

Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes

Andra Setiawan¹, Ryan Randy Suryono^{1,*}

¹ Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia

* Correspondence: ryan@teknokrat.ac.id

Copyright: © 2024 by the authors

Received: 4 April 2024 | Revised: 13 April 2024 | Accepted: 20 April 2024 | Published: 20 Juni 2024

Abstrak

Kebijakan Pemerintah dalam pemindahan Ibu Kota Nusantara (IKN) dipandang kontroversial, hal ini menimbulkan beragam tanggapan dari masyarakat, khususnya pada media sosial X. Penelitian ini bertujuan untuk menganalisis sentimen *tweet* yang berkaitan dengan IKN dan melakukan perbandingan dua algoritma. Pada eksperimen ini berhasil mengumpulkan data *tweet* mengenai IKN pada aplikasi X sebanyak 5128 data, jumlah data IKN tersebut diklasifikasikan kedalam sentimen positif sebanyak 2598 data dan sentimen negatif sebanyak 1659 data. Untuk menganalisis kedua sentimen tersebut diperlukan pendekatan menggunakan teknik *Text Mining* dengan membandingkan dua algoritma, yaitu algoritma *Support Vector Machine (SVM)* dan *Naive Bayes*. Pada penelitian ini dalam meningkatkan kinerja kedua algoritma untuk mempelajari suatu data diperlukan optimasi *SMOTE* yang bertujuan mengatasi ketidakseimbangan data. Hasil temuan kami menunjukkan bahwa algoritma *SVM* mencapai akurasi sebesar 84%, sedangkan algoritma *Naive Bayes* mencapai akurasi sebesar 77%. Dengan demikian, dapat disimpulkan bahwa algoritma *SVM* lebih unggul dibandingkan dengan algoritma *Naive Bayes*. Selain itu, penggunaan optimasi *SMOTE* juga terbukti meningkatkan kemampuan kedua algoritma dalam mengidentifikasi sentimen positif, seperti yang terlihat dari nilai *precision*, *recall*, dan *F1-Score*. Algoritma *SVM* mencapai *precision* sebesar 82%, *recall* sebesar 86%, dan *F1-Score* sebesar 84%, sementara algoritma *Naive Bayes* memiliki *precision* sebesar 71%, *recall* sebesar 92%, dan *F1-Score* sebesar 80%.

Kata kunci: ibu kota nusantara; support vector machine; naïve bayes; smote; analisis sentimen

Abstract

The Government's policy of moving the Indonesian capital city (IKN) is considered controversial, sparking various responses from the public, especially on the social media platform X. This research aims to analyze tweet sentiment related to IKN and compare two algorithms. In this experiment, we collected 5,128 tweets regarding IKN from the X application. The dataset was classified into 2,598 positive sentiments and 1,659 negative sentiments. To analyze these sentiments, we used Text Mining techniques, comparing the Support Vector Machine (SVM) and Naive Bayes algorithms. To improve the performance of these algorithms in analyzing the data, SMOTE optimization was employed to address data imbalance. Our findings show that the SVM algorithm achieves an accuracy of 84%, while the Naive Bayes algorithm achieves an accuracy of 77%. Thus, it can be concluded that the SVM algorithm is superior to the Naive Bayes algorithm. Furthermore, the use of SMOTE optimization proved to enhance the ability of both algorithms to identify positive sentiment, as evidenced by the precision, recall, and F1-Score values. The SVM algorithm achieved a precision of 82%, recall of 86%, and F1-Score of 84%, while the Naive Bayes algorithm achieved a precision of 71%, recall of 92%, and F1-Score of 80%.

Keywords: ibu kota nusantara; support vector machine; naïve bayes; smote; sentiment analysis



PENDAHULUAN

Pemindahan Ibu Kota bukan merupakan wacana baru di Indonesia, berawal pada tahun 1957 Presiden Soekarno pertama kali mengusulkan ide pemindahan Ibu Kota ke Palangkaraya dengan latar belakang kota yang berlokasi ditengah negara kepulauan Indonesia, namun ide itu tidak pernah terlaksana dan tahun 1964 Soekarno menetapkan Jakarta sebagai Ibu kota. Pada masa Presiden Soeharto tahun 1990 muncul wacana pemindahan ibu kota ke Jonggol. Kemudian pada masa Presiden Susilo Bambang Yudhoyono ide pemindahan ibu kota kembali mencuat dikarenakan berbagai permasalahan yang ada di Jakarta (Wahyudi et al., 2023).

Pada rapat yang dilaksanakan tanggal 29 April 2019 Presiden Joko Widodo menetapkan pemindahan ibu kota indonesia ke luar pulau jawa, tepatnya di wilayah Kabupaten Penajam Paser Utara Provinsi Kalimantan Timur (Septiani et al., 2022). Menindak lanjuti keputusan Presiden tersebut, tanggal 17 januari 2022 Kepala Bappenas Suharso Monoarfa menetapkan nama ibu kota baru sebagai "Nusantara" pada rapat Rancangan Undang-Undang Ibu Kota Negara (RUU IKN) (Sitio et al., 2024).

Pemindahan Ibu Kota Nusantara (IKN) mengundang beragam perspektif dari masyarakat. Dimulai dari penentuan lokasi yang dipandang kontroversial, penetapan Undang-Undang yang dianggap terlalu tergesa-gesa dan belakangan ini, pemerintah juga mengajak partisipasi aktif masyarakat Indonesia dalam mendukung pembangunan Ibu Kota Nusantara (Huda & Yel, 2024). Banyak masyarakat yang mengungkapkan pandangan mereka, baik dalam diskusi langsung maupun melalui media sosial seperti Twitter. Twitter yang sekarang ini telah berganti nama menjadi platform X adalah salah satu platform media sosial yang memungkinkan pengguna mengirim pesan singkat atau "*tweet*" dengan batasan 280 karakter (Lutfiyanto et al., 2021). Pengguna memanfaatkan platform ini untuk membagikan informasi, serta terlibat dalam percakapan daring mengenai berbagai berita terkini, dan isu-isu aktual, termasuk dalam konteks pemindahan ibu kota (Normawati & Prayogi, 2021). maka dari itu diperlukan analisis sentimen untuk mengetahui arah dan tingkat intensitas sentimen positif maupun negatif masyarakat terhadap kebijakan pemindahan Ibu Kota.

Analisis sentimen merupakan salah satu bagian penting dari proses *text mining*. *Text mining* adalah proses ekstraksi informasi berharga dari teks tidak terstruktur menggunakan algoritma dan analisis bahasa alami untuk mengungkap pola dan informasi penting yang tersembunyi di dalam teks (Putri et al., 2022). Pada penelitian ini metode klasifikasi digunakan dalam proses analisis sentimen. Klasifikasi dalam *text mining* adalah proses mengelompokkan dokumen atau teks ke dalam kelas atau kategori tertentu berdasarkan pada karakteristik tertentu yang terdapat dalam teks tersebut (Sari & Suryono, 2024). Metode klasifikasi ini sering digunakan untuk menganalisis teks dalam skala besar, seperti klasifikasi sentimen, klasifikasi topik, atau identifikasi spam dalam email. Algoritma klasifikasi seperti *Naïve Bayes* dan *Support Vector Machine* (SVM), sering digunakan dalam *text mining* untuk memprediksi kategori atau label yang tepat untuk setiap dokumen berdasarkan pada fitur-fitur yang diekstrak dari teks (Singgalen, 2022).

SVM merupakan algoritma klasifikasi yang kuat, cocok untuk data linier dan non-linier, serta cenderung menghasilkan batas keputusan yang optimal (Sihombing & Yulianti, 2021). Naive Bayes adalah sebuah algoritma klasifikasi dalam pembelajaran mesin yang berdasarkan pada Teorema Bayes. Algoritma ini dianggap "*naive*" karena asumsinya yang sederhana dan independen antar fitur, meskipun di dunia nyata fitur-fitur mungkin saling terkait (Hanafiah et al., 2023). *Naïve Bayes* salah satu algoritma klasifikasi yang sederhana namun efektif, sering memberikan hasil yang memuaskan terutama dalam kasus dengan jumlah fitur yang relatif besar dibandingkan dengan jumlah sampel (Sihombing et al., 2021). Dengan memanfaatkan fitur-fitur yang diekstrak dari teks, keduanya dapat memberikan prediksi sentimen yang akurat. Pada penelitian terdahulu terhadap opini masyarakat Indonesia mengenai pemindahan ibu kota melalui platform Twitter menggunakan metode *Naïve Bayes* dengan *TF-IDF*. Hasil analisis

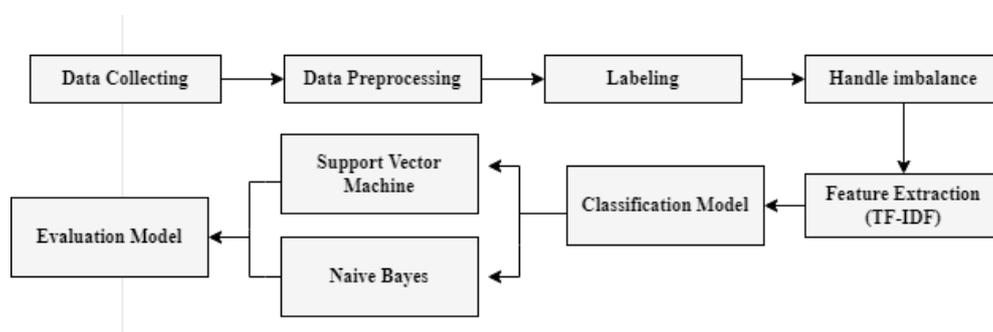
menunjukkan sentimen negatif merupakan yang paling dominan dalam opini masyarakat terkait pemindahan ibu kota. Akurasi tertinggi sebesar 60.66%, Presisi sebesar 66,67% dan *Recall* sebesar 66,67% Dengan membagi data menjadi 80% untuk data training dan 20% untuk data testing. (Kurniawan & Waluyo, 2022).

Pada penelitian lainnya yang membahas analisis sentimen terhadap Popularitas Aplikasi Moodle Dan Edmodo dengan penggunaan algoritma *Support Vector Machine*. Dapat disimpulkan bahwa pengguna cenderung memberikan sentimen positif lebih tinggi terhadap aplikasi Edmodo daripada aplikasi Moodle. Persentase sentimen positif pada Edmodo adalah 67%, dengan akurasi 83,7%, *precision* 93%, *recall* 82%, dan *f1-score* 87%. Sementara Moodle, persentase sentimen negatif adalah 67%, dengan akurasi 81,6%, *precision* 79%, *recall* 100%, dan *f1-score* 88% (Yolanda et al., 2022). Dari hasil beberapa temuan sebelumnya dapat dilihat bahwa penelitian analisis sentimen sebelumnya hanya menggunakan satu metode, dan tidak menggunakan optimasi *SMOTE* untuk mengatasi ketidakseimbangan data, oleh karena itu penelitian saat ini membandingkan kedua metode algoritma SVM dan Naïve bayes.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan pemindahan ibu kota dengan melakukan perbandingan algoritma SVM dan Naïve Bayes yang dikombinasikan dengan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani masalah ketidakseimbangan data, dan berupaya untuk menemukan algoritma klasifikasi yang paling optimal dan perbandingan akurasi, presisi, *recall* dan *F1-Score* dari kedua algoritma tersebut.

METODE

Penelitian ini menggunakan pendekatan kuantitatif, dalam topik analisis sentimen sampel data diperoleh dari platform twitter menggunakan kata kunci “Ibu Kota Nusantara”. Peneliti akan mengeksplorasi beragam perspektif, dan pandangan yang tercermin dalam *tweet-tweet* pengguna. Adapun alur dalam penelitian ini mencakup 7 tahapan, yaitu *Data collection*, *Data preprocessing*, *Labelling*, kemudian *feature extraction* dengan *TF-IDF*, *Handle imbalance* dengan penerapan optimasi *SMOTE*, lalu *Classification Model* menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*, dan tahapan terakhir *Evaluation Model*. Gambar 1 merupakan skema dari penelitian ini.



Gambar 1. Skema penelitian

Pada eksperimen ini diperoleh dataset melalui proses *crawling* dari platform media sosial twitter menggunakan *library Harvest* dengan memanfaatkan Bahasa pemrograman *Pyhton*. Data berupa *tweet* pengguna menggunakan Bahasa Indonesia dengan kata kunci “Ibu Kota Nusantara”. Pengambilan data dimulai dari rentang waktu januari hingga maret 2024 dengan jumlah data yang didapatkan sebanyak 5128 data *tweet*. Setelah data didapatkan selanjutnya pada tahap awal dalam analisis data yaitu *preprocessing*. Tahap ini mencakup berbagai tindakan seperti *data cleaning* yang berfungsi untuk membersihkan data dari tanda baca, simbol, Uniform Resource Locator (URL), data duplikat, dan data kosong. *Case folding*

merupakan proses untuk mengubah semua huruf dalam teks menjadi huruf kecil. *Tokenizing* berfungsi untuk memecah suatu kalimat menjadi unit-unit kecil atau *token*. *Filtering* digunakan untuk meyaring suatu kata yang tidak dibutuhkan, dan proses terakhir dalam preprocessing yaitu *stemming* merupakan proses untuk menghilangkan awalan atau akhiran kata dalam sebuah kata sehingga hanya tersisa bentuk dasarnya atau kata dasar (Hendrawan et al., 2022). Setelah melalui tahapan *preprocessing* data akan diberikan kategori yang berbentuk label.

Proses pelabelan data menggunakan pustaka *texblob* yang menghasilkan label dengan sentimen positif dan negatif. Pada label sentimen negatif menghasilkan 2598 data dan sentimen positif 1695 data. Dengan hasil yang diperoleh dari proses pelabelan terjadi ketidakseimbangan data oleh karena itu perlu dilakukan proses *Handle Imbalance* menggunakan SMOTE. SMOTE merupakan teknik yang digunakan dalam pembelajaran mesin untuk mengatasi ketidakseimbangan kelas dalam data (Hidayatullah et al., 2023). Pada tahapan selanjutnya untuk menilai pentingnya suatu kata dalam sebuah dokumen dilakukan proses ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode *TF-IDF* memberikan bobot yang lebih besar untuk kata-kata yang muncul sering dalam sebuah dokumen tetapi jarang di seluruh koleksi dokumen. Ini membantu menentukan kata-kata yang paling penting dalam sebuah dokumen dengan memperhitungkan kedua faktor tersebut (Rahayu et al., 2022). Pada proses klasifikasi model digunakan algoritma *Support Vector Machine* dan *Naïve Bayes*.

SVM merupakan algoritma pada pemodelan klasifikasi yang bertujuan untuk memisahkan dua kelas dengan membuat hyperplane yang memiliki margin maksimum di antara kelas-kelas tersebut dalam ruang fitur. SVM bekerja dengan menemukan *hyperplane* terbaik yang memaksimalkan jarak antara titik-titik data terdekat dari kedua kelas, yang disebut *support vectors*. Algoritma ini efektif dalam menangani *dataset* dengan dimensi tinggi dan dapat bekerja dengan baik dalam kasus jumlah fitur lebih besar dari jumlah sampel data. Selanjutnya algoritma *Naive bayes* merupakan sebuah algoritma klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi independensi antara fitur-fitur (Saputra & Noor Hasan, 2023). *Naive bayes* memanfaatkan perhitungan probabilitas untuk memprediksi kelas dari sebuah data berdasarkan pada distribusi probabilitas dari fitur-fiturnya.

Tahap akhir penelitian ini mengevaluasi kinerja model klasifikasi, menggunakan *confusion matrix* untuk membandingkan prediksi model dengan nilai sebenarnya dari data uji. Dengan menggunakan *confusion matrix*, peneliti mampu mengukur berbagai metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-Score*, yang memberikan pemahaman yang lebih holistik tentang kemampuan model dalam mengklasifikasikan berbagai kelas.

HASIL DAN PEMBAHASAN

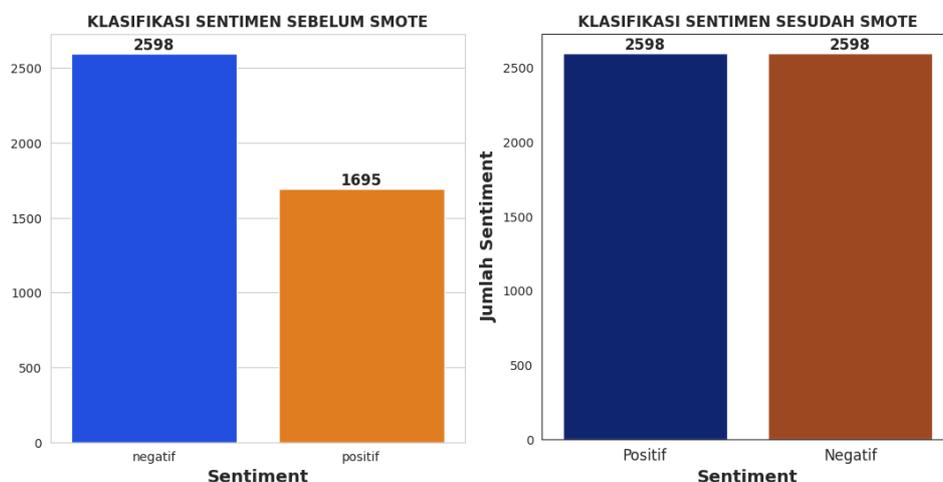
Hasil

Data yang digunakan dalam penelitian ini diperoleh dari tweet pengguna media sosial Twitter dengan menggunakan metode *crawling* menggunakan kata kunci "Ibu Kota Nusantara". Total terdapat 5128 *tweet* yang dikumpulkan dari rentang waktu Januari hingga Maret 2024. Selanjutnya data yang didapatkan akan masuk pada tahap preprocessing. Tabel 1 menunjukkan hasil dari proses *preprocessing* data yang melibatkan beberapa tahapan. Pada *data cleaning* merupakan tahapan menghapus tanda baca, simbol dan URL, kemudian *case folding* mengubah semua huruf menjadi huruf kecil, lalu tahapan *tokenizing* memecah kalimat menjadi kata individu, kemudian tahapan *filtering* yang menghapus kata-kata yang tidak penting pada *tokenizing* dan yang terakhir tahapan *stemming* yang menghapus kata imbuhan dan menjadi kata dasar. Setelah melalui proses *preprocessing* data teks dapat menjadi lebih terstruktur, bersih, dan siap untuk proses analisis sentimen Hal ini membantu meningkatkan kualitas dan akurasi hasil analisis serta efisiensi dalam pengolahan data.

Tabel 1. Hasil preprocessing

Tahapan	Hasil
Data <i>Tweet</i>	@SadikGalib Dukungan terhadap IKN menunjukkan bahwa masyarakat Indonesia memiliki semangat untuk mengatasi berbagai permasalahan dan menggapai cita-cita bersama
Data <i>cleaning</i>	Dukungan terhadap IKN menunjukkan bahwa masyarakat Indonesia memiliki semangat untuk mengatasi berbagai permasalahan dan menggapai cita cita bersama
<i>Case folding</i>	dukungan terhadap ikn menunjukkan bahwa masyarakat indonesia memiliki semangat untuk mengatasi berbagai permasalahan dan menggapai cita cita bersama
<i>Tokenizing</i>	['', 'dukungan', 'terhadap', 'ikn', 'menunjukkan', 'bahwa', 'masyarakat', 'indonesia', 'memiliki', 'semangat', 'untuk', 'mengatasi', 'berbagai', 'permasalahan', 'dan', 'menggapai', 'cita', 'cita', 'bersama']
<i>Filtering</i>	['', 'dukungan', 'terhadap', 'ikn', 'menunjukkan', 'bahwa', 'masyarakat', 'indonesia', 'memiliki', 'semangat', 'mengatasi', 'berbagai', 'permasalahan', 'menggapai', 'cita', 'cita', 'bersama']
<i>Stemming</i>	dukung hadap ikn menunjuk bahwa masyarakat indonesia miliki semangat atasi bagai masalah gapai cita cita bersama

Setelah melalui semua tahapan selanjutnya dalam penelitian, jumlah data berkurang menjadi 4293 *tweet*. Dari jumlah tersebut, terdapat 1695 *tweet* yang diklasifikasikan sebagai sentimen positif dan 2598 *tweet* sebagai sentimen negatif. Dengan hasil tersebut Ketidakseimbangan data antara sentimen positif dan negatif dapat mengakibatkan model algoritma seperti *Support Vector Machine* dan *Naïve Bayes* cenderung lebih terlatih untuk mengenali kata-kata yang bersifat negatif. Hal ini dapat meningkatkan akurasi dan performa klasifikasi untuk sentimen negatif. Oleh karena itu dalam penelitian ini, dilakukan optimasi menggunakan metode *SMOTE* untuk menyeimbangkan jumlah data antara sentimen positif dan negatif, sehingga model tidak cenderung mendominasi satu sentimen tertentu. Dengan menerapkan optimasi *SMOTE*, jumlah data minoritas akan diperbesar sehingga seimbang dengan jumlah data mayoritas. Berikut gambar hasil dari perbandingan klasifikasi dengan optimasi *SMOTE*.

**Gambar 2.** Perbandingan klasifikasi sentimen

Berdasarkan hasil yang terlihat pada gambar 2 menghasilkan klasifikasi sentimen data sebelum optimasi *SMOTE* dan sesudah optimasi *SMOTE*. Sebelum penerapan optimasi *SMOTE* hasil

sentimen negatif sebanyak 2598 data dan sentimen Positif sebanyak 1695 data, setelah penggunaan optimasi *SMOTE* jumlah sentimen akan sama yaitu 2598 data. Hal ini menyebabkan model algoritma menjadi lebih seimbang dalam mempelajari sentimen, tanpa adanya data yang mayoritas atau minoritas.

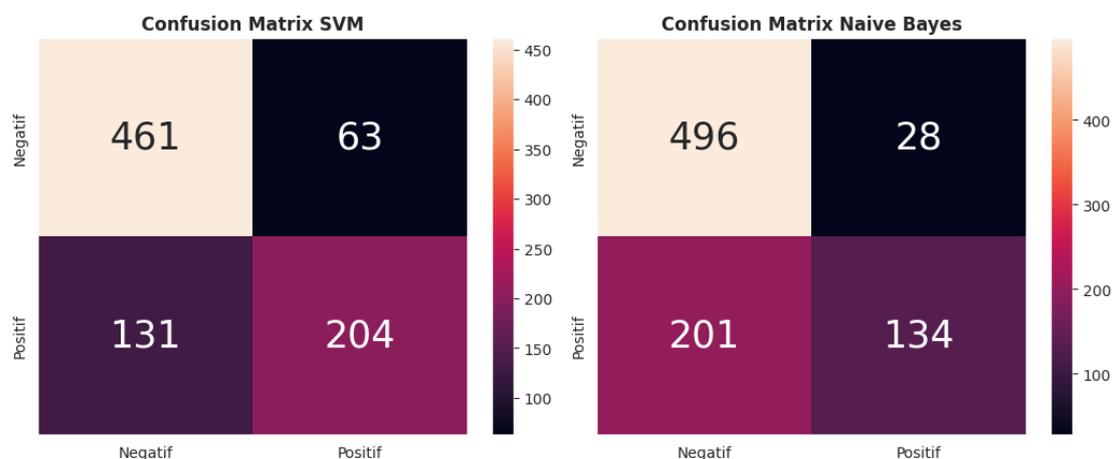
Pada tahap pengujian dalam penelitian ini dilakukan perbandingan antara algoritma SVM dan Naive bayes dengan penggunaan 80% sebagai data training dan 20% sebagai data testing. Data yang digunakan mencakup dua kondisi, yaitu sebelum dan setelah dilakukan optimasi *SMOTE* untuk menemukan kinerja algoritma terbaik. Berikut merupakan tabel hasil dari perbandingan klasifikasi SVM dan naïve bayes dengan *SMOTE* dan tanpa *SMOTE*.

Tabel 2. Hasil klasifikasi

Report	Support Vector Machine				Naïve Bayes			
	Sebelum SMOTE		Sesudah SMOTE		Sebelum SMOTE		Sesudah SMOTE	
	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif
Akurasi	0,77		0,84		0,73		0,77	
Presisi	0,76	0,78	0,82	0,85	0,83	0,71	0,71	0,88
Recall	0,61	0,88	0,86	0,81	0,40	0,95	0,92	0,61
F1-Score	0,68	0,83	0,84	0,83	0,54	0,81	0,80	0,72

Pada tabel 2 hasil eksperimen komparasi dari kedua algoritma menunjukkan bahwa pengaruh optimasi *SMOTE* terhadap performa kinerja model mengalami peningkatan. pada algoritma SVM terjadi peningkatan nilai akurasi 77% menjadi 84% setelah penerapan *SMOTE*. Disisi lain *algoritma naïve bayes* juga mengalami peningkatan nilai akurasi 73% menjadi 77% setelah penerapan *SMOTE*.

Pada algoritma SVM dengan sentimen positif juga mengalami peningkatan dalam hal nilai presisi 76% menjadi 82%, nilai recall 61% menjadi 86%, dan nilai F1-Score 68% menjadi 84%. Sedangkan pada algoritma *naïve bayes* terjadi penurunan dalam nilai presisi 83% menjadi 71%, namun terjadi peningkatan pada nilai recall 40% menjadi 92% dan nilai F1-Score 54% menjadi 80%. Selain itu, dalam eksperimen penelitian juga terdapat perbandingan *confusion matrix* antara kedua algoritma. untuk memahami lebih lanjut sejauh mana kinerja model dalam mengklasifikasikan sentimen. Ini memberikan gambaran menyeluruh tentang seberapa baik model dapat mengenali sentimen positif dan negatif.



Gambar 3. Confusion matrix sebelum smote

Terlihat pada gambar 3 bahwa sebelum penerapan optimasi *SMOTE* kedua algoritma lebih dominan dalam memprediksi kata dengan nilai sentimen negative terkait IKN. Pada

Dengan cara ini, *SMOTE* membantu mendorong model untuk belajar dengan lebih baik dari kelas minoritas, sehingga meningkatkan kemampuannya untuk membedakan kata yang jarang muncul (Anjani et al., 2023). Optimasi *SMOTE* berdampak signifikan dalam membantu algoritma membedakan kata, Hal ini terbukti dari gambar 4 perbandingan *confusion matrix* kedua algoritma menampilkan bahwa algoritma SVM dan Naïve bayes lebih optimal dalam menentukan kata dalam nilai True Positif.

Pada eksperimen dalam membandingkan algoritma SVM dan Naïve bayes dengan menerapkan optimasi *SMOTE* menunjukkan bahwa algoritma SVM lebih unggul, hal ini dikarenakan *SVM* cenderung memberikan lebih banyak perhatian pada kelompok minoritas dalam proses pembuatan keputusan karena fokusnya pada penemuan *hyperplane* yang memaksimalkan margin pada setiap sentimen. Ini berarti *SVM* secara alami lebih cenderung untuk membedakan antara kelas-kelas yang kurang umum. Akurasi algoritma SVM setelah dilakukan optimasi *SMOTE* mencapai 84%, merupakan yang tertinggi. Analisis juga menunjukkan bahwa kedua algoritma memiliki keunggulan pada presisi dan *recall* untuk kedua kelas sentimen positif dan negatif. Oleh karena itu, *F1-Score* digunakan sebagai indikator kinerja model dikarenakan *F1-Score* menghitung nilai dari presisi dan *recall*, sehingga nilai *F1-Score* dapat dijadikan indikator untuk menentukan algoritma terbaik dimana dalam penelitian ini *SVM* mencapai nilai *F1-Score* sebesar 84%, sehingga menunjukkan performa yang sangat baik dibandingkan Naïve Bayes yang memiliki nilai *F1-Score* 80%.

Pada penelitian yang dibahas, terdapat beberapa perbedaan signifikan dan peningkatan dibandingkan dengan penelitian sebelumnya yang hanya menggunakan algoritma Naïve Bayes (Kurniawan & Waluyo, 2022) dan algoritma SVM (Yolanda et al., 2022). Pertama, penelitian ini memperluas cakupan dengan membandingkan dua algoritma klasifikasi yang berbeda, yaitu SVM dan Naïve Bayes, dalam konteks analisis sentimen terkait pemindahan IKN. Selain itu, penggunaan teknik optimasi *SMOTE* untuk mengatasi ketidakseimbangan data memberikan kontribusi yang signifikan dalam meningkatkan performa kedua algoritma. Temuan ini menunjukkan bahwa algoritma SVM dapat mencapai akurasi sebesar 84% dan Naïve Bayes mencapai 77% setelah penerapan *SMOTE*, yang merupakan peningkatan yang cukup mencolok dibandingkan dengan penelitian sebelumnya yang hanya menggunakan *Naïve Bayes*. Dengan demikian, penerapan optimasi *SMOTE* mampu meningkatkan kinerja pada kedua model algoritma SVM dan Naïve Bayes. Eksperimen ini mendapatkan wawasan berupa visualisasi berbentuk wordcloud yang menampilkan visual terkait kata-kata yang sering dilontarkan oleh pengguna aplikasi X mengenai topik IKN seperti “bangun”, “Jakarta”, “dukung”, “rakyat”, “investor”, dan “duit” sehingga dapat diasumsikan dalam penelitian ini instansi terkait perlu memperhatikan pemindahan Ibu Kota ke IKN menimbulkan pro dan kontra pada rakyat terutama terkait dengan besarnya dana yang diperlukan untuk pembangunan.

SIMPULAN

Berdasarkan hasil dari analisis sentimen pada penelitian ini mengenai pemindahan Ibu Kota Nusantara (IKN) memicu beragam reaksi dari masyarakat. Data yang digunakan dalam penelitian ini berjumlah 5128 data mengenai Ibu Kota Nusantara. Melalui analisis sentimen menggunakan algoritma klasifikasi *Support Vector Machine (SVM)* dan *Naïve bayes*, diperoleh pemahaman mendalam tentang arah dan intensitas sentimen positif maupun negatif terhadap kebijakan tersebut. Hasil penelitian ini menunjukkan perbandingan nilai akurasi pada algoritma *Support Vector Machine* mencapai 77% dan algoritma *Naïve Bayes* mencapai 73%. Metode klasifikasi ini mampu mengidentifikasi sentimen dengan akurasi yang signifikan setelah penerapan teknik optimasi *SMOTE* untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa algoritma SVM terbukti lebih unggul dari *Naïve Bayes* dengan akurasi mencapai 84% sedangkan *Naïve bayes* dengan akurasi 77% setelah penerapan *SMOTE*. Selain itu, terdapat peningkatan yang signifikan dalam presisi, *recall*, dan *F1-Score* pada kedua

algoritma setelah penggunaan SMOTE. Pada algoritma SVM nilai presisi 82%, *Recall* 86% dan *F1-Score* 84%. Selain itu penelitian ini mendapatkan hasil berupa pendapat dari masyarakat terhadap pemindahan IKN yang berbentuk visualisasi *wordcloud* yang dapat digunakan oleh instansi terkait untuk menjadi masukan mengenai pemindahan IKN yang membutuhkan dana sangat besar sehingga menimbulkan pro dan kontra di masyarakat.

REFERENSI

- Anjani, A. F., Anggraeni, D., & Tirta, I. M. (2023). Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 163–172. <https://doi.org/10.25077/teknosi.v9i2.2023.163-172>
- Dimas Lutfiyanto, M., Setiawan, E. B., & Si, S. (2021). Expansion Feature dengan Word2Vec untuk Analisis Sentimen pada Opini Politik di Twitter dengan Klasifikasi Support Vector Machine, Naïve Bayes, dan Random Forest. *EProceedings of Engineering*, 8(5), 10399–10410.
- Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1), 34–40. <https://doi.org/10.23960/jitet.v10i1.2262>
- Hanafiah, A., Nasution, A. H., Arta, Y., Wandri, R., Nasution, H. O., & Mardafora, J. (2023). Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier. *INTECOMS: Journal of Information Technology and Computer Science*, 6(1), 230–236. <https://doi.org/10.31539/intecom.v6i1.5845>
- Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 143–149. <https://doi.org/10.29408/edumatic.v6i1.5613>
- Hidayatullah, H., Purwantoro, P., & Umaidah, Y. (2023). Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1546–1553. <https://doi.org/10.36040/jati.v7i3.6887>
- Huda, C., & Yel, M. B. (2024). Analisa Sentimen Tentang Ibu Kota Nusantara (IKN) Dengan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), 126–130. <https://doi.org/10.55338/jikomsi.v7i1.2846>
- Kurniawan, A., & Waluyo, S. (2022, September). Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter. In *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 1(1), 455–461.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Nugroho, A., & Rilvani, E. (2023). Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan. *Techno.Com*, 22(1), 207–214. <https://doi.org/10.33633/tc.v22i1.7527>
- Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Saputra, A., & Noor Hasan, F. (2023). Analisis Sentimen Terhadap Aplikasi Coffee Meets Bagel Dengan Algoritma Naïve Bayes Classifier. *SIBATIK JOURNAL: Jurnal Ilmiah*

- Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 2(2), 465–474. <https://doi.org/10.54443/sibatik.v2i2.579>
- Sari, P. K., & Suryono, R. R. (2024). Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse. *Jurnal Mnemonic*, 7(1), 31–39. <https://doi.org/10.36040/mnemonic.v7i1.8977>
- Septiani, R. K., Anggraeni, S., & Saraswati, S. D. (2022). Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes. *TEKNIKA: Jurnal Ilmiah Bidang Ilmu Rekayasa*, 16(2), 245–254.
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233–242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 417–426. <https://doi.org/10.30812/matrik.v20i2.1174>
- Singgalen, Y. A. (2022). Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1634–1646. <https://doi.org/10.47065/bits.v4i3.2766>
- Sitio, G. Y., Rumapea, S. A., & Lumbanraja, P. (2023). Analisis Sentimen Pemandangan Ibu Kota Negara Di Media Sosial Twitter Menggunakan metode Convolutional Neural Network (CNN). *METHOTIKA: Jurnal Ilmiah Teknik Informatika*, 3(2), 97-104.
- Wahyudi, F. S., Kunci, K., Ibu, P., Baru, K., Ekonomi, P., & Pemerintahan, T. K. (2023). Humantech Jurnal Ilmiah Multi Disiplin Indonesia Proyek Pemandangan Ikn: Upaya Tata Kelola Pemerintahan Yang Baik Ataukah Untuk Pengembangan Ekonomi. *Jurnal Ilmiah Multidisiplin Indonesia*, 2(9), 1890–1908.
- Yolanda, N., Santi, I. H., & Permadi, D. F. H. (2022). Analisis Sentimen Analisis Sentimen Popularitas Aplikasi Moodle dan Edmodo Menggunakan Algoritma Support Vector Machine. *Jurnal Algoritme*, 3(1), 48–59. <https://doi.org/10.35957/algoritme.v3i1.3313>