROVINCE),” *Jurnal Kependudukan Indonesia* |, vol. 15, no. Desember, pp. 133–144, 2020.
- [3] F. Farida, “Indonesia’s capital city relocation: A perspective of regional planning,” *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah*, vol. 9, no. 3, pp. 221–234, Aug. 2021, doi: 10.22437/ppd.v9i3.12013.
- [4] A. Kodir, N. Hadi, I. K. Astina, D. Taryana, N. Ratnawati, and Idris, “The dynamics of community response to the development of the New Capital (IKN) of Indonesia,” in *Development, Social Change and Environmental Sustainability*, Routledge, 2021, pp. 57–61. doi: 10.1201/9781003178163-13.

- [5] D. Nugroho, "The Indonesian Journal of Politics and Policy Bentuk Ibu Kota Negara Nusantara Dalam Negara Kesatuan Republik Indonesia," *IJPP (The Indonesian Journal of Politics And Policy)* vol. 4, no. 1, 2022, [Online]. Available: <https://journal.unsika.ac.id/index.php/IJPP>
- [6] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [7] R. Y. Yanis, "Analisis Sentimen terhadap Debat Pemilihan Gubernur Jakarta Tahun 2017," *AITI*, vol. 15, no. 2, pp. 128–134, Oct. 2018, doi: 10.24246/aiti.v15i2.128-134.
- [8] N. Munasatya and S. Novianto, "Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan Multi-Layer Perceptron Natural Language Processing for President Jokowi Sentiment Analysis using Multi-Layer Perceptron." *Techno.COM*, Vol.19, No.3, Agustus 2020:237-244 [Online]. Available: <https://t.co/dV56DeVJSA>
- [9] F. A. Muttaqin and A. Mukaharil Bachtiar, "Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) IMPLEMENTASI TEKS MINING PADA APLIKASI PENGAWASAN PENGGUNAAN INTERNET ANAK 'DODO KIDS BROWSER'", 2016. [Online]. Available: <http://www.bing.com/>
- [10] J. Resti and F. Selva Jumeilah, "Terbit online pada laman web jurnal : <http://jurnal.iaii.or.id> Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," 2017. [Online]. Available: <http://jurnal.iaii.or.id>
- [11] A. Ashari Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmiah Ilmu Komputer*, vol. 2, no. 1, 2016, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [12] A. Kurniawan and S. Waluyo, "Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter," *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, pp. 35-41, 2022.
- [13] A. Deviyanto and M. D. R. Wahyudi, "PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 1, p. 1, Dec. 2018, doi: 10.14421/jiska.2018.31-01.
- [14] A. Rahman Hakim, W. Gata, A. Zevana Putri Widodo, O. Kurniawan, and A. Rama Syarif, "Analisis Perbandingan Algoritma Machine Learning Terhadap Sentimen Analisis Pemindahan Ibu Kota Negara," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, 2023, doi: 10.35870/jti.
- [15] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [16] N. Hayatin, G. I. Marthasari, and L. Nuraini, "Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization," *Jurnal Online Informatika*, vol. 5, no. 1, pp. 81–88, 2020, doi: 10.15575/join.v5i1.558.
- [17] M. Al Khadafi, Kurnia Paranitha Kartika, and Filda Febrinita, "PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN LEXICON BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 725–733, Oct. 2022, doi: 10.36040/jati.v6i2.5633.
- [18] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm," *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, vol. 4, no. 1, Mei 2022.
- [19] G. A. Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *CogITO Smart Journal*, vol. 6, no. 2, pp. 167–178, Dec. 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [20] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [21] D. Pramana, M. Afdal, and I. Permana, "Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1306–1314,

2023, doi: 10.30865/mib.v7i3.6523.

- [22] T. Cahya Herdiyani and A. U. Zailani, "Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia Sentiment Analysis Related to Transportation of Indonesian Capital City Using Random Forest Method Based On Tweet Of Indonesian Citizens," *Jurnal Teknologi Sistem Informasi*, Vol. 3, No. 2, Sept. 2022:154-165.
- [23] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," vol. 8, no. 1, pp. 147–156, Feb. 2021, doi: 10.25126/jtiik.202183944.
- [24] J. Teknika, R. K. Septiani, S. Anggraeni, and S. D. Saraswati, "Teknika 16 (02): 245-254 Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *IJCCS*, vol. x, No.x, pp. 1–5, Sept. 2022.