

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KLASIFIKASI SENTIMEN TERHADAP TOPIK PINDAH IBU KOTA NEGARA PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

JOZU DERMAWAN

NIM. 11751101941



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU**

PEKANBARU

2024

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

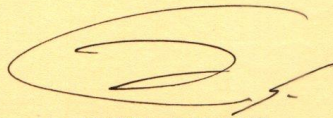
LEMBAR PERSETUJUAN**KLASIFIKASI SENTIMEN TERHADAP TOPIK PINDAH IBU
KOTA NEGARA PADA TWITTER MENGGUNAKAN
METODE NAÏVE BAYES CLASSIFIER****TUGAS AKHIR**

Oleh

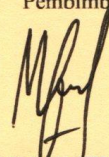
JOZU DERMAWAN**NIM. 11751101941**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 1 Juli 2024

Pembimbing I,

**YUSRA, S.T., M.T****NIP. 19840123 201503 2 001**

Pembimbing II,

**MUHAMMAD FIKRY, S.T., M.SC****NIP. 19801018 200710 1 002**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

**KLASIFIKASI SENTIMEN TERHADAP TOPIK PINDAH IBU
KOTA NEGARA PADA TWITTER MENGGUNAKAN
METODE NAÏVE BAYES CLASSIFIER**

Oleh

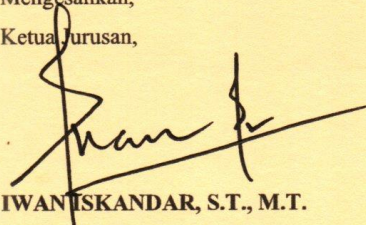
JOZU DERMAWAN


NIM. 11751101941

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

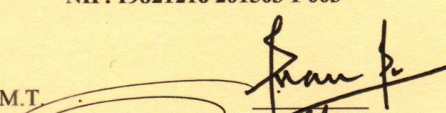
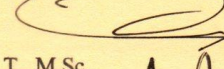
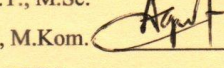
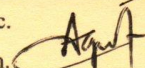
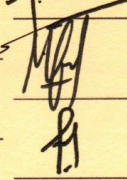
Pekanbaru, 1 Juli 2024

Mengesahkan,
Ketua Jurusan,


IWAN ISKANDAR, S.T., M.T.
NIP. 19821216 201503 1 003


Dr. HARTONO, M.Pd.
NIP. 19640301 199203 1 003

DEWAN PENGUJI

Ketua	: Iwan Iskandar, S.T., M.T.	
Pembimbing I	: Yusra, S.T., M.T.	
Pembimbing II	: Muhammad Fikry, S.T., M.Sc.	
Penguji I	: Surya Agustian, S.T., M.Kom.	
Penguji II	: Lola Oktavia, S.S.T., M.T.I.	



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis tertera dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 1 Juli 2024

Yang membuat pernyataan,

JOZU DERMAWAN

NIM. 11751101941

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah Rabbil'alamin, puji syukur kepada Allah Subhanahu wa ta'ala karena atas izin dan ridho-Nya saya dapat menyelesaikan tugas akhir saya ini. Shalawat serta salam yang selalu tercurah kepada baginda Rasulullah Muhammad Shalallaahu Alaihi Wassalaam”.

Saya Persembahkan Tugas Akhir ini Kepada kedua Orang Tua, Adik, Kakak, Keluarga Besar dan Kerabat yang selalu memberikan do'a, dukungan, kasih sayang, dan semangat tanpa henti. Terima kasih atas segala pengorbanan dan cinta kasih yang tiada batas.

Terima kasih saya ucapkan kepada Dosen Pembimbing dan Dosen Penguji yang dengan sabar memberikan bimbingan, ilmu, masukan, kritik, dan saran demi kesempurnaan tugas akhir ini. Terima kasih atas waktu dan perhatian yang telah diberikan.

Terima kasih kepada Teman-teman seperjuangan, yang telah menjadi teman berbagi, sumber inspirasi, dan dukungan moral selama proses belajar dan penulisan tugas akhir ini.

Pesan Moral, bagi siapapun yang tengah berjuang menyelesaikan studinya. Meskipun perjalanan akademik saya membutuhkan waktu yang lebih lama, saya ingin menyampaikan jangan pernah merasa putus asa, karena setiap usaha yang dilakukan dengan sungguh-sungguh pasti akan membuahkan hasil. Yakinlah bahwa setiap langkah kecil menuju impian adalah langkah besar dalam perjalanan hidup kita.

Semoga tugas akhir ini bermanfaat bagi penulis dan pembaca.

Aamiin Allahuma Aamiin.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Jozu Dermawan
NIM : 11751101941
Tempat / Tgl Lahir : Lubuk Basung / 22 April 1999
Fakultas : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi : Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara Pada Twitter Menggunakan Metode Naïve Bayes Classifier

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan Skripsi dengan judul sebagaimana tersebut diatas adalah hasil pemikiran penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebut sumbernya.
3. Oleh karena itu, Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai perundang-undangan.

Demikian Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksa pihak manapun juga.

Pekanbaru, 5 Juli 2024

Yang membuat pernyataan,



JOZU DERMAWAN
NIM. 11751101941



Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara Pada Twitter Menggunakan Metode Naïve Bayes Classifier

Jozu Dermawan^{*}, Yusra, Muhammad Fikry, Surya Agustian, Lola Oktavia

akultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11751101941@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id,

⁴surya.agustian@uin-suska.ac.id, ⁵lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: 11751101941@uin-suska.ac.id

Submitted: 10/02/2024; Accepted: 29/03/2024; Published: 31/03/2024

Abstrak—Menjelang pertengahan tahun 2019, Presiden Joko Widodo mengumumkan rencana pemindahan Ibu Kota Negara Indonesia. Hal ini menimbulkan pro dan kontra di masyarakat, yang banyak terpantau dalam berbagai media sosial. Untuk mengukur dengan cepat tingkat sentimen masyarakat terhadap kebijakan pemindahan Ibu Kota Negara (IKN) yang pembangunannya sudah berjalan tersebut, diperlukan suatu sistem klasifikasi yang memiliki performa yang baik. Penelitian ini mengusulkan klasifikasi sentimen masyarakat tentang topik tersebut menggunakan metode Naïve Bayes Classifier. Data yang digunakan pada penelitian ini berjumlah 4000 tweet yang telah diklasifikasi ke dalam dua kelas, yaitu 2000 data kelas positif dan 2000 data kelas negatif. Tujuan dari penelitian ini adalah bagaimana menerapkan metode Naïve Bayes Classifier dalam mengklasifikasikan sentimen terhadap topik pindah ibu kota negara serta mengetahui tingkat akurasi dari metode tersebut. Penerapan metode klasifikasi Naïve Bayes dengan menggunakan fitur TF-IDF untuk mengklasifikasi 10% data sebagai data testing menghasilkan akurasi sebesar 77,00%, untuk nilai precision sebesar 77,06%, recall 77,08% dan f1-score sebesar 77,00%. Berdasarkan hasil yang telah dicapai, metode Naïve Bayes Classifier baik dalam tugas klasifikasi teks, dengan tingkat akurasi yang cukup baik.

Kata Kunci: Klasifikasi Sentimen; Twitter; Naïve Bayes Classifier; TF-IDF; Pemindahan Ibu Kota Negara

Abstract—Towards the middle of 2019, President Joko Widodo announced plans to relocate Indonesia's capital city. This caused pros and cons in the community, which were widely observed in various social media. To quickly measure the level of public sentiment towards the policy of moving the National Capital City (IKN), whose construction is already underway, a classification system that has good performance is needed. This research proposes a classification of public sentiment on the topic using the Naïve Bayes Classifier method. The data used in this study amounted to 4000 tweets that have been classified into two classes, namely 2000 positive class data and 2000 negative class data. The purpose of this research is how to apply the Naïve Bayes Classifier method in classifying sentiment on the topic of moving the nation's capital and determine the accuracy level of the method. The application of the Naïve Bayes classification method using TF-IDF features to classify 10% of the data as testing data resulted in an accuracy of 77.00%, for a precision value of 77.06%, recall 77.08% and f1-score of 77.00%. Based on the results achieved, the Naïve Bayes Classifier method is good at text classification tasks, with a fairly good accuracy rate.

Keywords: Sentiment Classification; Twitter; Naïve Bayes Classifier; TF-IDF; Relocation of National Capital City

1. PENDAHULUAN

Dengan kemajuan teknologi informasi, masyarakat dapat dengan mudah mendapatkan berbagai informasi melalui internet [1]. Media sosial adalah salah satu aplikasi di internet yang paling banyak penggunanya, sehingga memungkinkan komunikasi terbuka dari masyarakat dengan berbagai latar belakang dengan pemangku kepentingan. Media sosial juga menjadi fasilitas yang efisien serta bermanfaat guna berbagi informasi kepada warga universal [2]. Berbagai macam aktifitas dilakukan masyarakat melalui media sosial, hal ini bisa menciptakan suatu informasi digital tiap harinya yang berasal dari interaksi sosial yang dibentuk lewat media sosial itu sendiri [3]. Salah satu media sosial yang banyak digunakan adalah X (Twitter). Berdasarkan laporan We Are Social, pengguna Twitter di Indonesia pada tahun 2022 mencapai 18,45 juta dari total 436 juta pengguna Twitter di seluruh dunia [4].

Melihat banyaknya pengguna media sosial Twitter, tidak salah jika Twitter sering digunakan sebagai sarana untuk melihat pendapat atau opini pengguna terkait program pemerintah atau isu yang sedang beredar di masyarakat. Twitter dapat digunakan untuk menyampaikan kritik, pendapat publik, dan saran [5]. Menjelang pertengahan tahun 2019, Presiden Joko Widodo mengumumkan rencana pemindahan Ibu Kota Negara Indonesia. Pemindahan Ibu Kota Negara (IKN) keluar pulau Jawa telah diputuskan pada tanggal 29 April 2019 oleh Presiden Joko Widodo, dan dicantumkan dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024. Tanggal 18 Januari 2022, telah disahkan Rancangan Undang-undang (RUU) tentang IKN menjadi Undang-undang (UU) oleh DPR RI dan Pemerintah. Dengan demikian, Indonesia akan mempunyai IKN yang baru menggantikan Jakarta. Hal ini menimbulkan berbagai macam opini di masyarakat. Opini yang disampaikan melalui tweet dapat dimanfaatkan untuk melihat sentimen masyarakat terhadap topik pindah Ibu Kota Negara Indonesia. Opini tersebut bisa diklasifikasikan ke dalam dua kelas yaitu positif dan negatif.

Perlu dibutuhkan sebuah proses untuk mengevaluasi pendapat masyarakat karena banyaknya informasi yang dihasilkan oleh media sosial. Salah satu alat efektif untuk menganalisis refleksi penting dari persepsi publik adalah

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Analisis sentimen [6]. Analisis sentimen digunakan untuk mengekstrak informasi penting dari data tidak terstruktur [7]. Analisis sentimen merupakan upaya guna mengungkap serta mengkategorikan komentar yang diungkapkan dalam bacaan guna memastikan apakah asumsi terhadap topik tertentu positif, negatif, ataupun netral [8]. Metode text mining dapat menyelesaikan masalah seperti pemrosesan, pengorganisasian, pengelompokan, dan analisis data yang tidak terstruktur [9].

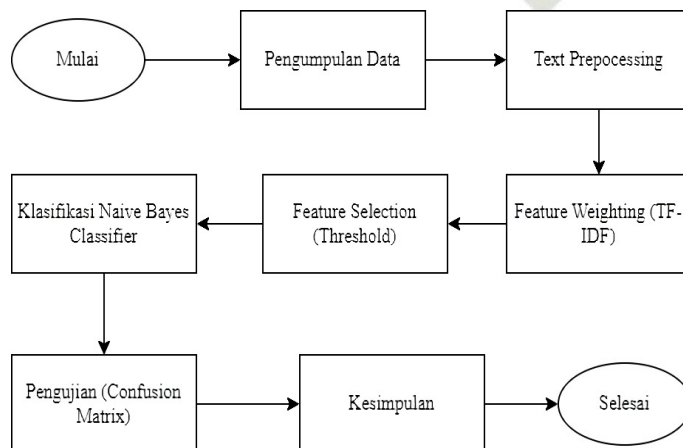
Penelitian sebelumnya tentang analisis sentimen topik pindah ibu kota negara pada twitter menggunakan metode Modified K-Nearest Neighbor pernah dilakukan oleh [10] dan mendapatkan akurasi 56,16%. Penelitian lain mengenai analisis sentimen pernah dilakukan sebelumnya oleh [11] tentang penerapan sistem plat nomor ganjil/genap pada twitter dengan metode klasifikasi Naïve Bayes dengan hasil riset tersebut dengan hasil accuracy 67,67%, presisi 71,43% serta recall 80,00%. Penelitian lain tentang analisis sentimen dan klasifikasi Tweets berbahasa Indonesia terhadap transportasi umum MRT Jakarta menggunakan Naïve Bayes Classifier yang dilakukan oleh [12] mendapatkan hasil akurasi sebesar 95,88%. Selanjutnya penelitian klasifikasi sentimen sara, provokasi, dan radikal pada postingan media sosial menggunakan algoritma Naïve Bayes Multinomial Text yang dilakukan oleh [13] mampu mengklasifikasikan sentimen dengan sangat akurat dengan akurasi 99,62%. Penelitian lain tentang implementasi data mining dengan algoritma Naïve Bayes untuk klasifikasi kelayakan penerima bantuan sembako yang dilakukan oleh [14] menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%. Penelitian lain tentang klasifikasi sentimen pergelaran MOTOGP di Indonesia menggunakan algoritma Correlated Naïve Bayes Classifier menghasilkan akurasi sebesar 82% [15].

Metode klasifikasi Naïve Bayes menggunakan algoritma yang sederhana tetapi sangat cepat dan akurat [16]. Hal tersebut dibuktikan dalam penelitian yang dilakukan oleh [17] yang membandingkan performa algoritma Naïve Bayes dan KNN dalam mengklasifikasikan sentimen masyarakat tentang kenaikan harga tiket pesawat pada Twitter, mendapatkan hasil bahwa algoritma Naïve Bayes memiliki akurasi yang lebih baik dengan nilai sebesar 90,70% dan algoritma KNN mendapatkan akurasi sebesar 62,79%. Pada penelitian lain yang dilakukan oleh [18] tentang implementasi orange data mining untuk klasifikasi kelulusan mahasiswa dengan model K-Nearest Neighbor, Decision Tree serta Naïve Bayes dan mendapatkan hasil bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang lebih tinggi dengan 89% dibandingkan dengan algoritma K-NN dengan akurasi 77% dan Decision Tree sebesar 74%. Algoritma Naïve Bayes memerlukan waktu yang lebih sedikit jika dibandingkan dengan algoritma Support Vector Machine (SVM) [19]. Penelitian lain [20] yang membandingkan performa dari algoritma Naïve Bayes Classifier (NBC) dan SVM mendapatkan hasil bahwa NBC lebih baik dengan akurasi 81,7%.

Penelitian dilakukan untuk menerapkan metode klasifikasi Naïve Bayes untuk mengklasifikasikan sentimen di masyarakat tentang topik pemindahan Ibu Kota Negara Indonesia. Data yang akan digunakan sebanyak 4000 tweet berbahasa Indonesia yang didapatkan dari penelitian sebelumnya yang dilakukan oleh A.D. Dian Kurniawan [10] dengan label positif dan negatif. Data yang diambil adalah tweet yang mengandung hastag #pindahibukota, #ibukotabaru, #pindahibukotauntuksiapa, #bukotapindah dan #indonesiaibukotabaru. Penelitian ini ingin menguji apakah metode klasifikasi Naïve Bayes lebih baik dari Modified K-Nearest Neighbor yang digunakan di dalam [10]. Tujuan dari penelitian ini adalah bagaimana menerapkan metode Naïve Bayes Classifier dalam mengklasifikasikan sentimen terhadap topik pindah ibu kota negara serta mengetahui tingkat akurasi dari metode tersebut.

2. METODOLOGI PENELITIAN

Tahapan pada penelitian ini mengikuti alur seperti pada Gambar 1, yang dimulai dari analisa data, selanjutnya masuk pada tahap text preprocessing, kemudian melakukan feature weighting menggunakan TF-IDF, feature selection menggunakan threshold, kemudian proses klasifikasi menggunakan metode Naïve Bayes Classifier, selanjutnya dilakukan pengujian menggunakan confusion matrix dan yang terakhir kesimpulan.



Gambar 1. Tahapan Penelitian

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



1. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dan melakukan pengecekan pada data apakah terdapat data yang kosong (null) dan memastikan data sesuai dengan kebutuhan.

2 Text Preprocessing

lanjutan adalah tahapan text preprocessing, terdapat beberapa proses di dalam text preprocessing yang bertujuan untuk mengubah data menjadi format yang lebih baik, sehingga data tersebut dapat diolah menjadi data yang berkualitas [21]. Text preprocessing dimulai dari:

1. Case Folding

Case folding ialah mengubah semua huruf dalam dokumen menjadi huruf kecil; hanya huruf dari alfabet a sampai z yang diizinkan, dan semua karakter lainnya dihilangkan [22].

2. Text Cleaning

Text cleaning data merupakan proses yang bertujuan untuk menghapus noise pada data [23]. Karakter dapat diidentifikasi melalui karakter-karakter tertentu seperti tanda baca, tanda pagar (#), tanda tanya (?), tanda seru (!), emoji, tautan situs web, nama pengguna (@), dan elemen-elemen lainnya.

3. Tokenizing

Tokenizing ialah proses untuk memecah kalimat menjadi beberapa kata.

4. Text Normalization

Text normalization ialah proses pengembalian perkata yang tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia. Normalization sangat penting untuk menyiapkan data teks untuk tugas pemrosesan bahasa alami (NLP) [24].

5. Negation handling

Negation handling ialah sesuatu proses pergantian kata negasi jadi kata yang sesungguhnya semacam kata "belum", "tidak", "tak", "jangan", dan "bukan".

6. Stopword Removal

Stopword removal ialah proses penghapusan kata yang dikira tidak bermanfaat.

7. Stemming

Stemming ialah proses mengurangi kata-kata ke bentuk aslinya. Ini dilakukan untuk membuat analisis teks lebih mudah, seperti untuk menganalisis sentimen, mengklasifikasikan teks, dan mencari Informasi

2.3 Feature Weighting (TF-IDF)

Feature weighting atau pembobotan fitur menggunakan metode pembobotan kata TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengukur relevansi kata-kata dalam dokumen. Kata-kata yang sering muncul dan kata-kata yang jarang muncul diberi nilai yang lebih tinggi daripada kata-kata yang hanya muncul dalam beberapa dokumen [25]. Pembobotan TF-IDF terdiri dari dua elemen: frekuensi kata muncul dalam suatu dokumen dan kebalikan dari frekuensi kata muncul dalam dokumen secara keseluruhan [26].

2.4 Feature Selection (Threshold)

Feature selection atau seleksi fitur merupakan tahapan untuk mengurangi dimensi data tekstual sehingga hasil text mining berkualitas lebih baik [27]. Fitur yang akan dihapus adalah semua fitur dengan varian lebih rendah dari nilai ambang batas yang ditentukan.

2.5 Klasifikasi Naïve Bayes Classifier

Langkah berikutnya melibatkan penggunaan Naïve Bayes Classifier, suatu teknik klasifikasi yang berakar pada metode probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris, Thomas Bayes. Metode ini mendasarkan prediksi peluang di masa depan pada pengalaman masa lalu, sehingga sering disebut sebagai Teorema Bayes [28] merupakan algoritma yang digunakan untuk mengklasifikasikan data uji ke dalam kategori yang paling sesuai. Metode klasifikasi Naïve Bayes menggunakan algoritma yang sederhana tetapi sangat cepat dan akurat [16]. Pada tahap ini data dibagi menjadi dua bagian, yaitu 90% data latih (3600 data) dan 10% data uji (400 data).

2.6 Pengujian (Confusion Matrix)

Pada tahap ini dilakukan pengujian menggunakan Confusion Matrix. Confusion matrix merupakan alat yang bermanfaat untuk mengevaluasi sejauh mana klasifikasi yang telah dilakukan mampu mengidentifikasi elemen-elemen dari kelas yang berbeda. [29].

Tabel 1. Confusion Matrix

	Prediksi	
	Positif	Negatif
Aktual	TP	FN
	FP	TN

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.2.1 Case Folding

Case folding ialah proses pengubahan huruf kapital menjadi huruf kecil. Hasil dari proses case folding dapat dilihat pada Tabel 2 berikut.

Tabel 2. Case Folding

Sebelum Case Folding	Setelah Case Folding
Jika pemindahan ibu kota negara terealisasi, Presiden Jokowi tidak akan pernah dilupakan rakyat Indonesia di masa datang. #IbuKotaPindah	jika pemindahan ibu kota negara terealisasi, presiden jokowi tidak akan pernah dilupakan rakyat indonesia di masa datang. #ibukotapindah
http://m.jpnn.com/amp/news/seperti-bung-karnopak-jokowi-bakal-dikenang-rakyat-indonesiaâ€¦	http://m.jpnn.com/amp/news/seperti-bung-karno-pak-jokowi-bakal-dikenang-rakyat-indonesiaâ€¦
#IbuKotaPindah"	#ibukotapindah

3.2 Text Preprocessing

Text preprocessing teks adalah tahapan yang bertujuan untuk menyesuaikan teks dengan kebutuhan pembentukan fitur klasifikasi. Beberapa tahap yang umum dilakukan adalah case folding, text cleaning, tokenizing, text normalization, negation handling, stopword removal, dan stemming. Di dalam penelitian ini, diselidiki aspek-aspek text preprocessing dalam pembentukan fitur apakah mempengaruhi hasil klasifikasi.

The screenshot shows a spreadsheet with the following columns: 'id', 'tweet', and 'sentimen'. The data rows contain various tweets related to the relocation of the capital city of Indonesia, such as 'Indonesia menuju negara maju', 'Yuk, Beli Tanah di Kaltim.', and 'Bagus lah agar ibukota nya lebih aman dan tenang kata fz.'. The 'sentimen' column contains values like 'Positif', 'Negatif', and 'Netral'.

Gambar 2. Data Penelitian

3. Pengumpulan Data

Data yang digunakan pada penelitian ini berjumlah 4000 data, yang terdiri dari dua kelas yaitu 2000 data kelas positif dan 2000 data kelas negatif yang disimpan dalam format .csv, bisa dilihat pada gambar 2. Pada bagian ini juga dilakukan pengecekan pada data apakah terdapat data yang kosong (null) dan memastikan data sesuai dengan kebutuhan. Data yang digunakan dapat dilihat pada gambar 2:

3. HASIL DAN PEMBAHASAN

3.2.1 Case Folding

Case folding ialah proses pengubahan huruf kapital menjadi huruf kecil. Hasil dari proses case folding dapat dilihat pada Tabel 2 berikut.

Tabel 2. Case Folding

Sebelum Case Folding	Setelah Case Folding
Jika pemindahan ibu kota negara terealisasi, Presiden Jokowi tidak akan pernah dilupakan rakyat Indonesia di masa datang. #IbuKotaPindah	jika pemindahan ibu kota negara terealisasi, presiden jokowi tidak akan pernah dilupakan rakyat indonesia di masa datang. #ibukotapindah
http://m.jpnn.com/amp/news/seperti-bung-karnopak-jokowi-bakal-dikenang-rakyat-indonesiaâ€¦	http://m.jpnn.com/amp/news/seperti-bung-karno-pak-jokowi-bakal-dikenang-rakyat-indonesiaâ€¦
#IbuKotaPindah"	#ibukotapindah



3.2.2 Cleaning

Cleaning data merupakan proses yang bertujuan untuk menghapus “noise” pada data teks. Beberapa yang dihapus dalam proses cleaning ini adalah karakter-karakter khusus seperti tanda baca, tanda pagar (#), tanda tanya (?), tanda seru (!), dan emoji. Selain itu, tautan situs web (URL), nama pengguna (@), dan karakter yang bukan teks seperti tabet juga dihapus dari teks. Contoh hasil dari proses cleaning dapat dilihat pada Tabel 3.

Tabel 3. Cleaning

Sebelum Cleaning	Setelah Cleaning
jika pemindahan ibu kota negara terealisasi, presiden jokowi tidak akan pernah dilupakan rakyat indonesia di masa datang. #ibukotapindah http://m.jpnn.com/amp/news/seperti-bung-karno-pak-jokowi-bakal-dikenang-rakyat-indonesiaâ€¦ #ibukotapindah	jika pemindahan ibu kota negara terealisasi presiden jokowi tidak akan pernah dilupakan rakyat indonesia di masa datang

3.2.3 Tokenizing

Tokenizing ialah proses untuk memecah kalimat menjadi token berbentuk kata. Contoh hasil dari proses tokenizing dapat dilihat pada Tabel 4.

Tabel 4. Tokenizing

Sebelum Tokenizing	Setelah Tokenizing
jika pemindahan ibu kota negara terealisasi presiden jokowi tidak akan pernah dilupakan rakyat indonesia di masa datang	['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'tidak', 'akan', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']

3.2.4 Normalization

Normalization ialah proses pengembalian kata yang tidak baku atau typo ke dalam bahasa baku dalam kamus bahasa Indonesia. Hasil dari proses normalization bisa dilihat pada tabel 5.

Tabel 5. Normalization

Sebelum Normalization	Setelah Normalization
['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'tidak', 'akan', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']	['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'tidak', 'akan', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']

3.2.5 Negation Handling

Negation handling ialah sesuatu proses mengubah kata yang mengandung negasi menjadi lawan katanya. Kata-kata negasi “belum”, “tidak”, “tak”, “jangan”, dan “bukan” yang mengikuti suatu kata tertentu (misal kata benda, kata sifat, kata kerja, kata keterangan) akan diubah keduanya menjadi lawan kata dari kata yang diikuti tadi. Contoh hasil dari proses negation handling bisa di lihat pada Tabel 6.

Tabel 6. Negation Handling

Sebelum Negation Handling	Setelah Negation Handling
['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'tidak', 'akan', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']	['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'menolak', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']

3.2.6 Stopword Removal

Stopword removal ialah proses penghapusan kata di dalam teks yang terdapat di dalam stoplist. Kata-kata yang digolongkan ke dalam stopword pada umumnya adalah kata-kata yang terlalu sering muncul di dalam suatu korpus dokumen yang besar. Untuk penelitian ini, digunakan daftar stoplist yang sudah umum untuk bahasa Indonesia, ditambah sejumlah kata yang memiliki frekuensi tertinggi di dalam dataset sentimen ini, seperti kata pindah, ibu, kota, dan negara. Contoh hasil dari proses stopword removal bisa dilihat pada Tabel 7.

Tabel 7. Stopword Removal

Sebelum Stopword Removal	Setelah Stopword Removal
['jika', 'pemindahan', 'ibu', 'kota', 'negara', 'terrealisasi', 'presiden', 'jokowi', 'menolak', 'pernah', 'dilupakan', 'rakyat', 'indonesia', 'di', 'masa', 'datang']	['terrealisasi', 'presiden', 'jokowi', 'menolak', 'dilupakan', 'rakyat', 'indonesia']

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.7 Stemming

Stemming adalah proses mengembalikan kata-kata ke bentuk dasarnya. Penelitian ini menggunakan stemming bahasa Indonesia dari library sastrawi. Contoh hasil dari proses stemming bisa dilihat pada Tabel 8.

Tabel 8. Stemming

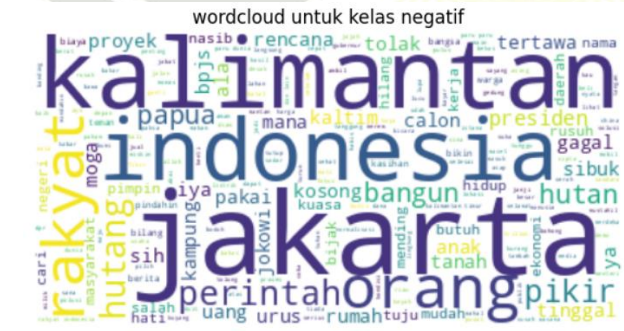
Sebelum Stemming	Setelah Stemming
'teralisasi', 'presiden', 'jokowi', 'menolak', 'dilupakan', 'rakyat', 'indonesia']	realisasi presiden jokowi tolak lupa rakyat indonesia

2.7 Word Cloud

Setelah seluruh data tweet mengalami text preprocessing, profil kata-kata di dalam dataset dapat digambarkan dalam bentuk word cloud untuk kelas positif dan kelas negatif, seperti terlihat pada Gambar 3 dan Gambar 4. Dari kedua gambar tersebut, kita dapat melihat kecenderungan kata-kata yang banyak terdapat pada masing-masing kelas, dengan melihat kepada ukuran hurufnya. Pada bagian ini juga sudah dilakukan percobaan untuk menghilangkan kata dengan frekuensi terbanyak seperti jakarta, kalimantan dan indonesia, tetapi hasil akurasi yang dihasilkan menjadi lebih rendah dibandingkan tidak menghilangkan kata-kata tersebut.



Gambar 3. Word Cloud kelas positif



Gambar 4. Word Cloud kelas negatif

3.3 Feature Weighting (TF-IDF)

Proses selanjutnya adalah pembentukan vektor bag of words dengan pembobotan kata menggunakan TF-IDF. Untuk proses pembobotan kata tersebut, digunakan library tfidf vectorizer dari sklearn [30].

3.4 Feature Selection

Feature selection atau seleksi fitur merupakan tahapan untuk mengurangi dimensi vektor input pada machine learning, agar proses komputasi lebih efisien dan hasil klasifikasi lebih baik, karena membuang fitur-fitur yang kurang signifikan dalam klasifikasi. Fitur yang menjadi objek komputasi dari metode Naïve Bayes dalam penelitian ini adalah daftar kata-kata yang terbentuk dari proses sebelumnya (tfidf tokenizer), dengan nilai berupa bobot TF-IDF. Fitur yang akan dipilih adalah semua fitur dengan nilai bobot lebih tinggi dari nilai ambang batas (threshold) yang ditentukan.

Pada penelitian ini, feature selection menggunakan library modul variance threshold dari sklearn [30]. Sebelum dilakukan feature selection, jumlah fitur berjumlah 26.514 token kata. Kemudian dilakukan beberapa percobaan untuk menentukan nilai terbaik dari threshold, dengan memilih nilai threshold dari 0,0001 sampai dengan 0,007. Nilai threshold terbaik yang diperoleh untuk penelitian ini adalah 0,0001. Setelah proses feature selection dengan threshold 0,0001 jumlah fitur menjadi 22.929 token.



Klasifikasi Naïve Bayes Classifier

Setelah melakukan proses feature selection, tahap berikutnya adalah klasifikasi menggunakan algoritma Naïve Bayes Classifier. Data dibagi menjadi dua bagian, yaitu 90% data latih (3600 data) dan 10% data uji (400 data).

3.5.1 Experiment Setup

Experiment setup dirancang untuk menemukan model yang paling optimal untuk Naïve Bayes Classifier, yaitu model yang menghasilkan performa paling baik, ditinjau dari F1-score hasil prediksinya. Pada tahap ini dilakukan beberapa experiment untuk menyelidiki pengaruh penerapan beberapa langkah text preprocessing, yaitu stemming, normalization, negation handling, dan stopword removal sebagaimana yang dilakukan oleh [31]. Hal ini dilakukan, karena keempat proses tersebut sangat memakan resource komputasi, karena prosesnya yang berulang-ulang memeriksa setiap kata pada teks, apakah terdapat di dalam kamus kata dari masing-masing proses. Komposisi eksperimen untuk menemukan model terbaik dari tahap ini bisa dilihat pada Tabel 9.

Tabel 9. Variasi eksperimen dalam menguji pengaruh aspek text preprocessing dalam klasifikasi

ID Experiment	Stemming	Stopword	Normalization	Negation Handling
1	✓	x	x	x
2	✓	✓	x	x
3	✓	✓	✓	x
4	✓	✓	✓	✓
5	x	✓	x	x
6	✓	x	✓	x
7	x	x	x	✓
8	x	✓	✓	x
9	x	✓	✓	✓
10	✓	x	✓	✓
11	✓	✓	x	✓
12	x	x	✓	✓
13	✓	x	✓	x
14	x	x	x	x
15	x	x	✓	x
16	✓	x	x	✓

Pada Tabel 9 di atas, tanda centang menyatakan tahapan tersebut diterapkan, dan tanda silang menyatakan tahapan tersebut tidak diterapkan pada proses text preprocessing, setelah itu baru dilakukan proses vectorizer.

3.5.2 Proses Training NB

Tahap menentukan kelas dari metode Naïve Bayes, dimulai dari proses training dengan melakukan proses fitting pada dataset training, menggunakan library multinomialNB dari sklearn [30]. Multinomial NB adalah pendekatan yang paling cocok dengan kasus klasifikasi teks dengan banyak fitur. Setelah proses training, terbentuk model NB yang berisikan parameter-parameter probabilitas untuk setiap fitur pada setiap kelas klasifikasi. Selanjutnya model ini dipakai untuk melakukan prediksi terhadap data testing. Beberapa contoh penerapan pada prediksi data teks tweet yang sudah mengalami proses seleksi fitur dapat dilihat pada Gambar 5 di bawah ini.

```

Teks: dampak negatif bijak presiden joko widodo kaltim narkoba masuk
Kelas Asli: negatif
Kelas Prediksi: negatif

Teks: wajar tolak nyaman jakarta
Kelas Asli: negatif
Kelas Prediksi: negatif

Teks: orang jakarta tahap uji coba orang kampung
Kelas Asli: positif
Kelas Prediksi: negatif
    
```

Gambar 5. Hasil Prediksi Kelas

3.5.3 Pencarian model NB optimal

Pencarian model optimal dilakukan dengan skema Experiment Setup yang telah dijelaskan pada Tabel 9. Pada tahapan ini juga dilakukan penerapan threshold dan pengujian tanpa threshold. Nilai threshold yang diterapkan adalah 0,0001, sesuai dengan hasil rekomendasi dari proses seleksi fitur sebelumnya, yaitu bahwa semakin besar nilai threshold yang digunakan, maka nilai akurasi dan f1-score akan semakin kecil.

Data hasil eksperimen disajikan pada Tabel 10 menampilkan nilai threshold 0,0001 dan tanpa threshold, dari hasil pengujian setiap variasi kombinasi tahap text preprocessing. Proses training (fit) dilakukan terhadap data training yang telah mengalami proses pengolahan teks sesuai dengan skenario pengujian. Kemudian akurasi dan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



score-nya dicatat ke dalam Tabel 10. Prediksi terhadap data validasi, yaitu data yang tidak terlihat saat proses training dilakukan, dan hasilnya dicatat di dalam Tabel 10 pada kolom Data Validasi. Hasil performa terbaik dari model naive bayes classifier dari skema pengujian di atas, akan dipilih sebagai model final dalam penelitian ini.

Dari pengujian ini, dipilih skema model dari ID experiment 14, yaitu proses stemming, stopword removal, negation handling dan normalisasi teks tidak dilakukan. Hanya dilakukan pemrosesan teks dasar, yang meliputi text cleaning, case folding, tokenization dan menghapus karakter-karakter khusus, angka, mention, dan URL pada model yang paling optimal. Di samping itu, penerapan threshold 0.0001 pada data fitur (token) juga diterapkan karena menghasilkan performa yang lebih baik daripada menggunakan semua fiturnya. Akurasi dan F1-score yang diperoleh pada data validasi adalah berturut-turut sebesar 78,61% dan 78,58%.

Tabel 10. Evaluation Score dari eksperimen pencarian model optimal

id experiment	Threshold	Data Training		Data Validasi	
		Accuracy	f1 score	Accuracy	f1 score
1	0	89.10%	89.10%	75.00%	75.00%
	0.0001	85.77%	85.77%	75.28%	75.27%
2	0	90.49%	90.49%	72.78%	72.69%
	0.0001	86.54%	86.54%	73.33%	73.32%
3	0	88.80%	88.79%	73.06%	72.92%
	0.0001	86.14%	86.14%	74.72%	74.68%
4	0	88.98%	88.98%	74.72%	74.67%
	0.0001	86.39%	86.39%	75.56%	75.54%
5	0	91.79%	91.79%	77.50%	77.33%
	0.0001	87.04%	87.04%	77.78%	77.75%
6	0	88.43%	88.43%	78.61%	78.57%
	0.0001	86.05%	86.04%	77.78%	77.77%
7	0	91.48%	91.48%	78.33%	78.25%
	0.0001	86.51%	86.51%	78.33%	78.31%
8	0	91.11%	91.11%	77.22%	76.97%
	0.0001	87.19%	87.19%	78.61%	78.47%
9	0	91.05%	91.05%	78.06%	77.87%
	0.0001	87.93%	87.93%	78.61%	78.49%
10	0	88.46%	88.46%	78.33%	78.32%
	0.0001	86.02%	86.01%	77.78%	77.78%
11	0	90.34%	90.34%	73.33%	73.27%
	0.0001	86.60%	86.60%	74.44%	74.44%
12	0	90.49%	90.49%	78.61%	78.54%
	0.0001	86.88%	86.88%	77.50%	77.47%
13	0	88.43%	88.43%	78.61%	78.57%
	0.0001	86.05%	86.04%	77.78%	77.77%
14	0	91.42%	91.42%	78.33%	78.25%
	0.0001	86.51%	86.51%	78.61%	78.58%
15	0	90.31%	90.31%	77.50%	77.45%
	0.0001	86.30%	86.29%	78.33%	78.32%
16	0	89.01%	89.01%	76.39%	76.35%
	0.0001	85.77%	85.77%	76.67%	76.67%

*Threshold 0 menyatakan tidak menggunakan threshold

3.5.4 Pengujian Final terhadap Data Testing

Setelah diperoleh model NB yang paling optimal, yaitu model pada ID Experiment 14, maka dilakukan pengujian terhadap data testing. Hasil prediksi kelas sentimen pada data testing dituliskan pada Tabel 11 berikut. Tabel ini juga memperlihatkan hasil yang diperoleh pada penelitian sebelumnya oleh A.D.Dian Kurniawan [10] dengan algoritma Modified K-Nearest Neighbor (M-KNN), yang mendapatkan akurasi 56,16%, precision 100%, dan recall 56,16%. Perbandingan hasil antara penelitian ini dan penelitian sebelumnya bisa dilihat pada tabel 11.

Tabel 11. Perbandingan Penelitian

Metode	Text Preprocessing	Akurasi	F1-score
Modified K-NN	(1) Basic,	56,16%	-
	(2) Normalization		
	(3) Negation Handling		
	(4) Stopword Removal		
	(5) Stemming		

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Metode	Text Preprocessing	Akurasi	F1-score
Penelitian ini (NB)	Basic	77,00%	77,00%

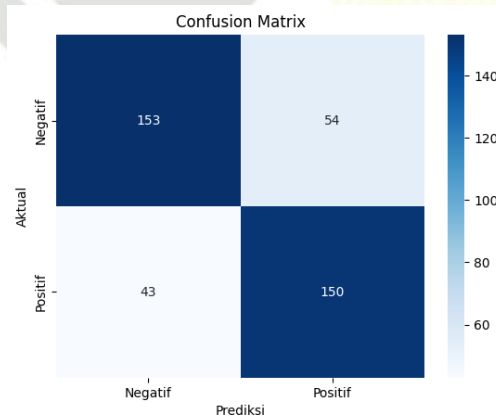
Dari tabel 11 dapat terlihat bahwa penelitian ini dengan hanya menggunakan text preprocessing dasar (basic) dapat menghasilkan akurasi yang jauh lebih baik dari pada penelitian sebelumnya yang menerapkan semua proses text preprocessing, pada pengujian data testing 10%. Peningkatan akurasi yang dapat dicapai dari metode NB melebihi 20% dibandingkan metode modified K-NN yang digunakan dalam [10]. Sedangkan untuk F1-score [10] tidak melaporkannya.

5.5 Confusion Matrix

Pemampuan model NB mengklasifikasi kedua kelas (positif dan negatif), dapat diamati dari grafik confusion matrix pada Gambar 6. Dari gambar tersebut, dapat dijelaskan hasil evaluasi sebagai berikut :

1. True Positive (TP) 150, berarti ada 150 data tweet yang diklasifikasikan dengan benar oleh model sebagai kelas positif.
2. True Negative (TN) 153, berarti ada 153 data tweet, yang diklasifikasikan dengan benar oleh model sebagai kelas negatif.
3. False Positive (FP) 54, berarti ada 54 data tweet yang sebenarnya adalah kelas negatif, tetapi keliru diklasifikasikan oleh model sebagai kelas positif.
4. False Negatif (FN) 43, berarti ada 43 data tweet yang sebenarnya adalah kelas positif, tetapi keliru diklasifikasikan oleh model sebagai kelas negatif.

Berdasarkan hasil pada confusion matrix tersebut, metode Naïve Bayes Classifier dalam pengujian terhadap data testing menghasilkan tingkat akurasi sebesar 77,00% , dengan precision sebesar 77,06%, recall 77,08%, dan f1-score sebesar 77,00%. Hasil ini sudah cukup baik, namun masih perlu ditingkatkan lagi karena pada dasarnya metode pembelajaran mesin (machine learning) sangat baik dalam melakukan klasifikasi untuk 2 kelas (binary classification). Potensi peningkatan hasil dapat dilakukan dengan menguji berbagai aspek, misalnya penerapan fitur word occurency (jumlah kemunculan kata) di dalam dokumen dibandingkan TF-IDF, atau penerapan fitur boolean (ada atau tidak) kata di dalam dokumen. Model NB yang dipakai juga bisa diganti dengan model NB lainnya, misalnya ada model Bernoulli NB, ada model Gaussian NB, dan lainnya. Selain itu dapat juga dengan menganalisa data hasil pelabelan, karena apabila kualitas data tidak baik (banyak inkonsistensi dalam penetapan label kelas), maka hasil klasifikasi juga akan tidak bagus.



Gambar 6. Confusion Matrix dari hasil prediksi pada data testing

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dengan metode Naïve Bayes Classifier pada proses pengklasifikasian sentimen yang diambil dari Twitter, yaitu mengenai topik pindah Ibu Kota Negara berhasil dilakukan. Penelitian dilakukan menggunakan 4000 data yang dibagi menjadi dua kelas, yaitu kelas positif dan kelas negatif, dengan masing-masing sebanyak 2000 data. Pengujian dilakukan dengan dua cara, penggunaan experiment setup dan penerapan threshold. Data dibagi menjadi dua bagian, yaitu 90% data latih (3600 data) dan 10% data uji (400 data). Hasil pengujian terbaik didapatkan dari experiment setup dengan id experiment 14, yaitu hanya menggunakan text preprocessing basic (cleaning, case folding, dan tokenizing) dan penggunaan threshold dengan nilai 0,0001. Akurasi yang didapatkan sebesar 77,00% dengan precision 77,06%, recall 77,08% dan F1-Score 77,00%. Hasil akurasi yang tidak tinggi disebabkan oleh banyaknya kata yang digunakan untuk melatih model tidak cukup mewakili sentimen yang sebenarnya, hal ini bisa dilihat pada word cloud pada gambar 3 dan gambar 4. Berdasarkan hasil riset memverifikasi kalau metode Naïve Bayes Classifier menghasilkan nilai ataupun tingkat akurasi yang cukup baik.



REFERENCES

1. R. N. Fahmi, Nursyifa, and A. Primajaya, "Analisis Sentimen Pengguna Twitter Terhadap Kasus Penembakan Laskar Fpi Oleh Polri Dengan Metode Naive Bayes Classifier," 2021.
2. A. Mustofa and R. Novita, "Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter," *Technology and Science (BITS)*, vol. 4, no. 1, 2022, doi: 10.47065/bits.v4i1.1628.
3. F. Septianingrum and A. S. Y. Irawan, "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 799, Jul. 2021, doi: 10.30865/mib.v5i3.2983.
4. M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," *dataindonesia.id*. Accessed: Dec. 13, 2022. [Online]. Available: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>
5. F. Amirullah, S. Alam, and M. I. Sulisty, "Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naive Bayes," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 69–76, Aug. 2023, doi: 10.55123/storage.v2i3.2293.
6. A. Faisal et al., "Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization," *Jointecs (Journal of Information Technology and Computer Science)*, vol. 5, no. 2, pp. 61–70, 2020.
7. A. P. Giovani, 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.
8. S. Kurniawan, W. Gata, D. Ayu Puspitawati, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 176–183, 2019.
9. Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naive Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.
10. A. D. D. Kurniawan, "Analisis Sentimen Topik Pindah Ibu Kota Pada Twitter Menggunakan Metode Modified K-Nearest Neighbor," *UIN Suska Riau, Pekanbaru*, 2021.
11. N. Ruhjana, "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes," *Jurnal IKRA-ITH Informatika*, vol. 3, no. 1, pp. 94–99, 2019, [Online]. Available: www.situs.com
12. D. Ikarari, Y. Fajarwati, and Widiastuti, "Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum Mrt Jakarta Menggunakan Naive Bayes Classifier," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 1, pp. 64–75, 2020, doi: 10.35760/ik.2020.v25i1.2427.
13. F. E. Purwiantono and A. Aditya, "Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text," *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, pp. 68–73, 2020.
14. A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *Jurnal Riset Komputer*, vol. 8, no. 6, pp. 2407–389, 2021, doi: 10.30865/jurikom.v8i6.3655.
15. R. Indransyah, Y. H. Chrisnanto, and P. N. Sabrina, "Klasifikasi Sentimen Pergelaran MotoGP Di Indonesia Menggunakan Algoritma Correlated Naive Bayes Clasifier," *Infotech*, vol. 8, pp. 60–66, 2022, doi: 10.31949/infotech.v8i2.3103.
16. L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naive Bayes untuk Analisis Sentimen pada Ulasan Aktual di Google Play Store," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, Aug. 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
17. A. M. Raya, F. Nurbaiti, and S. Detin, "Klasifikasi Sentimen Masyarakat Terhadap Kenaikan Harga Tiket Pesawat Pada Twitter Menggunakan Naive Bayes," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 4, no. 2, pp. 239–246, 2019.
18. S. Alim, Anwari, and Hozairi, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Jurnal Ilmiah NERO*, vol. 6, no. 2, pp. 133–144, 2021.
19. F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 19–25, Jul. 2021, doi: 10.31603/komtika.v5i1.5185.
20. E. S. R. Br.Situmorang, M. K. Anam, R. Rahmadden, and A. N. Ulfah, "Perbandingan Algoritma SVM Dan NBC Dalam Analisa Sentimen Pilkada Pada Twitter," *CSRID (Computer Science Research and Its Development Journal)*, vol. 13, no. 3, p. 169, Nov. 2021, doi: 10.22303/csrid.13.3.2021.169-179.
21. A. R. Kardian and D. Gustiana, "Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naive Bayes Classifier," *Jurnal Ilmiah Komputasi*, vol. 20, no. 1, Mar. 2021, doi: 10.32409/jikstik.20.1.401.
22. I. Sari and D. Indarti, *Text Mining : Praktik Klasifikasi Dan Pemodelan Topik Dengan Phytion*. Ponorogo: Uwais Inspirasi Indonesia, 2022.
23. N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Dasar di Twitter Menggunakan Supervised Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3216.
24. V. Ayumi, H. Noprisson, M. Utami, E. D. Putra, and M. Purba, *Konsep Dasar Natural Language Processing (NLP)*. Sukabumi: CV Jejak, 2023.
25. R. Ariyanti, T. Misriati, and A. Sagiyanto, "Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 218–227, 2023, doi: 10.47065/josyc.v5i1.4562.

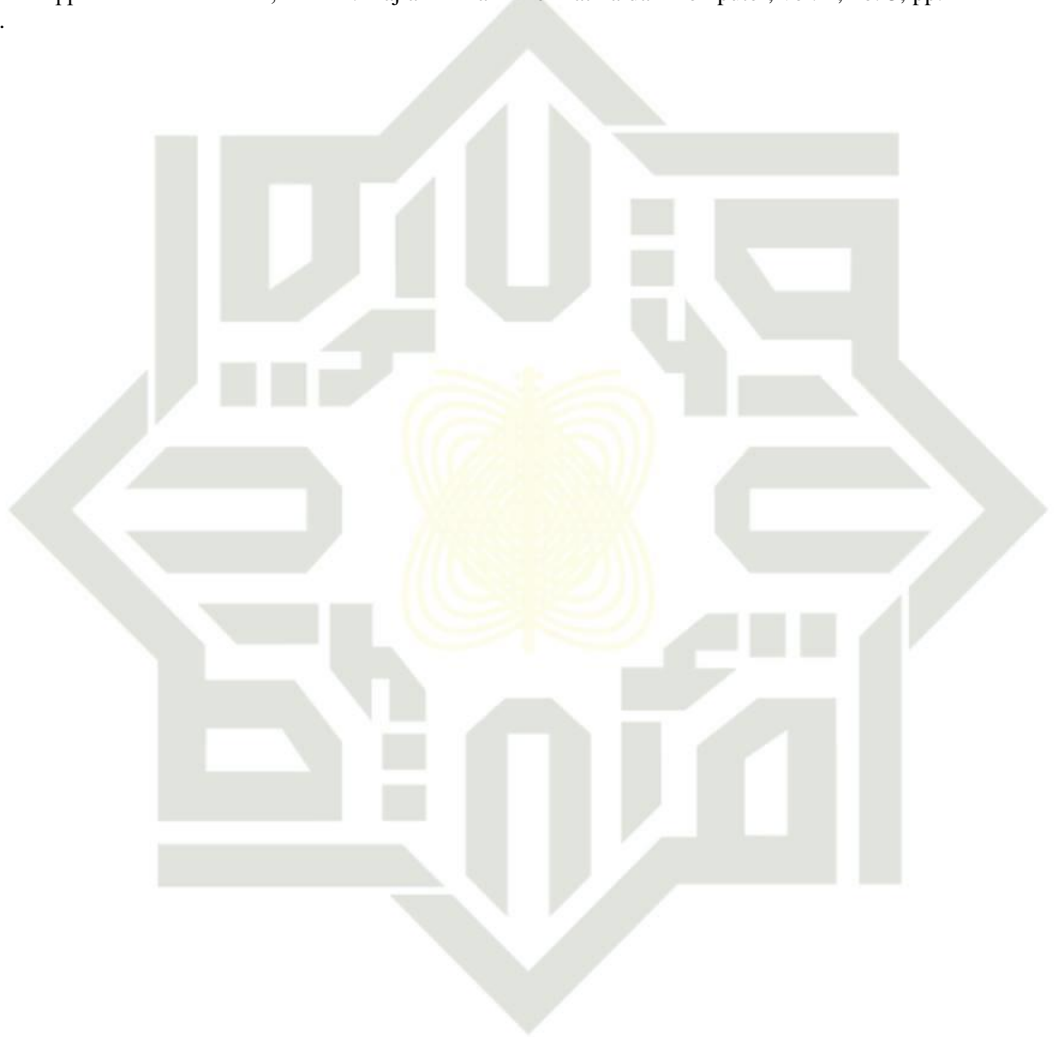
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

- 6] C. F. Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 4, pp. 1702–1708, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - 7] S. Fide, Suparti, and Sudarno, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (SVM) Dan Asosiasi," *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, 2021, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
 - 8] A. Kausar, A. Irawan, Wahyuddin, and I. Fernando, "Implementasi Algoritma Naïve Bayes Classifier Untuk Penilaian Kinerja Dosen," *Jurnal PROSISKO*, vol. 10, no. 2, 2023.
 - 9] S. W. U. Vitandy, A. A. Supianto, and F. A. Bachtiar, "Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naïve Bayes Classifier," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 6, pp. 6080–6088, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - 10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *JMLR* 12.
 - 11] S. Azhar, Yusra, M. Fikry, S. Agustian, and I. Afrianty, "Klasifikasi Sentimen Masyarakat di Twitter Terhadap Ganjar Pranowo dengan Metode Support Vector Machine," *KLK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1660–1667, Dec. 2023.
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Medan, 31 Maret 2024

274/JSON/LOA/III/2024

Surat Penerimaan Naskah Publikasi Jurnal

Kepada Yth,
Bapak/Ibu **Jozu Dermawan**
Di Tempat

Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada **Jurnal Sistem Komputer dan Informatika (JSON)** (eISSN 2685-998X), dengan judul:

Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara Pada Twitter Menggunakan Metode Naïve Bayes Classifier

Penulis: **Jozu Dermawan(*)**, Yusra, Muhammad Fikry, Surya Agustian, Lola Oktavia


Berdasarkan hasil review, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Volume 5, Nomor 3, Maret 2024**.

Informasi QR-Code dapat digunakan untuk melihat link LOA Jurnal Sistem Komputer dan Informatika (JSON), **Volume 5, Nomor 3, Maret 2024** yang telah dikeluarkan.

Sebagai informasi tambahan, saat ini **Jurnal Sistem Komputer dan Informatika (JSON)** telah **TERAKREDITASI** dengan Peringkat **SINTA 4** berdasarkan SK Kepmendikbudristek No. **158/E/PT/2021 tertanggal 09 Desember 2021** dimulai dari Volume 1 No 1. 2019.

Demikian informasi yang kami sampaikan, atas perhatiannya kami ucapkan terimakasih.

Hormat Kami,


Surya Darma Nasution, M.Kom
Managing Editor



Tembusan:

1. Editor JSON
2. Author
3. Bertinggal

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Surat Penerimaan Naskah Publikasi Jurnal
Kecamatan Hampar Mas, Kabupaten Langkat, Provinsi Sumatera Utara
Kantor: Jl. Sisingamangaraja No. 338, Medan, Sumatera Utara
Telp: (061) 451 2311
Email: jurnal.json@gmail.com
Website: <https://ejournal.stmik-budidarma.ac.id/index.php/json>