

Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes

Moch. Reinaldy Destra Fachreza ^{(1)*}, Suhartono ⁽²⁾, M. Ainul Yaqin ⁽³⁾

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : 17650053@student.uin-malang.ac.id, {suhartono,yaqinov}@ti.uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 27 Juni 2023, direvisi 12 September 2023, diterima 12 September 2023, dan dipublikasikan 30 September 2023.

Abstract

Some time ago, the House of Representatives passed Law (UU) Number 3 of 2022 concerning the National Capital City on January 18, 2022. Then, President Joko Widodo officially signed the IKN Law on February 15, 2022. Thus, the Indonesian capital will be moved to Penajam Paser Utara Regency and Kutai Kartanegara Regency, East Kalimantan Province. The public's response to the decision varies; many respond with supportive sentiments, but some react with unsupportive ideas. Nowadays, there are many ways to observe information collected on social media. Various responses submitted through social media can be used as sentiment classification research data. The Naïve Bayes method is commonly used for this type of research. Data was collected between February 15-25, 2023, with as many as 500 tweets. This research uses the Gaussian Naïve Bayes type because of the independence assumption made by this method. Features that do not significantly contribute to the classification can be ignored, thus reducing the impact of irrelevant features. This study aims to measure public sentiment on Twitter towards the process of moving the nation's capital. The system created provides the best trial results at 80% feature usage with 82.0% accuracy, 76.9% precision, and 100% recall.

Keywords: IKN, Naïve Bayes, Sentiments, Classification, Twitter

Abstrak

Beberapa waktu lalu, DPR mengesahkan Undang-Undang (UU) Nomor 3 Tahun 2022 tentang IKN pada 18 Januari 2022. Kemudian Presiden Joko Widodo resmi telah menandatangani UU IKN pada 15 Februari 2022. Dengan demikian, ibu kota Indonesia IKN akan dipindahkan ke Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Respon masyarakat terkait keputusan tersebut beraneka ragam, banyak yang menanggapi dengan sentimen mendukung, namun ada pula yang menanggapi dengan sentimen tidak mendukung. Dewasa ini, banyak cara untuk mengamati informasi yang terhimpun pada media sosial. Berbagai tanggapan yang disampaikan melalui media sosial dapat digunakan sebagai data penelitian klasifikasi sentimen. Metode Naïve Bayes lumrah digunakan untuk penelitian jenis ini. Pengumpulan data dilakukan antara tanggal 15-25 Februari 2023 sebanyak 500 tweets. Penelitian ini menggunakan jenis Gaussian Naïve Bayes karena asumsi independensi yang dilakukan oleh metode ini. Fitur-fitur yang tidak memberikan kontribusi signifikan pada klasifikasi dapat diabaikan, sehingga mengurangi dampak dari fitur-fitur yang tidak relevan. Penelitian ini bertujuan untuk mengukur sentimen masyarakat di media sosial Twitter terhadap proses pemindahan ibu kota negara. Sistem yang dibuat memberikan hasil uji coba terbaik pada penggunaan fitur sebanyak 80% dengan *accuracy* 82,0%, *precision* 76,9%, dan *recall* 100%.

Kata Kunci: IKN, Naïve Bayes, Sentimen, Klasifikasi, Twitter

1. PENDAHULUAN

Ibu kota adalah wilayah letak lokasi pusat pemerintahan suatu negara, tempat himpunan segala administratif, yaitu lembaga eksekutif, lembaga legislatif, dan lembaga yudikatif (Tim Pustaka Phoenix, 2013). Ibu kota negara diambil dari bahasa latin *caput* yang bermakna kepala (*head*), dan terpaut dengan tata *capitol* yang berarti pusat pokok pemerintahan. Jakarta merupakan Ibu



Kota Indonesia, mempunyai kedudukan vital sebagai inti sebuah negara dan pusat pemerintahan yang diatur dalam Undang-Undang. Jakarta sebagai ibu kota pastinya berbeda dari daerah lainnya meskipun mempunyai kewenangan yang serupa sebagai sebuah provinsi. Kota Jakarta sebagai ibu kota memiliki kedudukan eksklusif sebagai ibu kota, yang mempunyai tugas sebagai tempat penguasaan politik, ekonomi, pembangunan, dan bidang krusial lainnya kepada tiap-tiap provinsi yang lain. Jakarta dijadikan sebagai ibu kota negara diatur dalam Undang-Undang Republik Indonesia Nomor 10 tahun 1964 Tentang Pernyataan Daerah Khusus Ibukota Jakarta Raya Tetap Sebagai Ibu Kota Negera Republik Indonesia (Hutasoit, 2019).

DPR mengesahkan Undang-Undang (UU) Nomor 3 Tahun 2022 pada 18 Januari 2022 tentang IKN, lalu Presiden Joko Widodo telah menandatangani secara resmi UU tentang IKN pada tanggal 15 Februari 2022. Dengan demikian, ibu kota Indonesia IKN direlokasikan ke Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur (Hadi & Ristawati, 2020). Barometer dipilihnya Kalimantan Timur untuk dibangun IKN yang baru adalah akses lokasi yang mudah, lebih dekat dengan kota besar disekitar wilayah tersebut, penduduk yang terbuka dan heterogen, rendahnya potensi terjadinya konflik, didukung oleh tri matra darat, laut, dan udara, infrastruktur yang mumpuni, air alami didapatkan dari 3 waduk, 2 waduk yang diproyeksikan akan dibangun, 4 sungai strategis, dan 4 daerah aliran sungai. (Saraswati & Adi, 2022).

Keputusan pemindahan ibu kota negara oleh pemerintah menimbulkan perbincangan dalam masyarakat, khususnya pada sosial media. Media sosial yang memiliki banyak pengguna khususnya di Indonesia adalah Twitter. Krisdiyanto (2021) menjelaskan bahwa Twitter merupakan media sosial yang dikembangkan oleh Jack Dorsey yang berfungsi untuk mengunggah pesan yang dikenal sebagai kicauan (*tweet*). Twitter digunakan oleh kalangan masyarakat sebanyak 59% dari total media sosial yang diakses serta menjadi peringkat ke-5 media sosial yang sering diakses pada tahun 2020. Maka dari itu, Twitter menjadi platform media sosial yang berpengaruh bagi para pengguna aktif di Indonesia. Menurut Zuhdi et al. (2019), data yang dihimpun pada Twitter dapat bermanfaat jika dianalisis dikarenakan data yang diolah dari Twitter dapat dimanfaatkan sebagai informasi penting melalui *opinion mining*. Kumpulan *tweet* yang ada pada Twitter menjadikannya sumber data dalam bentuk opini, maka diperlukan teknik *text mining* untuk mengumpulkan data.

Text mining adalah teknik yang dimanfaatkan sebagai proses klasifikasi, *text mining* adalah ragam *data mining* yang berfokus untuk mencari pola menarik dari kumpulan data bersifat tekstual dalam jumlah yang besar (Ratniasih et al., 2017). Data dari *tweet* yang berupa teks kemudian diolah menggunakan klasifikasi. Salah satu Algoritma klasifikasi adalah Naïve Bayes. Apabila disejajarkan dengan metode klasifikasi lainnya, Naïve Bayes dapat mengolah data yang cukup sedikit untuk proses latihnya. Performa data latih yang dilakukan pada informasi yang dikumpulkan relatif cepat, menggunakan sedikit *storage* pada proses latih, dan klasifikasi yang cenderung ringan untuk digunakan serta mempunyai lebih sedikit kriteria dari metode lain seperti Jaringan Neural dan *Support Vector Machine* (Al-Aidaros et al., 2010).

Herdhianto (2020) melakukan penelitian *sentiment analysis* menggunakan *naïve bayes classifier* pada *tweet* tentang zakat. Studi tersebut menggunakan seleksi fitur *term-frequency* dan metode *lexicon based* di mana hasil studi menunjukkan bahwa metode Naïve Bayes dan *lexicon based* dengan menggunakan seleksi fitur *term-frequency* memiliki tingkat akurasi sebesar 74% dikarenakan jumlah sentiment positif yang lebih dominan. Hal tersebut terjadi karena terdapat ketidakseimbangan pada data pada kamus *lexicon* positif yang lebih tinggi daripada kamus *lexicon* negatif yang cenderung lebih sedikit.

Ashraf et al. (2018) melakukan penelitian analisis sentiment opini publik terhadap pekan olahraga nasional pada Instagram menggunakan metode Naïve Bayes *classifier*. Studi tersebut menggunakan pembagian data sebesar 90% data latih dan 10% data uji. Hasil yang didapatkan adalah akurasi sebesar 75% dengan total data sebesar 2.000 dokumen menunjukkan hasil negatif yang lebih tinggi karena data uji yang cenderung mengarah ke konotasi negatif.



Klasifikasi sentimen pemindahan ibu kota Indonesia pada media sosial Twitter telah dieksplorasi dalam penelitian. Penelitian yang dilakukan oleh Sutoyo & Almaarif, (2020) berfokus pada analisis sentimen pemindahan ibu kota Indonesia dengan menggunakan data Twitter. Sutoyo & Almaarif (2020) menerapkan teknik analisis sentimen menggunakan metode Naïve Bayes untuk klasifikasi sentimen. Penelitian ini bertujuan untuk mengetahui polaritas sentimen dalam tweet tersebut, baik positif maupun negatif, terkait pemindahan ibu kota.

Berdasarkan hasil pengamatan, Naïve Bayes dan penggunaan seleksi fitur mampu memberikan hasil akurasi yang baik mengenai klasifikasi sentimen pada media sosial dengan jumlah data yang cukup. Meskipun penelitian tentang metode Naïve Bayes untuk klasifikasi sentimen cukup menghasilkan *output* yang baik, masih terdapat kekurangan yang dibutuhkan untuk meningkatkan model prediksi yang meliputi proses *preprocessing* yang mempengaruhi tingkat kesesuaian data dalam kelas pada model serta penggunaan seleksi fitur lebih lanjut dalam sistem klasifikasi sentimen. Oleh karena itu, penelitian lebih lanjut diperlukan dalam membangun model menggunakan metode Naïve Bayes dan seleksi fitur untuk mengklasifikasikan sentimen. Hal tersebut menjadi landasan dalam penelitian ini dengan tujuan untuk mengukur sentiment masyarakat terhadap proses pemindahan ibu kota negara Indonesia pada media sosial Twitter menggunakan metode Naïve Bayes.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan sebagai objek penelitian dikumpulkan dari sosial media Twitter. Data objek penelitian ini adalah tweet dari pengguna aktif Twitter. *Tweet* yang mengandung opini atau komentar warga negara Indonesia dikumpulkan menggunakan teknik *scraping*. Data dikumpulkan dengan menggunakan kata kunci “ibu kota pindah, ibukotanegara, pemindahan ibukota” serta tagar #ikn, dan #ibukotabaru yang menggunakan bahasa Indonesia. Keseluruhan data yang telah dikumpulkan melalui proses *scraping* pada periode tanggal 15 Februari 2023 hingga 25 Februari 2023 sebanyak 500 *tweets*. Data yang berhasil dikumpulkan kemudian disimpan menjadi *output* di dalam berkas .csv. Data yang telah disimpan kemudian diolah dengan tahapan pelabelan yang bertujuan menentukan label pada tiap kelas tweet yang telah melalui proses *scraping*. Pada tahap ini, data yang telah dikumpulkan melalui proses *scraping* akan dilakukan pelabelan secara manual. Tidak semua data akan melewati tahap pelabelan, namun dilakukan pemilihan sampel secara acak untuk kemudian diberi label. Metode pelabelan ini melibatkan tinjauan opini secara individu yang dilakukan oleh ahli, di mana opini tersebut akan dikategorikan sebagai sentimen positif atau negatif.

Tabel 1 Pelabelan Manual

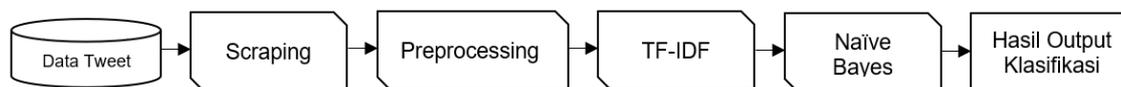
Dokumen	<i>Tweet</i>	Label
Dokumen 1	@dragon3mpro @maryshelparaiso warga dukung kota masa depan ikn #IbukotaNegara	Mendukung
Dokumen 2	@edyaldo1 @maryshelparaiso ih keren bgt desain IKN ini #IbukotaNegar	Mendukung
Dokumen 3	IKN MANGKRAK Menghamburkan Uang Negara, Kereta Cepat Yang Katanya Tidak Menggunakan APBN dll Apabila Di Usut Tuntas.	Tidak Mendukung

2.2 Desain Sistem

Rancangan pada sistem diawali melalui tahap pencarian data pada Twitter dengan menggunakan kata kunci yang telah ditentukan, kemudian data dikumpulkan menggunakan teknik *scraping*. Setelah data disimpan, tahap berikutnya adalah proses *preprocessing* dan proses TF-IDF. Setelah melalui tahap pengolahan data, kemudian data tersebut melalui proses klasifikasi dengan menggunakan Naïve Bayes untuk selanjutnya menuju tahap akhir berupa hasil klasifikasi.



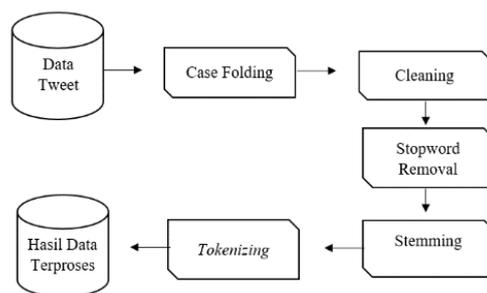
Rancangan pada sistem diawali melalui tahap pencarian data pada Twitter dengan menggunakan kata kunci yang telah ditentukan, kemudian data dikumpulkan menggunakan teknik *scraping*. Setelah data disimpan, tahap berikutnya adalah proses *preprocessing* dan proses TF-IDF. Setelah melalui tahap pengolahan data, kemudian data tersebut melalui proses klasifikasi dengan menggunakan Naïve Bayes untuk selanjutnya menuju tahap akhir berupa hasil klasifikasi.



Gambar 1 Desain Sistem (Sutoyo & Almaarif, 2020)

2.3 Preprocessing

Preprocessing merupakan proses untuk mengubah sebuah suatu dokumen abstrak menjadi terstruktur dengan cara menghilangkan atribut yang tidak dibutuhkan sehingga data menjadi sistematis dan mengurangi *noise*. Langkah pertama dalam tahapan *preprocessing* adalah melakukan proses *case folding*. *Case folding* merupakan proses untuk mengubah semua karakter dalam dokumen menjadi huruf kecil (*lowercase*). *Case folding* bertujuan untuk menyeragamkan karakter yang tidak sesuai. Alur kerja *case folding* dimulai dengan memasukkan data awal kemudian diproses dengan *case folding* sehingga menghasilkan data hasil yang diinginkan. Berikutnya, proses *cleaning* yang merupakan proses untuk menghilangkan karakter yang tidak relevan dalam klasifikasi serta tidak ada kaitannya dengan hasil klasifikasi serta tidak diperlukan dalam tahapan klasifikasi. Langkah selanjutnya *stopword removal*, yakni langkah untuk menghilangkan kata-kata yang sering muncul dalam jumlah yang besar pada dokumen namun tidak memiliki makna penting. Penghapusan kata-kata ini tidak mempengaruhi hasil klasifikasi. Tujuan *stopword removal* untuk mempermudah proses klasifikasi. Langkah setelahnya yakni *stemming* yang merupakan langkah dalam mengembalikan kata-kata dalam data menjadi bentuk kata dasarnya. Proses ini diperlukan untuk menghapus sufiks, prefiks, dan konfiks dalam teks Bahasa Indonesia yang dapat mengubah kata dasar menjadi bentuk yang berbeda. Tahap *tokenizing* merupakan proses yang membagi kata pada kalimat menjadi kata terpisah, sehingga terdapat segmentasi kata yang membentuk kalimat. Kata-kata yang telah dipisahkan tersebut disebut sebagai token. Token-token yang telah dibuat selanjutnya akan digunakan untuk proses pembobotan.



Gambar 2 Alur Preprocessing (Sutoyo & Almaarif, 2020)

2.4 TF - IDF

TF-IDF merupakan sebuah algoritma statistik numerik yang menghasilkan bobot nilai untuk setiap atribut dalam sebuah dokumen teks (Ogura et al., 2022). Metode ini bertujuan untuk menentukan bobot kata dalam teks dengan mengukur tingkat kepentingan suatu kata dalam kumpulan dokumen. Tingkat kepentingan kata tersebut meningkat sejalan dengan jumlah total kemunculannya dalam dokumen teks. Langkah selanjutnya dalam penelitian ini adalah tahap seleksi fitur, di mana penulis menggunakan algoritma TF-IDF seperti yang dituliskan pada Pers. (3). TF yang dituliskan pada Pers. (1), merupakan jumlah total kata yang muncul dalam sebuah



dokumen. Di mana $f_{i,j}$ dalam persamaan tersebut merupakan jumlah eksistensi (i) *term* pada dokumen (j) dan $\sum_{\kappa} f_{i,j}$ adalah jumlah (i) *term* yang terdapat dalam dokumen (j). Besar kecilnya nilai TF tergantung pada seberapa sering kata tersebut muncul dalam dokumen. Sedangkan IDF yang dituliskan pada Pers. (2), merupakan perhitungan seberapa banyak sebuah kata didistribusikan dalam semua dokumen yang tersedia. Di mana N dalam persamaan tersebut merupakan jumlah dokumen dan n_i adalah jumlah dokumen (i) *term* atau kata tersebut. Nilai IDF akan semakin tinggi jika frekuensi kemunculan kata semakin sedikit, dan nilainya berbanding terbalik dengan TF (Ogura et al., 2022).

$$TF_{ij} = \frac{f_{ij}}{\sum_{\kappa} f_{i,j}} \quad (1)$$

$$IDF_i = \log\left(\frac{N}{n_i}\right) + 1 \quad (2)$$

$$TFIDF_{ij} = TF_{ij} \times IDF_i \quad (3)$$

2.5 Naïve Bayes

Langkah berikutnya, yang merupakan langkah inti dari penelitian ini, adalah langkah klasifikasi. Dalam penelitian ini, digunakan algoritma Naïve Bayes *classifier* sebagai algoritma utama untuk melakukan pengklasifikasian. Sistem dibangun dengan pendekatan statistik yang disebut teori Bayes yang menggunakan probabilitas bersyarat. Menurut Suhartono et al. (2018), probabilitas bersyarat adalah perhitungan probabilitas suatu peristiwa, A, ketika peristiwa lain, B, telah terjadi, yang dicatat sebagai $P(A|B)$ yang menggabungkan probabilitas A dan B. Keunggulan utama metode ini terletak pada asumsi yang kuat tentang ketidakterikatan antar variabel. Proses pengklasifikasian melibatkan dua tahap, yaitu pelatihan data dan pengujian data. Tahap pelatihan digunakan untuk memperkenalkan dan melatih *dataset* pada sistem, sedangkan tahap pengujian digunakan untuk menguji *dataset* dengan mengacu pada perhitungan probabilitas yang diperoleh dari tahap pelatihan. Secara umum, teorema Bayes direpresentasikan dengan Pers. (4) (Suhartono et al., 2018). Di mana persamaan $P(A|B)$ merupakan probabilitas *posterior*, $P(B|A)$ adalah estimasi probabilitas, $P(A)$ yaitu probabilitas prior dari kelas, dan $P(B)$ adalah probabilitas *prior* dari *predictor*.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (4)$$

Sistem menggunakan algoritma Gaussian Naïve Bayes, yang sering digunakan untuk mengklasifikasikan data dengan atribut yang bersifat kontinu. Algoritma ini mengandalkan distribusi normal Gaussian sebagai fungsi dasarnya. Algoritma ini memiliki dua parameter utama, yaitu *mean* (μ) dan *variants* (σ). *Mean* adalah nilai rata-rata dari atribut kontinu dengan rumus yang ditunjukkan pada Pers. (5). *variants* adalah kuadrat dari simpangan nilai data pada atribut kontinu dengan rumus yang ditunjukkan pada Pers. (6) (Jahromi & Taheri, 2017). Algoritma ini memiliki ciri khas dalam menggunakan parameter-parameter tersebut. Pada persamaan *mean* (μ), $\in X$ merupakan total nilai TF-IDF dan N menunjukkan total data. Sedangkan pada persamaan *variants* (σ), X adalah nilai data TF-IDF, μ nilai rata-rata data TF-IDF dan N merupakan jumlah data.

$$\mu = \frac{\in X}{N} \quad (5)$$

$$\sigma = \frac{\in (X - \mu)^2}{N - 1} \quad (6)$$

Apabila menghadapi data kontinu, diasumsikan secara umum bahwa nilai kontinu yang berkorelasi dengan setiap kelas memiliki distribusi Gaussian. Perhitungan pada data pelatihan dilakukan dengan memisahkan data berdasarkan kelas yang ada, dan kemudian dihitung nilai rata-rata dan standar deviasi untuk setiap kelas. Selanjutnya, dilakukan proses untuk menghitung



probabilitas dari himpunan data kontinu menggunakan Pers. (7). Di mana $P(B|A)$ adalah estimasi probabilitas, σ menunjukkan variants dari TF-IDF, π memiliki nilai sebesar 3,14, dan μ merupakan nilai rata-rata data TF-IDF.

$$P(B|A) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \quad (7)$$

2.6 Skenario Uji

Penentuan data dilakukan dengan cara mengambil tweet menggunakan teknik *scraping* (You et al., 2021). Data *tweet* berjumlah 500 data kemudian diolah dan diberi label oleh ahli sehingga menghasilkan 319 jumlah *tweet* berlabel mendukung dan 181 jumlah *tweet* berlabel tidak mendukung. Pada langkah ini, *dataset* yang tersedia dipecah menjadi dua bagian. Bagian pertama adalah data latih, yang digunakan untuk melatih model dan mengajarkan pola-pola sentimen yang terdapat dalam teks. Bagian kedua adalah data uji, yang digunakan untuk menguji seberapa baik model dapat memprediksi sentimen dengan tingkat akurasi yang tinggi (You et al., 2021). Pembagian data uji dan data latih dilakukan berulang kali hingga menemukan akurasi terbaiknya.

Tabel 2 Hasil Akurasi Tiap Persentase Dokumen Latih

Persentase Dokumen Latih	Akurasi	Akurasi (%)
10%	0,8	80%
20%	0,64	64%
30%	0,71	71%
40%	0,7	70%
50%	0,72	72%
60%	0,73	73%
70%	0,77	77%
80%	0,74	74%
90%	0,74	74%

Dari tiap poin persentase tersebut ditemukan bahwa hasil akurasi maksimal terdapat pada jumlah persentase dokumen uji dari total dokumen keseluruhan yang berjumlah 10% dari total 500 dokumen. Sehingga dalam penelitian ini, menggunakan perbandingan 90:10 yang artinya 90% merupakan dokumen latih dan 10% merupakan dokumen uji. Untuk meningkatkan akurasi dalam melakukan prediksi klasifikasi, cara yang diterapkan yaitu mengatur penggunaan jumlah fitur. Pengaturan jumlah fitur ini membantu memperbaiki keakuratan model klasifikasi dan mengurangi risiko adanya *overfitting* pada data. Dengan demikian, penggunaan fitur-fitur yang tepat dapat memperkuat kemampuan model dalam mengklasifikasikan data secara lebih efektif dan akurat. Maka dari itu diperlukan pengaturan jumlah fitur yang dimuat dalam prediksi klasifikasi.

2.7 Evaluasi Model

Setelah memperoleh hasil dari data yang telah diklasifikasikan, langkah berikutnya adalah menghitung total akurasi dalam penerapan klasifikasi menggunakan metode Naïve Bayes. Hasil pengujian skenario akan digunakan untuk menghitung nilai akurasi, presisi, dan *recall* menggunakan Matriks Konfusi (*Confusion Matrix*) (Sivakumari et al., 2009). Perhitungan jumlah akurasi, presisi, dan *recall* dijalankan dengan mengacu pada Pers. (8), (9), dan (10).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} 100\% \quad (8)$$

$$Precision = \frac{TP}{TP + FP} 100\% \quad (9)$$

$$Recall = \frac{TP}{TP + FN} 100\% \quad (10)$$



Tabel 3 Keterangan *Confusion Matrix*

Keterangan	Definisi
<i>True Positive</i> (TP)	Banyaknya dokumen konkret berlabel labelmendukung (1) lalu hasil dari prediksinya juga berlabel mendukung (1) atau data positif yangterprediksi positif.
<i>True Negative</i> (TN)	Banyaknya dokumen konkret berlabel tidak mendukung (0) namun hasil prediksinya berlabel tidak mendukung (0) atau data negatif yang terprediksi negatif.
<i>False Positive</i> (FP)	Banyaknya dokumen konkret berlabel tidak mendukung (0) namun hasil prediksinya berlabel mendukung (1) atau data negatif yang terprediksi positif.
<i>False Negative</i> (FN)	Banyaknya dokumen konkret berlabel mendukung (1) namun hasil prediksinya berlabel negatif (0). Atau data positif yang terprediksi negatif.

3. HASIL DAN PEMBAHASAN

Selanjutnya, dalam pengujian sistem, dilakukan pembagian *dataset* menjadi dua bagian, yaitu data latih dan data uji. Proporsi pembagiannya adalah 10% dari *dataset* untuk data latih dan 90% untuk data uji. Selain itu, *dataset* juga dibagi berdasarkan persentase penggunaan fitur yang akan dijelaskan pada poin berikutnya. Tujuan dari tahap ini adalah untuk menentukan jumlah data yang akan digunakan dalam proses pelatihan dan pengujian, sehingga memastikan adanya data yang cukup untuk kedua proses tersebut. Pembagian 90% dan 10% dari total dokumen dipilih karena memiliki nilai akurasi tertinggi pada skenario pengujian. Setelah itu, tindakan berikutnya adalah menentukan persentase penggunaan fitur yang akan digunakan dalam pengujian, dengan variasi persentase yang berbeda dari total fitur yang ada. Proses pembagian ini bertujuan untuk menguji tingkat akurasi sistem secara lebih komprehensif melalui beberapa variasi pengujian.

Langkah awal dalam proses pelatihan data dimulai dengan membagi *dataset* berdasarkan jumlah kelas yang diinginkan. Dalam pengujian ini, *dataset* dibagi menjadi dua kelas, yaitu "mendukung" dan "tidak mendukung". Setelah *dataset* terbagi dalam kedua kelas tersebut, langkah berikutnya adalah mengolah nilai rata-rata (*mean*) dan standar deviasi pada setiap kata yang terdapat dalam *dataset*.

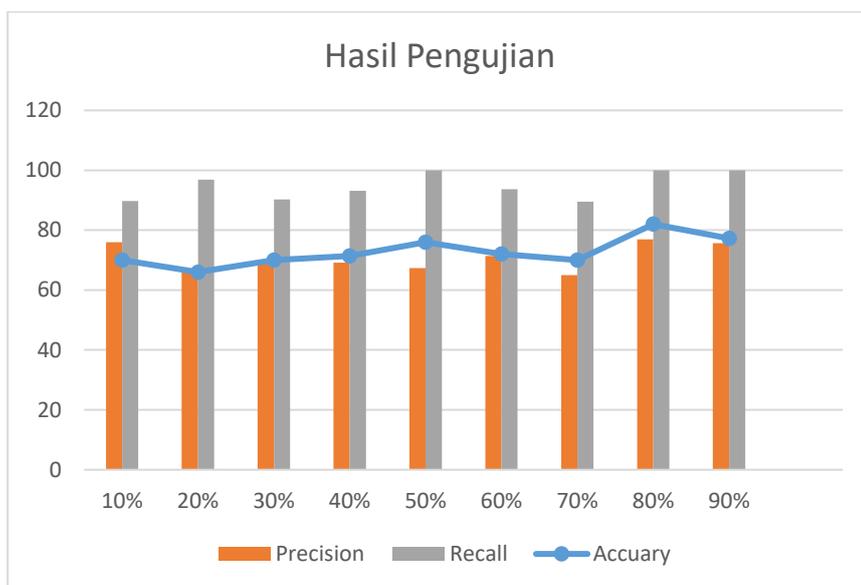
Pengujian data merupakan fase terakhir dari algoritma yang diterapkan dalam sistem. Pada tahap ini, dilakukan perhitungan untuk mendapatkan hasil klasifikasi dari dokumen uji menerapkan persamaan dasar probabilitas Gaussian Naïve Bayes sebagaimana disajikan dalam Pers. (7). Variabel x mewakili bobot data uji, sementara nilai rata-rata (*mean*) dan *variants* diambil dari perhitungan pada data pelatihan. Selanjutnya, yang diperoleh untuk setiap kata dikalikan dengan nilai probabilitas *prior* dari masing-masing kelas untuk memperoleh probabilitas hasil pada setiap dokumen dalam setiap kelas. Hasil probabilitas tertinggi kemudian dijadikan sebagai klasifikasi akhir dari dokumen tersebut.

Melalui proses uji yang dijalankan dengan variasi persentase fitur yang telah disebutkan, diperoleh hasil yang dapat mengindikasikan keakuratan sistem terhadap *class* konkret yang sudah ditentukan. Hasil tersebut disajikan berdasarkan setiap persentase total fitur. Dengan menggunakan data ini, dapat dievaluasi sejauh mana keselarasan dari kelas prediksi dengan kelas konkret, yang memungkinkan dilakukannya perhitungan untuk mengevaluasi kemampuan sistem melalui *confusion matrix*. Matriks kebingungan digunakan untuk mengukur sejauh mana klasifikasi yang dilakukan oleh sistem sesuai dengan kelas konkret, dengan memberikan informasi tentang total *true positive*, *true negative*, *false positive*, dan *false negative*. Oleh sebab itu, penggunaan metode Gaussian Naïve Bayes dengan jumlah fitur yang telah dioptimalkan dapat memberikan hasil yang akurat dan dapat dievaluasi secara objektif menggunakan *confusion matrix* untuk mengukur kemampuan sistem.



Tabel 4 Hasil Klasifikasi Tiap Persentase Penggunaan Fitur

No.	Persentase	TP	TN	FP	FN	Akurasi	Presisi	Recall
1	10%	35	0	11	4	70,0%	76,0%	89,7%
2	20%	32	1	16	1	66,0%	66,6%	96,9%
3	30%	28	7	12	3	70,0 %	70,0%	90,3%
4	40%	27	8	12	2	71,4 %	69,2%	93,1%
5	50%	31	4	15	0	76 %	67,3%	100%
6	60%	30	6	12	2	72 %	71,4%	93,7%
7	70%	26	6	14	4	70,0 %	65,0%	89,5%
8	80%	30	11	9	0	82,0 %	76,9%	100%
9	90%	25	8	17	0	77,3 %	75,7%	100%



Gambar 3 Visualisasi Hasil Uji

Berdasarkan hasil evaluasi sistem dalam melakukan klasifikasi sentiment masyarakat terhadap proses pemindahan Ibu Kota Negara (IKN) Indonesia pada media sosial twitter menggunakan metode Naïve Bayes menunjukkan hasil terbaik pada nilai *accuracy* sebesar 82%, *precision* sebesar 76,9%, nilai *recall* sebesar 100%. Nilai *accuracy* 82% menunjukkan sistem dapat melakukan klasifikasi dengan baik.

4. KESIMPULAN

Peneliti telah membangun sistem klasifikasi sentiment masyarakat terhadap proses pemindahan Ibu Kota Negara (IKN) Indonesia pada media sosial Twitter menggunakan metode Naïve Bayes. Pengujian pada skenario kedua menggunakan total dokumen sebanyak 500 dokumen. Pengujian dilakukan untuk mengetahui proporsi pembagian total dokumen latih serta total dokumen uji yang terbaik. *Output* pengujian kedua menunjukkan proporsi terbaik terdapat pada 10:90 atau 90% dokumen yang dilatih dan 10% dokumen yang diuji mencapai hasil nilai akurasi sebanyak 80%. Kesimpulan berdasarkan nilai *accuracy* sebesar 82% menyatakan model prediksi dapat melakukan klasifikasi dengan baik sehingga penggunaan metode Naïve Bayes dapat diterima dan dapat digunakan untuk mengklasifikasi sentiment masyarakat terhadap proses pemindahan Ibu Kota Negara (IKN) Indonesia. Saran untuk penelitian selanjutnya dapat menggunakan tambahan metode optimasi seleksi fitur lainnya untuk menghasilkan nilai akurasi yang lebih optimal seperti *chi-square*, *information gain*, dan lain-lain. Hal ini dapat membantu meningkatkan performa sistem dalam memilih fitur-fitur yang paling relevan untuk klasifikasi.



DAFTAR PUSTAKA

- Al-Aidarroos, K. M., Bakar, A. A., & Othman, Z. (2010). Naïve Bayes variants in classification learning. *2010 International Conference on Information Retrieval & Knowledge Management (CAMP)*, 276–281. <https://doi.org/10.1109/INFRKM.2010.5466902>
- Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 44(1), 122–136. https://www.asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/4170
- Hadi, F., & Ristawati, R. (2020). Pemindahan Ibu Kota Indonesia dan Kekuasaan Presiden dalam Perspektif Konstitusi. *Jurnal Konstitusi*, 17(3), 530–557. <https://doi.org/10.31078/jk1734>
- Herdhianto, A. (2020). *Sentiment analysis menggunakan Naïve Bayes Classifier (NBC) pada tweet tentang zakat* [Universitas Islam Negeri Syarif Hidayatullah Jakarta]. <https://repository.uinjkt.ac.id/dspace/handle/123456789/53661>
- Hutasoit, W. L. (2019). Analisa Pemindahan Ibukota Negara. *DEDIKASI: Jurnal Ilmiah Sosial, Hukum, Budaya*, 39(2), 108–128. <https://doi.org/10.31293/DDK.V39I2.3989>
- Jahromi, A. H., & Taheri, M. (2017). A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, 209–212. <https://doi.org/10.1109/AISP.2017.8324083>
- Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945>
- Ogura, H., Hanada, Y., Amano, H., & Kondo, M. (2022). A stochastic model of word occurrences in hierarchically structured written texts. *SN Applied Sciences*, 4(3), 77. <https://doi.org/10.1007/s42452-022-04953-w>
- Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017). Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat. *Majalah Ilmiah Teknologi Elektro*, 16(3), 13. <https://doi.org/10.24843/MITE.2017.v16i03p03>
- Saraswati, M. K., & Adi, E. A. W. (2022). Pemindahan Ibu Kota Negara Ke Provinsi Kalimantan Timur Berdasarkan Analisis SWOT. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 6(2). <https://doi.org/10.58258/jisip.v6i2.3086>
- Sivakumari, S., Priyadarsini, R. P., & Amudha, P. (2009). Accuracy Evaluation of C4.5 and Naive Bayes Classifiers Using Attribute Ranking Method. *International Journal of Computational Intelligence Systems*, 2(1), 60. <https://doi.org/10.2991/jnmp.2009.2.1.7>
- Suhartono, S., Kurniawan, F., & Imran, B. (2018). Identification of virtual plants using bayesian networks based on parametric L-system. *International Journal of Advances in Intelligent Informatics*, 4(1), 40. <https://doi.org/10.26555/ijain.v4i1.157>
- Sutoyo, E., & Almaarif, A. (2020). Twitter sentiment analysis of the relocation of Indonesia's capital city. *Bulletin of Electrical Engineering and Informatics*, 9(4), 1620–1630. <https://doi.org/10.11591/eei.v9i4.2352>
- Tim Pustaka Phoenix. (2013). *Kamus Besar Bahasa Indonesia* (S. Wahyuni, Ed.). Pustaka Phoenix.
- You, J., Lee, J., & Kwon, H.-Y. (2021). A Complete and Fast Scraping Method for Collecting Tweets. *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 24–27. <https://doi.org/10.1109/BigComp51126.2021.00014>
- Zuhdi, A. M., Utami, E., & Raharjo, S. (2019). Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 dengan Metode K-NN. *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, 5(2), 1–7. <https://doi.org/10.46808/INFORMA.V5I2.73>

