DECODE: JURNAL PENDIDIKAN TEKNOLOGI INFORMASI
P-ISSN: 2775-2984 | E-ISSN: 2775-1813
Vol. 6, No 1, Maret 2026, Hal: 51-63
Doi: http://dx.doi.org/10.51454/decode.v6i1.1567
https://journal.umkendari.ac.id/index.php/decode

Analisis Sentimen Pemindahan Ibu Kota Nusantara (IKN) dengan Algoritma *Random Forest, Naïve Bayes,* dan *Support Vector Machine.*

Siti Azizah Aini[1], Afril Efan Pajri[2], Alif Yuanita Kartini[3]

[1]Program Studi Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Indonesia.
[2]Program Studi Sistem Komputer, Universitas Nahdlatul Ulama Sunan Giri, Indonesia
[3]Program Studi Statistika, Universitas Nahdlatul Ulama Sunan Giri, Indonesia.

Artikel Info

Kata Kunci:

Analisis Sentimen;
IKN;
Naïve Bayes;
Random Forest;
SVM.

Keywords:

Sentiment Analysis;
IKN;
Naïve Bayes;
Random Forest;
SVM.

Riwayat Artikel:

Submitted: 13 Januari 2026
Accepted: 17 Februari 2026
Published: 23 Februari 2026

Abstrak: Reaksi masyarakat Indonesia terhadap pemindahan Ibu Kota Nusantara (IKN) bervariasi, terutama di media sosial, yang merupakan platform utama untuk menyuarakan pendapat. Tujuan penelitian ini adalah untuk mengevaluasi persepsi publik tentang rencana pemindahan IKN dengan menggunakan tiga metode klasifikasi yaitu *Random Forest, Naïve Bayes,* dan *Support Vector Machine (SVM*). Data diproses melalui beberapa tahap preprocessing, Tahap-tahap ini termasuk tokenisasi, penghilangan kata umum, pemrosesan huruf, pembersihan teks, dan pengurangan kata dasar. Pelabelan data mengelompokkan komentar dalam kategori sentimen positif dan negatif, sedangkan ekstraksi fitur dilakukan dengan metode TF-IDF. Dataset tersebut kemudian dibagi menjadi dua, 20% untuk data pengujian dan 80% untuk data latihan. Hasil penelitian ini menunjukkan bahwa setiap model memiliki tingkat akurasi yang berbeda, dengan *Random Forest* mencapai 98%, *Naïve Bayes* 91%, dan SVM 97%, yang menunjukkan bahwa *Random Forest* adalah algoritma ensemble yang paling akurat dan dapat digunakan untuk analisis sentimen.Studi ini menemukan bahwa setiap model memberikan tingkat akurasi yang berbeda. Hasil penelitian ini menunjukkan bahwa *Random Forest* adalah algoritma ensemble terbaik untuk mengelompokkan sentimen publik tentang masalah pemindahan IKN. Kesimpulan penelitian ini menekankan bahwa menggunakan algoritma ensemble seperti *Random Forest* dapat menghasilkan analisis sentimen yang lebih akurat, dan ini dapat digunakan sebagai pedoman untuk proses pengambilan keputusan di masa mendatang tentang evaluasi kebijakan publik.

Abstract: *Indonesian public reactions to the planned transfer of the Indonesian Capital City (IKN) have varied, particularly on social media, which is a key platform for expressing opinions. The purpose of this study is to analyze public perceptions of the planned transfer of the IKN using three classification methods: Random Forest, Naive Bayes, and Support Vector Machine (SVM). Data were processed through several preprocessing stages, including tokenization, common word removal, case processing, text cleaning, and root word reduction. Data labeling grouped comments into positive and negative sentiment categories, while feature extraction was performed using the TF-IDF method. The dataset was then divided into two, 20% for testing data and 80% for training data. The results of this study show that each model has a different level of accuracy. Which indicates that Random Forest is the most accurate ensemble algorithm and can be used for sentiment analysis. This study found that each model provides a different level of accuracy, Random Forest reaching 98%, Naïve Bayes 91%, and SVM reaching 97%. The results of this study indicate that Random Forest is the best ensemble algorithm for classifying public sentiment regarding the*

relocation of the new capital city. The study's conclusions emphasize that using ensemble algorithms such as Random Forest can produce more accurate sentiment analysis, and this can be used as a guideline for future decision-making processes regarding public policy evaluation.

Corresponding Author:
Siti Azizah Aini
Email: azizahaini2233@gmail.com

PENDAHULUAN

Ibu Kota Negara Republik Indonesia direncanakan akan pindah ke Pulau Kalimantan pada tahun 2024 ditandai dengan akan dilaksanakannya upacara Hari Kemerdekaan Negara Kesatuan Republik Indonesia yang berlokasi di kabupaten Kalimantan Timur yang dinamakan Ibu Kota Nusantara (IKN). Rencana ini didasarkan pada pertimbangan bahwa kondisi Kota Jakarta dinilai tidak lagi memadai untuk berfungsi sebagai ibu kota. Gagasan pemindahan ibu kota ini telah memunculkan beragam tanggapan, baik dari kalangan politisi maupun masyarakat umum (Hadi & Sugiarto, 2025).

Ada banyak pendapat tentang relokasi Ibu Kota Nusantara. Ini terkait dengan pemilihan lokasi, anggapan bahwa proses pengesahan undang-undang terlalu cepat, dan upaya pemerintah untuk mendorong orang di seluruh Indonesia untuk mendukung pembangunan Ibu Kota Nusantara. Presiden Joko Widodo berpikir untuk memindahkan ibu kota dari Jakarta. Tujuannya adalah menyelesaikan penilaian alternatif tempat untuk ibu kota baru Indonesia sebelum akhir tahun 2017. Seluruh lembaga pemerintah akan pindah ke pusat pemerintahan yang baru, menurut Kementerian Perencanaan Pembangunan Nasional, yang diumumkan pada April 2019. Karena tidak terancam gempa bumi, tsunami, dan erupsi vulkanik, tiga provinsi di Kalimantan Selatan, Kalimantan Tengah, dan Kalimantan Timur diusulkan untuk menjadi ibu kota baru (Huda & Yel, 2024). Salah satu sumber data digital yang paling banyak adalah Media Sosial. Oleh karena itu, pendapat masyarakat saat ini menjadi sangat krusial karena bisa memengaruhi cara pandang publik terhadap keputusan pemerintah di Indonesia (Septianingrum & Irawan, 2021).

Penelitian sebelumnya yang didasarkan pada ulasan literatur yang berkaitan dengan kajian ini menunjukkan beberapa pendekatan penelitian yang memanfaatkan algoritma Data Mining. Ulasan pustaka tersebut mencakup artikel ilmiah dari (Huang & Wasito, 2024) yang ditujukan untuk mengelompokkan pemain berdasarkan posisi yang tepat dalam permainan mereka. Metode *Random Forest* prediksi dengan hasil akurasi sebesar 86,89%, tinjauan pustaka yang kedua, artikel dari (Hanifatun & Zahrotun, 2025) menerapkan metode *Naïve Bayes* dalam menentukan kelayakan kredit yang diajukan nasabah, hasil pengujian menggunakan 186 data dengan metode *confusion matrix* diperoleh akurasi 67%, dan tinjauan pustaka yang ketiga artikel dari (Hadi & Sugiarto, 2025) dengan topik menganalisis opini masyarakat terhadap keputusan pemerintah dalam membangun Ibu Kota Nusatara (IKN) performa dari algoritma *Support Vector Machine*, *Logistic Regression*, dan *Naïve Bayes*, Hasil yang dicapai menggunakan algoritma *Support Vector Machine* adalah 80%. Oleh karena itu, penelitian ini untuk mengisi celah penelitian (*research gap*) dengan melakukan analisis sentimen pemindahan IKN melalui perbandingan algoritma *Random Forest, Naïve Bayes*, dan *Support Vector Machine* guna memperoleh model klasifikasi sentimen dengan performa terbaik.

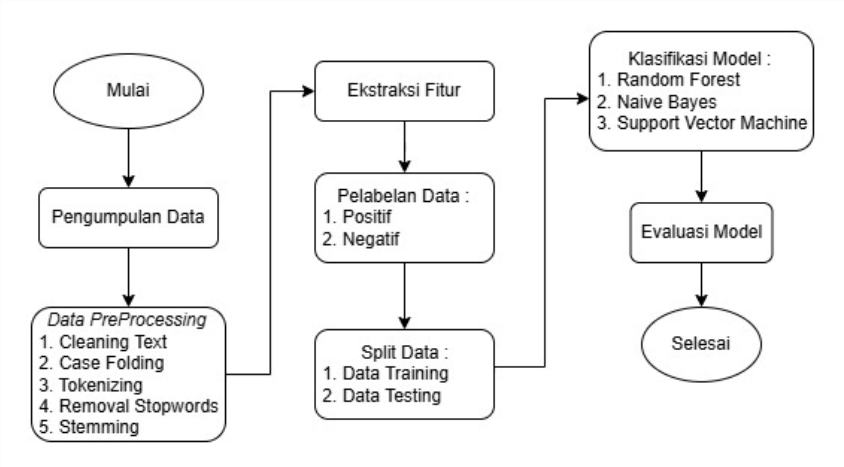
Perbedaan utama dengan penelitian terdahulu berfokus pada cara metodologis yang digunakan dalam klasifikasi data. Pada penelitian terdahulu, sebagian besar studi hanya mengandalkan satu jenis algoritma klasifikasi, seperti *Random Forest* saja atau *Naïve Bayes* saja, sehingga evaluasi performa model menjadi terbatas pada karakteristik dan keunggulan dari satu metode tertentu. Selain itu, penelitian sebelumnya menggunakan tiga algoritma klasifikasi, yaitu *Support Vector Machine (SVM*), *Logistic Regression*, dan *Naïve Bayes*, namun pemilihan metode tersebut masih berada dalam kelompok algoritma yang relatif serupa dari sisi pendekatan statistik dan linearitas model. Penelitian ini menggunakan pendekatan komparatif, tiga algoritma klasifikasi *data mining* yang memiliki karakteristik dan mekanisme kerja yang berbeda. Studi ini tidak hanya menekankan penggunaan satu

metode, tetapi juga melakukan analisis perbandingan kinerja di antara algoritma untuk menentukan bahwa metode ini akan menghasilkan hasil analisis yang lebih komprehensif dan objektif, serta memperkuat validitas hasil penelitian. Selain itu, dengan memanfaatkan berbagai algoritma yang memiliki sejumlah karakteristik, studi ini bisa memberikan pengembangan model klasifikasi yang lebih akurat dan dapat diandalkan.

Data mining adalah metode untuk menemukan pola-pola yang penting untuk menghasilkan informasi, yang dapat digunakan untuk melakukan analisis yang relevan dan memperbaiki proses pengambilan keputusan. Studi ini menyelidiki tentang relokasi ibu kota Negara Kesatuan Republik Indonesia (Hanifatun & Zahrotun, 2025). *Random Forest* merupakan metode *machine learning* yang umum diterapkan untuk mengklasifikasikan data (Huang & Wasito, 2024) . *Naïve bayes* adalah metode data mining yang bisa digunakan dalam membuat model secara cepat, membuat prediksi, dan menyediakan cara baru untuk mengeksplore dan memahami data. *Naive Bayes* merupakan suatu teknik pengelompokan sederhana yang menentukan beragam kemungkinan dengan metode untuk mengumpulkan beragam frekuensi dan kombinasi dari nilai-nilai yang ada dalam kumpulan data yang tersedia (Cesar et al., 2024; Kawani, 2019). Menurut klasifikasi metode *Naïve Bayes*, klasifikasi terdiri dari dua tahap, tahap pelatihan dan tahap pengujian. Tahap pelatihan melibatkan pembuatan model, dan tahap pengujian bertujuan untuk menguji dan mengevaluasi model yang telah dibuat pada tahap pelatihan.Pada metode *naïve bayes* ini digunakan teknik probalititas dalam pengitungannya (Hanifatun & Zahrotun, 2025). *Support Vector Machine (SVM)* memiliki sasaran untuk menemukan fungsi pemisah yang paling efisien di antara fungsi-fungsi yang tersedia untuk membedakan dua tipe objek (Styawati et al., 2021). SVM mampu mengkategorikan informasi baik yang berhubungan dengan garis lurus maupun yang tidak, sehingga sangat sesuai untuk pola data dari platform sosial yang beragam (Sarumpaet & Suryono, 2025; Pamungkas & Cahyono, 2024).

METODE

Studi ini mengkaji persepsi publik tentang pemindahan Ibu Kota Nusantara (IKN). Tujuannya adalah untuk mengetahui bagaimana masyarakat melihat kebijakan pemindahan ibu kota, maka teknik *text mining* dan *machine learning* digunakan untuk mengolah data teks dari sumber data. Pengolahan data dan pemodelan dalam penelitian ini dilakukan menggunakan bahasa pemrograman *Python*. Bahasa pemrograman ini dipilih karena memiliki banyak *library* pendukung yang dapat diandalkan untuk pengolahan data, pemrosesan bahasa alami, dan penggunaan algoritma klasifikasi. Alur lengkap dari proses penelitian ini disajikan pada Gambar 1, yang menjelaskan langkah-langkah penelitian satu per satu, dimulai dari pengumpulan data mentah hingga akhirnya diperoleh hasil pengelompokan sentimen.



Gambar 1. Alur Penelitian

Pengumpulan Data

Riset bertujuan menganalisis komentar publik terhadap pemindahan Ibu Kota Nusantara (IKN). Pada tahap awal, menyiapkan kumpulan data dari *Kaggle* yang didapat dari *https://www.kaggle.com/datasets/alamrrr/sentiment-dataset-for-pemindahan-ibukotaikn/code* berupa data teks komentar dan data diakses dan diperoleh pada tanggal 10 Oktober 2025 pada pukul 15.00 WIB.

Data Pre-Processing

Tahap ini bertujuan untuk memproses informasi, menentukan apakah ada data yang perlu diubah atau disusun ulang dengan metode tertentu yang bisa mempermudah langkah-langkah berikutnya (Marga et al., 2020 (Supriyanto et al., 2023)). Data mentah melalui *preprocessing* sebagai berikut:

1. *Cleaning Text*, yaitu proses penghapusan berkas serta pemilihan kata yang tidak berkaitan seperti html, emotikon, tagar, penyebutan, dan url.
2. *Case folding*, yaitu langkah untuk menyamakan semua kalimat dari teks data.
3. *Tokenizing*, yaitu tahapan dalam memilih dan memotong kata-kata dalam suatu kalimat. Pemisah seperti koma (,), titik (.), dan tanda lainnya.
4. *Removal Stopwords*, yaitu merujuk pada tindakan menghapus kata-kata yang dianggap tidak efektif.

Setelah dilakukan *preprocessing*, maka dataset siap diimplementasikan dengan beberapa model untuk analisis sentimen (Safira & Hasan, 2023).

Ekstraksi Fitur

Penelitian ini menggunakan ekstraksi fitur TF-IDF yang berfungsi untuk mengonversi kata-kata menjadi nilai numerik berdasarkan tingkat kepentingannya. TF-IDF memberikan bobot tinggi pada kata-kata yang sering muncul dalam suatu dokumen tetapi jarang ditemukan di dokumen lain, sehingga informasi yang relevan menjadi lebih penting, dan kata-kata umum diberi bobot rendah. Dengan pendekatan ini, setiap komentar atau teks diubah menjadi fitur yang diproses oleh *machine learning*, sehingga membantu model dalam mengenali pola dan membedakan kategori analisis secara lebih akurat.

Pelabelan Data

Pelabelan data dilakukan untuk menentukan data yang memiliki positif dan data yang memiliki nilai negatif (Darmawan et al., 2022). Daftar kata yang ada disusun dengan cara yang terorganisir untuk mencerminkan kecenderungan makna dalam teks, di mana istilah-istilah yang memiliki konotasi positif seperti baik, mendukung, berhasil, dan positif dipakai sebagai indikator sentimen positif, sedangkan istilah-istilah dengan konotasi negatif seperti buruk, gagal, menolak, dan masalah diterapkan sebagai indikator sentimen negatif. Setiap dokumen teks dievaluasi dengan menghitung seberapa sering kata-kata tersebut muncul, sehingga label sentimen diberikan berdasarkan kategori kata yang paling dominan. Metode ini dipilih karena memberikan proses pelabelan yang rapi, konsisten, dan terukur pada sekumpulan data yang besar, selain itu juga menjadi dasar untuk tahap modeling dan penilaian algoritma machine learning dalam analisis sentimen.

***Split* Data**

Untuk memastikan bahwa algoritma dapat diterapkan pada data baru, proses split data adalah tahapan penting dalam pembangunan model pembelajaran mesin. Pada saat ini, dataset secara keseluruhan dibagi menjadi dua kelompok, data pelatihan dan data pengujian. Model menggunakan data pelatihan untuk mengidentifikasi pola, struktur, dan hubungan antar variabel. Data pengujian, di sisi lain, berfungsi sebagai informasi yang tidak terlibat dalam proses pelatihan dan digunakan untuk mengukur kinerja model secara objektif. Untuk menghindari kesalahan data, pemisahan ini harus dilakukan secara mandiri. Penilaian kinerja secara nyata mencerminkan kapasitas model dalam memprediksi informasi yang belum pernah dikenali sebelumnya. Dalam studi ini, rasio 80:20 digunakan, dengan 80 persen digunakan sebagai data pelatihan dan 20 persen sebagai data pengujian. Diharapkan rasio ini dapat memberikan keseimbangan antara proses pembelajaran model dan evaluasi akurasi (Shafa et al., 2024).

Klasifikasi Model

Data yang telah melalui proses pelabelan dengan baik selanjutnya diuji menggunakan beberapa algoritma untuk memperoleh tingkat akurasi yang paling optimal (Ni Made Tara Okta Adriana et al., 2023). Pada penelitian ini, tiga model algoritma diterapkan. Ketiga algoritma tersebut dipilih karena memiliki karakteristik berbeda dalam proses klasifikasi, sehingga memungkinkan dilakukan pembandingan performa secara komprehensif. Pada tahap ini, masing-masing model dianalisis untuk menilai kemampuan dalam mengenali pola dari data yang tersedia serta mengidentifikasi model yang paling efektif dan akurat dalam menyelesaikan permasalahan penelitian.

Random Forest mengklasifikasikan informasi dengan cara mengenali pola yang terdapat di dalam data (Agtira et al., 2023; Yaqin et al., 2025). *Random Forest* digunakan sebagai model *ensemble* berbasis pengambilan sampel membangun *decision tree* secara bersamaan, dengan meminimalisir variasi prediksi dan tahan gangguan data (Ais & Sanjaya, 2025). *Naïve Bayes* menggunakan perhitungan kemungkinan untuk meramalkan kategori dari suatu data, dengan dasar pada sebaran peluang dari atribut-atributnya (Setiawan & Suryono, 2024; Mulyani & Pajri, 2024). SVM adalah sebuah metode dalam klasifikasi yang membedakan dua kategori dengan menghasilkan *hyperplane* dengan jarak maksimum diantara kedua kategori. SVM beroperasi dengan tujuan untuk mencari hyperplane yang paling optimal yang memperbesar jarak antara titik data yang paling dekat dari dua kategori yang berbeda, yang disebut sebagai *support vectors*. Metode yang berguna untuk mengelola dataset yang dimensi tinggi dan berfungsi baik ketika jumlah fitur melebihi jumlah sampel data (Saputra & Hasan, 2023). *Support Vector Machine,* Perhitungan model SVM terhadap bobot setiap kata dalam analisis sentimen sangat berpengaruh pada kategori negatif dan positif, serta ketepatan yang diperoleh dari proses pemodelan (Arsi & Waluyo, 2021).

Evaluasi Model

Evaluasi terhadap model-model dilakukan dengan tujuan menilai performa tiga algoritma yang diterapkan dalam penelitian ini. Proses evaluasi ini bertujuan untuk mengidentifikasi algoritma yang bisa menghasilkan performa optimal dalam menilai pandangan masyarakat tentang relokasi Ibu Kota Nusantara. Seluruh rangkaian proses penelitian dilaksanakan dengan cara yang sistematis untuk memastikan alur kerja yang teratur dan dapat dipertanggungjawabkan. Hasil evaluasi menunjukkan bahwa masing-masing algoritma menghasilkan kinerja yang berbeda, sehingga analisis ini berfungsi sebagai dasar untuk memilih model yang paling efektif dan cocok digunakan sebagai model utama dalam penelitian ini.

Model dikatakan paling efektif apabila mampu menghasilkan nilai metrik evaluasi tertinggi dan paling seimbang, khususnya pada *F1-score* yang mencerminkan keseimbangan antara kemampuan model dalam mengklasifikasikan secara tepat. Selain itu, stabilitas model dalam memprediksi kelas mayoritas maupun minoritas juga menjadi pertimbangan penting. Dengan demikian, metode yang menunjukkan performa paling unggul secara konsisten pada data uji dibuktikan melalui hasil evaluasi tersebut dipilih sebagai model utama karena dinilai paling mampu merepresentasikan pola opini publik terhadap pemindahan Ibu Kota Nusantara (IKN) secara akurat dan dapat dipertanggungjawabkan secara ilmiah.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dataset yang digunakan merupakan kumpulan data dari *kaggle* terkait pemindahan Ibu Kota Nusantara (IKN). Data yang berhasil dikumpulkan selanjutnya disimpan dalam bentuk tabel berupa kolom teks yang berisi komentar atau pendapat masyarakat. Data mencerminkan beragam respons publik, baik berupa dukungan, kritik, maupun pandangan terhadap kebijakan pemindahan IKN, sehingga dapat digunakan sebagai bahan untuk mengetahui kecenderungan opini masyarakat secara umum. Tabel 1 adalah kumpulan data teks yang diterapkan dalam penelitian ini, sumber data berasal dari Kaggle yang meliputi komentar teks mengenai pemindahan Ibu Kota Nusantara (IKN) dengan total sebanyak 5811 data yang mencakup kolom teks berisi opini komentar mengenai IKN.

Tabel 1. Raw data

	text
0	KELISTRIKAN IKN NUSANTARA Green, Smart &amp B...
1	@BlueNeo81 Kota ku
2	@kumparan Contoh negara2 agearis mcm Vietnam,d...
3	@bana_cotta @deeswast @detikcom Si ibu salah k...
4	Ikan Kakap Ikan Teri Dimakan dengan sambal ter...
5	Basa-basi Pemerintah soal Setop Investor Asing...
6	Makan GRATIS program @prabowo hingga 450 T set...
7	Presiden Jokowi letakan batu pertama pembangun...
8	Presiden Jokowi letakan batu pertama pembangun...
9	Ryeowook ke bandung, ini gaada yg mau tiba2 ke...
…	…
5811	Presiden Instruksikan Pembangunan SDM di IKN

Data Pre-Processing

Selanjutnya, hasil tersebut akan lanjut pada tahap *preprocessing* yang melalui beberapa tahapan yaitu *cleaning text, case folding, tokenizing,* dan *removal stopwords.* Hasil *cleaning text* pada tabel 2, proses pembersihan mencakup penghapusan URL, penghapusan karakter dan simbol yang tidak perlu, penghapusan penyebutan, dan pengubahan semua huruf menjadi huruf kecil, serta kata-kata yang tidak relevan untuk analisis juga dihapus.

Tabel 2. *Cleaning text*

	text	cleaned_text
0	KELISTRIKAN IKN NUSANTARA Green, Smart &amp B...	kelistrikan ikn nusantara green smart amp beau...
1	@BlueNeo81 Kota ku	kota ku
2	@kumparan Contoh negara2 agearis mcm Vietnam,d...	contoh negara agearis mcm vietnamdllbangun per...
3	@bana_cotta @deeswast @detikcom Si ibu salah k...	si ibu salah kamar ini bukan investasi tapi su...
4	Ikan Kakap Ikan Teri Dimakan dengan sambal ter...	ikan kakap ikan teri dimakan dengan sambal ter...
5	Basa-basi Pemerintah soal Setop Investor Asing...	basabasi pemerintah soal setop investor asing ...

Hasil *case folding* pada tabel 3, *kolom text* menunjukkan data asli yang masih mengandung huruf kapital, variasi penulisan, serta unsur-unsur lain seperti URL, mention, dan tanda baca. Setelah diterapkan *case folding*, hasilnya terlihat pada kolom *case_folding*, di mana setiap teks telah berubah menjadi huruf kecil sepenuhnya.

Tabel 3. *Case Folding*

	text	case_folding
0	KELISTRIKAN IKN NUSANTARA Green, Smart &amp B...	kelistrikan ikn nusantara green smart amp beau...
1	@BlueNeo81 Kota ku	kota ku
2	@kumparan Contoh negara2 agearis mcm Vietnam,d...	contoh negara agearis mcm vietnamdllbangun per...
3	@bana_cotta @deeswast @detikcom Si ibu salah k...	si ibu salah kamar ini bukan investasi tapi su...

4	Ikan Kakap Ikan Teri Dimakan dengan sambal ter...	ikan kakap ikan teri dimakan dengan sambal ter...
5	Basa-basi Pemerintah soal Setop Investor Asing...	basabasi pemerintah soal setop investor asing ...

Hasil *Tokenizing* pada tabel 4, *kolom text* kemudian diubah menjadi daftar kata yang lebih terstruktur pada kolom *tokenizing*. Setiap kalimat diuraikan menjadi potongan kata yang dipisahkan dengan tanda koma, sehingga struktur teks menjadi lebih mudah diproses oleh model analisis berikutnya.

Tabel 4. *Tokenizing*

	text	tokenizing
0	KELISTRIKAN IKN NUSANTARA Green, Smart &amp B...	[kelistrikan, ikn, nusantara, green, smart, am...
1	@BlueNeo81 Kota ku	[kota, ku]
2	@kumparan Contoh negara2 agearis mcm Vietnam,d...	[contoh, negara, agearis, mcm, vietnamdllbangu...
3	@bana_cotta @deeswast @detikcom Si ibu salah k...	[si, ibu, salah, kamar, ini, bukan, investasi,...
4	Ikan Kakap Ikan Teri Dimakan dengan sambal ter...	[ikan, kakap, ikan, teri, dimakan, dengan, sam...
5	Basa-basi Pemerintah soal Setop Investor Asing...	[basabasi, pemerintah, soal, setop, investor, ...

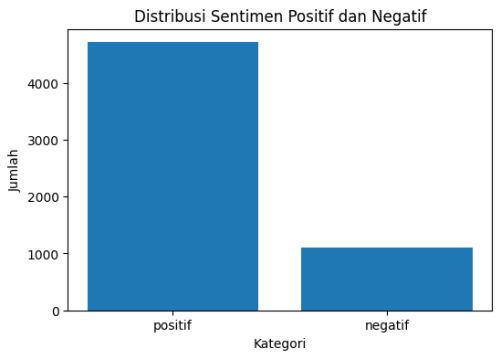
Hasil *Removal Stopwords* pada tabel 5, kolom text memperlihatkan teks asli, sedangkan kolom *stopword_removal* menunjukkan hasil setelah kata-kata yang tidak informatif tersebut dihilangkan. Pada hasil pembersihan, terlihat bahwa teks menjadi lebih ringkas dan hanya menyisakan kata-kata yang memiliki bobot makna lebih kuat terkait analisis sentimen atau topik.

Tabel 5. *Removal Stopwords*

	text	stopword_removal
0	KELISTRIKAN IKN NUSANTARA Green, Smart &amp B...	[kelistrikan, ikn, nusantara, green, smart, am...
1	@BlueNeo81 Kota ku	[kota, ku]
2	@kumparan Contoh negara2 agearis mcm Vietnam,d...	[contoh, negara, agearis, mcm, vietnamdllbangu...
3	@bana_cotta @deeswast @detikcom Si ibu salah k...	[si, salah, kamar, investasi, sumbangan, rakya...
4	Ikan Kakap Ikan Teri Dimakan dengan sambal ter...	[ikan, kakap, ikan, teri, dimakan, sambal, ter...
5	Basa-basi Pemerintah soal Setop Investor Asing...	[basabasi, pemerintah, setop, investor, asing,...

Pelabelan data

Setelah semua tahapan *preprocessing* selanjutnya pelabelan data, Sentimen positif diberikan pada teks yang berisi dukungan, apresiasi, atau pandangan menguntungkan terhadap pemindahan Ibu Kota Nusantara (IKN), sedangkan sentimen negatif diberikan pada teks yang mengandung kritik, penolakan, atau kekhawatiran terhadap kebijakan tersebut.



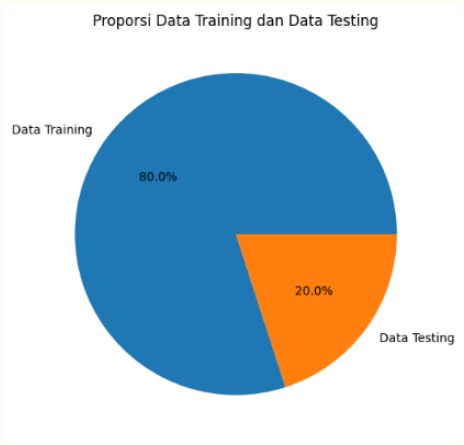
Gambar 2. Distibusi Pelabelan

Pada proses pelabelan pada gambar 2, diperoleh sebanyak 4.708 data sentimen positif dan 1.103 data sentimen negatif yang digunakan pada tahap analisis selanjutnya. Pada tabel 6, text sudah memiliki sentimen positif atau negatif sesuai dengan klasifikasi. Kata-kata positif seperti bagus, baik, mantap, puas, keren, dan bermanfaat digunakan untuk mengidentifikasi opini yang mengandung dukungan atau apresiasi, sedangkan kata-kata negatif seperti buruk, jelek, kecewa, parah, bohong, dan merugikan digunakan untuk menandai teks yang bernada keluhan, kritik, atau ketidakpuasan.

Tabel 6. *Sentimen text*

	cleaned_text	Sentimen
0	kelistrikan ikn nusantara green smart amp beau...	positif
1	kota ku	positif
2	contoh negara agearis mcm vietnamdllbangun per...	positif
3	si ibu salah kamar ini bukan investasi tapi su...	positif
4	ikan kakap ikan teri dimakan dengan sambal ter...	positif
50	gambaran jika indonesia di pimpin oleh anak mu...	negatif
181	semoga aja korupsi atas nama hutang negara ikn...	negatif
183	alasan pks tetap menolak ikn nusantara mardani...	negatif
203	melalui pembangunan ikn nusantara jokowi wujud...	negatif
204	melalui pembangunan ikn jokowi akan wujudkan i...	negatif

***Split* data**



Gambar 3. *Split* data

Pada gambar 3, data dibagi menjadi data pelatihan dan pengujian, yang pertama terdiri dari 4648 data pelatihan (80 persen) dan 1163 data pengujian (20 persen). Seluruh model diuji dengan data pengujian, yang mencakup 1163 data, atau 20% dari total dataset hasil *preprocessing*, yang memiliki 5811 data mentah.

Klasifikasi Model

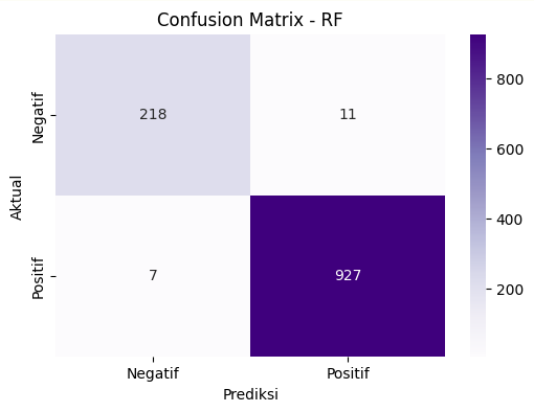
Pelabelan dan split data selesai, selanjutnya dilakukan klasifikasi model menggunakan tiga algoritma sebagai berikut.

1. Random Forest

	precision	recall	f1-score	support
negatif	0.97	0.95	0.96	229
positif	0.99	0.99	0.99	934
accuracy			0.98	1163
macro avg	0.98	0.97	0.98	1163
weighted avg	0.98	0.98	0.98	1163

Gambar 4. Laporan Klasifikasi RF

Pada gambar 4, Algoritma *Random Forest* menunjukkan hasil yang sangat memuaskan, dengan tingkat akurasi mencapai 98,5%. Dari analisis klasifikasi, kategori positif menunjukkan angka *precision* dan *recall* yang sangat tinggi, masing-masing mencapai 0,99, yang menunjukkan bahwa model ini mampu mengenali sentimen positif dengan tingkat kesalahan yang sangat rendah. Untuk kelas negatif, model ini menunjukkan nilai *precision* 0,97 dan *recall* 0,95, yang berarti bahwa mayoritas data negatif terklasifikasi dengan benar walaupun masih terdapat sejumlah kecil kesalahan dalam prediksi. Secara keseluruhan, nilai rata-rata *macro* dan *weighted average* yang keduanya mencapai 0,98 menegaskan bahwa model beroperasi dengan konsisten pada kedua kelas, serta mampu memberikan prediksi yang akurat dan stabil dalam konteks analisis sentimen.



Gambar 5. Confusion Matrix RF

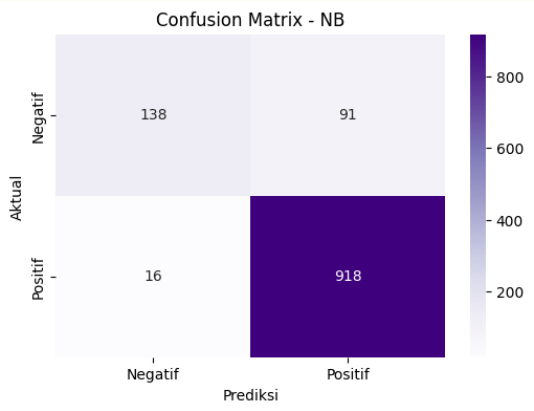
Pada gambar 5, dalam kategori Negatif, model dapat mengklasifikasikan 218 data dengan tepat (*True Negative*) dan hanya melakukan 11 kesalahan dalam mengidentifikasi data negatif sebagai positif (*False Positive*). Sementara itu, pada kategori Positif, hasilnya sangat memuaskan dengan 927 data yang berhasil diidentifikasi secara akurat (*True Positive*) dan hanya ada tujuh data positif yang salah yang dianggap negatif (*False Negative*).

2. Naïve Bayes

	precision	recall	f1-score	support
negatif	0.90	0.60	0.72	229
positif	0.91	0.98	0.94	934
accuracy			0.91	1163
macro avg	0.90	0.79	0.83	1163
weighted avg	0.91	0.91	0.90	1163

Gambar 6. Laporan Klasifikasi NB

Berdasarkan gambar 6, metode *Naive Bayes* berhasil mendapatkan tingkat ketepatan sebesar 90,8% dalam pengkategorian sentimen. Dengan precision 0.91 dan recall 0.98, kategori positif menunjukkan kinerja yang luar biasa, menunjukkan bahwa model dapat mengidentifikasi hampir semua data dengan label positif. Sementara itu, untuk kategori negatif, kinerjanya masih perlu perbaikan dengan *recall* hanya 0.60, yang berarti banyak data negatif yang keliru diklasifikasikan sebagai positif, meskipun *precision* di kategori ini cukup tinggi dengan nilai 0.90. Nilai *macro average f1-score* yang mencapai 0.83 menandakan adanya ketidakseimbangan kinerja antara kategori, sedangkan *weighted average f1-score* 0.90 menunjukkan bahwa secara keseluruhan model tetap memberikan performa yang baik, terutama karena jumlah data positif lebih besar. Oleh karena itu, *Naïve Bayes* berfungsi dengan baik untuk dataset dengan distribusi kelas yang tidak merata.



Gambar 7. Confusion Matrix NB

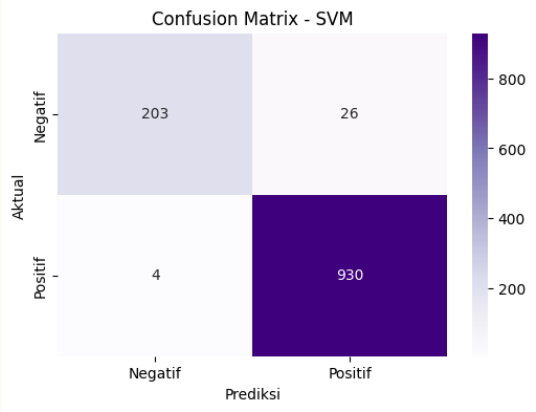
Berdasarkan gambar 7, Model menunjukkan hasil yang sangat memuaskan dalam kategori negatif, menemukan 138 data yang benar (*True Negative*), tetapi 91 data negatif yang salah dianggap positif (*False Negative*). Di sisi lain, model menunjukkan hasil yang sangat memuaskan dalam kategori positif, menemukan 918 data yang akurat (*True Positive*), dan hanya 16 data positif yang salah dianggap negatif (*False Negative*).

3. Support Vector Machine

	precision	recall	f1-score	support
negatif	0.98	0.89	0.93	229
positif	0.97	1.00	0.98	934
accuracy			0.97	1163
macro avg	0.98	0.94	0.96	1163
weighted avg	0.97	0.97	0.97	1163

Gambar 8. Laporan Klasifikasi SVM

Berdasarkan gambar 8, menunjukkan kinerja klasifikasi SVM yang luar biasa, dalam kategori positif, SVM menunjukkan kinerja yang sangat baik dengan *precision* 0.98 dan *f1-score* 0.93, dalam kategori negatif, model hampir sempurna dengan *recall* 1,00 dan *f1-score* 0.98, tetapi *recall* tetap berada pada nilai 0.89, menunjukkan bahwa sebagian kecil data dengan label negatif dapat diklasifikasikan dengan tepat.



Gambar 9. Confusion Matrix SVM

Berdasarkan gambar 9,, Untuk kelompok negatif, model mengidentifikasi 203 data yang akurat (*True Negative*), tetapi 26 diprediksi negatif meskipun sebenarnya positif (*False Negative*). Untuk kelompok positif, SVM memiliki kinerja yang sangat baik dengan mengidentifikasi 930 data positif dengan akurat, dan hanya 4 data positif yang salah dianggap negatif.

Evaluasi model

Pada tabel 7, hasil dari klasifikasi yang dilakukan dengan tiga algoritma menunjukkan bahwa *Random Forest* memiliki akurasi sebesar 98%, diikuti oleh *Naïve Bayes* dengan akurasi 91%, dan *Support Vector Machine* mencapai 97%. Ketiga algoritma yang dievaluasi menunjukkan hasil yang cukup memuaskan, meskipun dengan variasi dalam tingkat akurasi. Algoritma *Random Forest* mendapatkan hasil terbaik dengan akurasi mencapai 98%, yang mengindikasikan kemampuan algoritma ini dalam mengenali pola data lebih efektif melalui proses pengambilan keputusan berbasis banyak pohon (*ensemble*). Model SVM sedikit tertinggal dengan akurasi 97%, menunjukkan bahwa pemisahan data menggunakan *hyperplane* berfungsi dengan baik. Di sisi lain, model *Naïve Bayes* menunjukkan akurasi yang lebih rendah, sekitar 91%. Hasil ini dapat dimengerti karena *Naïve Bayes* mengasumsikan fitur-fitur yang tidak saling bergantung, sehingga dalam dataset dengan adanya korelasi antar variabel, kinerjanya cenderung menurun. Secara keseluruhan, grafik ini menandakan bahwa *Random Forest* adalah algoritma yang paling efektif untuk dataset yang dianalisis dalam penelitian ini.

Tabel 7. Akurasi Model

Model	Akurasi
Random forest	98%
Naïve Bayes	91%
Support Vector Machine	97%

Fungsi *wordcloud* memungkinkan memvisualisasikan data atau menemukan kata-kata yang sering digunakan (Idris et al., 2023).

Gambar 10. *Wordcloud*

KESIMPULAN

Berdasarkan hasil dan pembahasan metode klasifikasi Algoritma *Random Forest, Naïve Bayes*, dan *Support Vector Machine*. Data dibagi menjadi 2 yaitu 80% data pelatihan dan 20% data pengujian. Hasil performa akurasi dalam analisis mendapat nilai akurasi 98% untuk *Random Forest,* akurasi 91% untuk *Naïve Bayes,* dan akurasi 97% untuk *Support Vector Machine.* Dengan demikian, *Random Forest* memiliki nilai akurasi yang paling baik yaitu 98%. Hasil penelitian ini menunjukkan peningkatan kinerja yang signifikan dibandingkan penelitian terdahulu yang memperoleh akurasi sebesar 86,89%. Pada penelitian yang dilakukan saat ini, model mampu mencapai akurasi hingga 98%, yang berarti terjadi peningkatan ketepatan prediksi sebesar lebih dari 11%. Peningkatan ini dapat disebabkan oleh beberapa faktor, seperti teknik preprocessing yang lebih optimal, pemilihan algoritma yang lebih sesuai, serta penggunaan metode ekstraksi fitur yang mampu merepresentasikan teks secara lebih efektif. Perbandingan itu menunjukkan bahwa metode yang digunakan dalam studi ini dapat menghasilkan hasil yang lebih tepat dan konsisten dalam proses pengelompokan sentimen. Saran untuk penelitian berikutnya adalah peneliti diharapkan mampu memanfaatkan lebih banyak variasi data atau menggunakan algoritma yang berbeda. Langkah itu dapat menghasilkan temuan penelitian dan hasil yang berbeda.

DAFTAR PUSTAKA

Agtira, B. H., Handayani, H. H., & Masruriyah, A. F. N. (2023). Perbandingan Algoritma NBC dan Decision Tree pada Sentimen Analisis Mengenai Vaksinasi Covid-19 Di Indonesia. *Remik*, *7*(1), 704–712. https://doi.org/10.33395/remik.v7i1.12151

Ais, S. R., & Sanjaya, U. P. (2025). Perbandingan Algoritma Random Forest, XGBoost, dan Logistic Regression untuk Prediksi Risiko Kekambuhan Kanker Tiroid. *Edumatic: Jurnal Pendidikan Informatika*, *9*(1), 236–245. https://doi.org/10.29408/edumatic.v9i1.29644

Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(1), 147-156. https://doi.org/10.25126/jtiik.0813944

Cesar, W., Saputra, R. R., & Triyono, G. (2024). Perancangan Model Sistem Pendukung Keputusan Untuk Menentukan Formasi CASN Menggunakan Naïve Bayes dan Simple Additive Weighting. *Decode: Jurnal Pendidikan Teknologi Informasi*, *4*(1), 239–250. https://doi.org/10.51454/decode.v4i1.260

Darmawan, R., Indra, I., & Surahmat, A. (2022). Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru di Twitter. *Jurnal Kajian Ilmiah*, *22*(2), 143–152. https://doi.org/10.31599/jki.v22i2.1130

Hadi, N., & Sugiarto, D. (2025). Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes. *Jurnal Informatika: Jurnal Pengembangan IT*, *10*(1), 37–49. https://doi.org/10.30591/jpit.v10i1.7106

Hanifatun, F., & Zahrotun, L. (2025). Penerapan Data Mining Dalam Pemberian Kelayakan Kredit Nasabah Pada Badan Usaha Milik Desa Gedong Gincu Dengan Metode Naïve Bayes. *Jurnal Informatika: Jurnal Pengembangan IT*, *10*(1), 226-236 . https://doi.org/10.30591/jpit.v10i1.5939

Huang, D., & Wasito, I. (2024). Analisis Peran Atlet Dota 2 Dengan Algoritma Random Forest. *Decode: Jurnal Pendidikan Teknologi Informasi*, *4*(1), 107–115. https://doi.org/10.51454/decode.v4i1.216

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*, *7*(1), 126–130. https://dx.doi.org/10.55338/jikomsi.v7i1.2846

Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, *5*(1), 32–35. https://doi.org/10.37905/jjeee.v5i1.16830

Kawani, G. (2019). Implementasi Naive Bayes Untuk Menentukan Wadah Limbah B3 Sesuai Karakteristik. *Journal of Informatics Information System, Software Engineering and Applications*, *1*(2), 73–81. https://doi.org/10.20895/inista.v1i2.73

Marga, N. S., Isnain, A. R., & Alita, D. (2020). Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(4), 453–463. https://doi.org/10.33365/jatika.v2i4.1602

Mulyani, S., & Pajri, A. E. (2024). Klasisifikasi Tingkat Kemiskinan Di Indonesia Menggunakan Algoritma Naive Bayes. *Journal Of Computer Sciences and Informatics*, *1(2),* 53-57. https://doi.org/10.34304/scientific.v1i2.333

Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 1–10. https://doi.org/10.29408/edumatic.v8i1.24114

Safira, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Sistem Informasi*, *5*(1), 59–70. https://doi.org/10.31849/zn.v5i1.12856

Saputra, A., & Hasan, F. N. (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

Sarumpaet, L. H., & Suryono, R. R. (2025). Analisis Sentimen Publik Program PPPK di Media Sosial X menggunakan Naïve Bayes dan SVM. *Edumatic: Jurnal Pendidikan Informatika*, *9*(2), 362–371. https://doi.org/10.29408/edumatic.v9i2.30065

Septianingrum, F., & Irawan, A. S. Y. (2021). Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review. *Jurnal Media Informatika Budidarma*, *5*(3), 799. https://doi.org/10.30865/mib.v5i3.2983

Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 183–192. https://doi.org/10.29408/edumatic.v8i1.25667

Shafa, B., Handayani, H., Lestari, S. A. P., & Cahyana, Y. (2024). Prediksi Kanker Paru dengan Normalisasi menggunakan Perbandingan Algoritma Random Forest, Decision Tree dan Naïve Bayes. *Decode: Jurnal Pendidikan Teknologi Informasi*, *4*(3), 1057–1070. https://doi.org/10.51454/decode.v4i3.779

Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *4*(1), 74–80. https://doi.org/10.33365/jatika.v4i1.2468

Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(3). https://doi.org/10.30591/jpit.v6i3.2870

Yaqin, A. A., Barata, M. A., & Mahmudah, N. (2025). Implementation of the Random Forest Algorithm with Optuna Optimization in Lung Cancer Classification. *Jurnal Sistem Informasi*, *14*, 561–569.