



Analysis of the Effectiveness of the Naïve Bayes Method on Public Opinion Sentiment in X Regarding the Relocation of the Indonesian Capital City (IKN)

¹Mukhtar*, ²Nuridin, ³Taufiq

¹²³Program Studi Magister Teknologi Informasi, Fakultas Teknik, Universitas Malikussaleh
Jl. Batam, Bukit Indah - Lhokseumawe, Aceh, Indonesia

*e-mail: mukhtaranas@ugp.ac.id

Abstract. *Selesainya Rancangan Undang-Undang (RUU) pemindahan ibu kota pada awal September 2021, pembangunan Ibu Kota Nusantara (IKN) resmi dimulai pada bulan Juli tahun 2022. Persoalan ini mengundang pro dan kontra di masyarakat, karena kepentingan dan pandangan masing-masing. perkembangan dunia digital semakin canggih, informasi lebih mudah didapatkan melalui portal dan media sosial. Media sosial merupakan salah satu tempat untuk menyampaikan opini masyarakat. Melalui text mining di X kita dapat memahami gambaran orang dalam persepsi mereka terhadap kebijakan pemerintah, baik Positif, Netral dan Negatif. Analisis sentimen ini perlu dilakukan, oleh karena itu penelitian Analisis Efektivitas Metode Naïve Bayes Pada Sentimen Opini Publik di X Terhadap Pemindahan Ibu Kota Nusantara (IKN) dilakukan untuk melihat polarisasi opini masyarakat dan menguji efektivitas metode Naïve Bayes terhadap dataset IKN. Rancangan penelitian cross sectional karena data yang diambil dari aplikasi X sejak tanggal 18 Januari 2022 sampai dengan 3 Maret 2024. Hasil screping data didapatkan sebanyak 3775 tweet, kemudian dicleaning sehingga menghasilkan 2778 data tweet. Dalam scrapping dan analisis data tweet alat yang digunakan google colab dan bahasa pemograman phyton. Data dibagi menjadi data latih sebanyak 80% dan data uji 20%. Polarisasi sentimen tweet positif sebanyak 1153 (39.32%) Netral 991 (33.80%) dan Negatif 788 (26.88%). Tingkat akurasi untuk tiap tiap jenis naïve bayes, GaussianNB 0.5168, MultinomialNB 0.6133 dan bernoulliNB 0.6115. hasil uji akurasi menggunakan confusion matrix menggambarkan metode Naïve Bayes dengan jenis MultinomialNB mendapatkan akurasi yang lebih tinggi dalam mengklasifikasi dataset IKN.*

Keywords: *Classification, Naïve Bayes, X, Ibu Kota Nusantara (IKN).*

1 LATAR BELAKANG

Pada tahun 2019 Presiden Republik Indonesia Ir. Joko Widodo, pada tanggal 23 Agustus melalui pidato kenegaraan menyampaikan bahwa ibu kota negara Republik Indonesia akan dipindahkan ke pulau Kalimantan tepatnya di provinsi Kalimantan Timur [1]. *Persoalan ini mengundang pro dan kontra di masyarakat, karena kepentingan dan pandangan masing-masing.* Pihak pro atau setuju dengan alasan pembangunan Ibu Kota Nusantara (IKN) mampu mewujudkan pemerataan ekonomi terutama masyarakat di luar pulau jawa, dan juga ada yang mengatakan bahwa Kalimantan lebih aman dari gempa bumi dibandingkan dengan pulau jawa. Disisi lain yang tidak setuju atau kontra karena permasalahan lingkungan, khawatir kerusakan lingkungan akan terjadi akibat pembangunan kota bau. Dewasa ini perkembangan dunia digital semakin canggih, sehingga informasi lebih mudah didapatkan melalui portal dan media sosial. Media sosial merupakan salah satu tempat untuk menyampaikan opini masyarakat. Opini masyarakat dapat berupa pujian atau ujaran kebencian dan *hoax* yang dapat menimbulkan perdebatan di media sosial. Salah satu media diskusi adalah melalui aplikasi *twitter* atau X.

[2]. *Twitter* atau *X* adalah sebuah media sosial dan layanan *microblogging* yang diciptakan oleh Jack Dorsey dan dioperasikan oleh Twitter, Inc. pada tahun 2006 untuk difokuskan terhadap kebutuhan komunikasi yang cepat. *Microblogging* sendiri memiliki arti sebuah situs dimana penggunaanya dapat menyiarkan pesan atau tulisan dalam ukuran lebih kecil. [3] Penelitian yang serupa juga pernah dilakukan oleh Fachreza dkk. dengan judul *Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes* dan memberikan hasil uji coba terbaik pada penggunaan fitur sebanyak 80% dengan *accuracy* 82,0%, *precision* 76,9%, dan *recall* 100%. Penelitian ini menggunakan jenis *Gaussian Naïve Bayes* karena asumsi independensi [4]. Kemudian juga ada penelitian *Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN* Dalam penelitian ini metode yang digunakan adalah *Naïve Bayes (NB)* dan *K-Nearest Neighbor (KNN)*. Hasil penelitian menyajikan hasil komparasi kinerja metode tersebut bahwa metode *Naïve Bayes* memberikan tingkat akurasi analisis sentimen sebesar 82.27%, nilai *Precision* sebesar 86.36% dan nilai *Recall* sebesar 76.93%. Kinerja metode *KNN* juga menyajikan hasil analisis dengan tingkat akurasi sebesar 88,12%, *Precision* sebesar 93.%98 dan nilai *recall* sebesar 81.53%. Berdasarkan hasil analisis tersebut maka proses analisis menggunakan metode *KNN* mengungguli metode *NB* dalam mengukur sentimen terhadap pemindahan Ibu Kota Nusantara [5]. Melalui *text mining* di *X* kita dapat memahami gambaran orang dalam persepsi mereka terhadap kebijakan pemerintah dalam hal ini adalah pemindahan ibu kota negara, baik Positif, Netral dan Negatif. Analisis sentimen ini perlu dilakukan, oleh karena itu penelitian *Analisis Efektivitas Metode Naïve Bayes Pada Sentimen Opini Publik di X Terhadap Pemindahan Ibu Kota Nusantara (IKN)* dilakukan untuk melihat polarisasi opini masyarakat dan menguji efektivitas metode *Naïve Bayes* terhadap dataset *IKN*.

2 KAJIAN TEORITIS

Beberapa penelitian sebelumnya yang menjadi rujukan peneliti dalam penelitian ini adalah dengan judul *Implementasi Metode Klasifikasi Naïve Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web*. [6]. Tujuan dari penelitian ini untuk membuat sistem analisis sentimen yang bisa menjadi acuan dalam penentuan kebijakan baik pemerintah atau umum dan untuk mengetahui kapan dan apa penyebab perubahan sentimen di media sosial *Twitter*.

Penelitian kedua adalah dengan judul *Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritma Naïve Bayes*. [7]. Penelitian ini dilakukan sebuah analisa sentimen terhadap bakal calon Presiden 2024 dengan algoritma *naïve bayes*, Pada proses penelitian ini menggunakan metode *Knowledge Discovery from Data* dengan tahapannya yakni *Data Cleaning*, *Data Transformation*, *Data Mining*, *Pattern Evaluation*, dan *Knowledge Presentation*. Selain penggunaan metode tersebut ditambahkan satu tahapan metode yakni *Data Collection* yang bertujuan untuk melakukan pengumpulan data pada media sosial *twitter*.

Penelitian kedelapan dilakukan oleh [8]. dalam penelitian ini mengklasifikasikan sentimen masyarakat Indonesia mengenai isu *Omnibus Law* menggunakan metode *Naïve Bayes* dan *Particle Swarm Optimization (PSO)* dengan tiga kali pengujian. Penggunaan *PSO* ini bertujuan untuk mengoptimalkan hasil akurasi.

Penelitian ke Sembilan dilakukan oleh [9]. dengan judul penelitiannya *Penerapan Naïve Bayes dan Support Vector Macine pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter* dan untuk memastikan hasil penelitiannya dilakukan juga pengujian dengan *Support Vector Machine* yang menghasilkan nilai akurasi sebesar 79.96%.

Penelitian kesepuluh dilakukan oleh [10]. *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional*. Hasil Penelitian ini masuk kedalam *finned grained sentiment analysis* yaitu analisis pada suatu kalimat komentar. Data tersebut akan

diproses menggunakan text mining, kemudian dilanjutkan dengan mengklasifikasikan *Tweet* kedalam tiga kelas, yaitu positif, negatif, dan netral.

Penelitian kesebelas dilakukan oleh [11]. Penelitiannya adalah Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network.

Penelitian ke dua belas yang peneliti kutip dalam literature review ini adalah [12]. dalam pemaparannya dilakukan teknik seperti transform, tokenize, stemming, classification, dan lain-lain sangat berpengaruh pada akurasi. Gata Framework digunakan untuk preprocessing, dan Rapidminer juga digunakan untuk menganalisa dan membandingkan tiga metode klasifikasi yaitu Naive Bayes, Support Vector Machine, dan k-NN.

Penelitian ke tiga belas peneliti kutip dari artikel yang terikdeks di jurnal internasional yang diteliti oleh [13] dengan judul Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking.

Penelitian ke empat belas dilakukan oleh [14] dengan judul Evaluation of Bernoulli Naive Bayes model for detection of distributed denial of service attacks. Dalam makalah ini menyajikan model prediktif yang mengidentifikasi distributed denial of service attacks (DDSA) menggunakan Bernoulli- Naive Bayes. Model yang dikembangkan dievaluasi berdasarkan kumpulan data Kaggle yang tersedia untuk umum, metode ini diuji dengan matriks konfusi, kurva receiver operating characteristics (ROC), dan akurasi untuk mengukur kinerjanya. Hasil percobaan menunjukkan akurasi 85,99% dalam mendeteksi DDSA dengan metode yang diusulkan.

Kemudian ke lima belas penelitian yang dilakukan oleh [15] dengan judul Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes. Hasil penelitian ini difokuskan pada model atau pola yang dihasilkan oleh Algoritma Multinomial Naive Bayes. Hasil klasifikasi *tweet* pengguna media sosial terhadap kebijakan new normal memperoleh hasil yang baik dengan nilai akurasi sebesar 90,25%. Setelah mengklasifikasikan *tweet* pengguna media sosial mengenai kebijakan new normal, diperoleh hasil lebih dari 70% setuju dan mendukung kebijakan new normal.

Penelitian ke enam belas adalah dilakukan oleh [16], Algoritma Naive Bayes Classifier memiliki kinerja yang baik dalam klasifikasi karya ilmiah (tugas akhir) ditunjukkan dengan hasil pengujian 20 data karya ilmiah berdasarkan parameter latar elakang menghasilkan 18 data diklasifikasikan dengan benar dan 2 data lainnya terdeteksi salah, dan 20 data pengujian tersebut didapat rata-rata waktu proses pengujian dengan menerapkan algoritma Naive Bayes Classifier yaitu 5,7406 detik/pengujian. Dan Tingkat akurasi dari pengujian tersebut yang diklasifikasikan kedalam 5 kelas didapatkan nilai rata-rata akurasi yang cukup baik yaitu 86,68%.

Penelitian ke tujuh belas adalah yang dilakukan oleh [17]. Hasil dari penelitian ini adalah bahwa system aplikasi untuk melacak klaim PKH telah dikembangkan untuk menentukan apakah suatu klaim sah atau tidak berdasarkan kriteria yang telah ditetapkan. Sistem ini dapat merekomendasikan tindakan kepada peserta PKH agar mereka tetap pada jalurnya. Dari hasil pengujian yang dilakukan dengan menggunakan data sebanyak 500, 400 as data training dan 100 as data testing, akurasi polanya sebesar 86,00 persen dan errornya 14,00 persen, yang jumlah data yang tepat sebanyak 86 dan yang tidak tepat sebanyak 14.

Penelitian ke delapan belas dilakukan oleh [18]. Menggunakan perbandingan metode Nave Bayes dan metode Dempster Shafer. Jumlah data yang digunakan dalam penelitian ini adalah 550 data, terdiri dari 500 data latih dan 50 data uji. Sedangkan variabel yang digunakan adalah gejala terkait COVID-19 sebanyak 17 gejala. Hasil uji persentase dengan cara membandingkan diagnosa sistem dengan diagnosa ahli, untuk metode nave bayes mempunyai akurasi sebesar 96% dengan 48 diagnosa menurut diagnosa ahli dari 50 data yang diuji. Sementara itu, Dempster Shafer metode ini memiliki akurasi sebesar 40% dengan 20 diagnosa menurut diagnosa ahli dari 50 data yang diuji.

Penelitian ke Sembilan belas adalah yang dilakukan oleh [19]. menganalisis penggunaan website pada Perguruan Tinggi (PTN) di Aceh, serta menganalisis kelebihan dan kelemahan masing-masing website. Kemudian melakukan analisis kinerja website menggunakan Google Page Speed Insight (GPSI) dan alat online GTMetrix. Pada kontrol variabel menggunakan Qualys SSL Lab dan SECURi. Hasil penelitian menunjukkan USK Informasi dengan nilai 4,26, Ekonomi 4,13, Efisiensi 4,4, dan Pelayanan 4,36. UINAR memperoleh nilai 4,26 pada mata kuliah Informasi, Ekonomi 4,13, Efisiensi 4,3, dan Pelayanan 4,36. UNIMAL memperoleh Informasi 4,26, Ekonomi 4,23, Efisiensi 4,3, dan Pelayanan 4,15. UNSAM dengan mata pelajaran Informasi memperoleh 4,1, Ekonomi 4,13, Efisiensi 4,19 dan Pelayanan 4,15 serta mata pelajaran Informasi UTU 4,24, mata pelajaran Ekonomi 4, Efisiensi 4,19 dan Pelayanan dengan nilai 4,21. Tes GPSI dan GTMetrix menunjukkan hasil yang signifikan dengan skor tertinggi UNSAM sebesar 97 pada GPSI.

Penelitian ke dua puluh yang peneliti kutip adalah tesis [20]. Han dkk. juga menyatakan *Naïve Bayes Naïve Bayes Classifier* mengasumsikan bahwa nilai suatu atribut pada kelas tertentu tidak bergantung pada nilai atribut lainnya. Asumsi ini disebut independensi kelas bersyarat. Hal ini dilakukan dengan tujuan menyederhanakan perhitungan yang rumit dan dalam pengertian ini dianggap “naïve”.

Penelitian terakhir adalah [21] dengan judul Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia, Berdasarkan hasil analisis sentimen tweet didapatkan sentimen positif sebanyak 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) tweet dan sentimen netral dengan jumlah 9.174 (21,1%) dari total 43.435 tweet. Nilai akurasi dengan algoritma Naïve Bayes didapatkan hasil nilai akurasi sebesar 94%.

3 METODE PENELITIAN

The method or stages carried out in this study are shown in Figure 1.

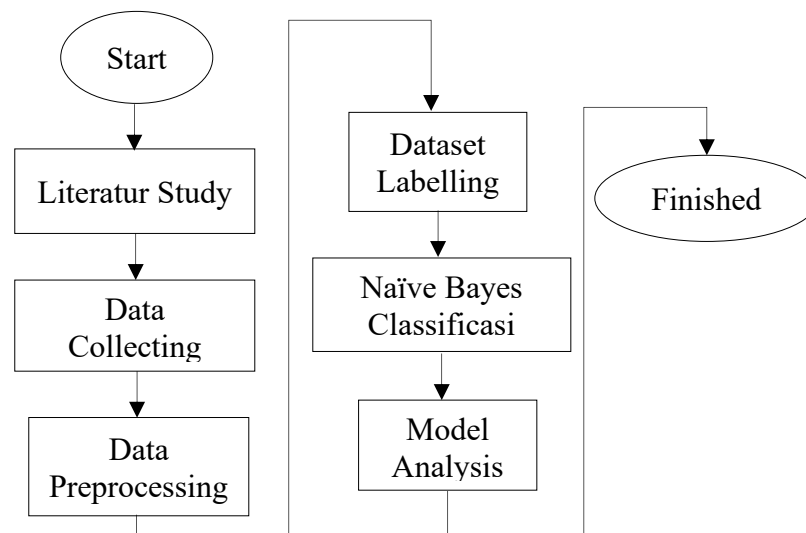


Figure 1. Research Methods or Stages

Berikut penjelasan gambar 1. Metode atau tahapan yang dilakukan dalam penelitian klasifikasi sentimen tweet terhadap pemindahan IKN.

a. Literature Study

Studi literatur adalah dasar teori yang digunakan dalam penelitian untuk menyelesaikan permasalahan dan merupakan referensi yang kuat dalam melakukan analisa. Pada tahap ini penulis menelusuri teori-teori yang terkait dengan penelitian ini dan menguraikan teori

yang menjadi pendukung penelitian ini seperti metode Naïve Bayes Clasifikasi (NBC), *Data Mining*, *Text Mining*, sentiment analisis dan lain-lain.

b. Pengumpulan Data

Pengumpulan data adalah untuk mendapatkan informasi berdasarkan opini masyarakat atau sentimen yang di ekspresikan melalui media sosial *twitter* atau X. Teknik pengumpulan data yang dilakukan pada penelitian ini adalah dengan melakukan *scrapping tweet* pada aplikasi media sosial *twitter* atau X menggunakan bahasa *phyton* di *google colab* mulai dari tanggal 18 Januari 2022 sampai dengan 3 Maret 2024 sebanyak 3754 data *tweet* dan disimpan dalam format *.csv*. Pada proses *scrapping tweet* ini penulis mendapatkan *tweet* dengan kata kunci Ibu Kota Baru sejumlah 776. *Tweet* dengan kata kunci *ikn* atau *IKN* di dapatkan sebanyak 1587 *tweet*, dan terakhir *tweet* dengan kata kunci ibu kota negara sebanyak 694 *tweet*.

c. Preprocessing Data

Sebelum data diproses pengklasifikasian menggunakan metode Naïve Bayes Classification (NBC). Preprocessing data yang dilakukan pada data penelitian ini adalah Hapus data duplikat, Pembersihan Data, Case folding, Normalisasi Kata, Tokenizing, Filtering, Stemming, dan Hapus data bernilai kosong.

d. Labeling Dataset

Tahap terakhir adalah melakukan pelabelan *tweet* sesuai dengan kelas yang di tentukan sebelumnya. Dataset yang telah berhasil dikumpulkan dari *twitter* merupakan data yang tidak mempunyai label (*unsupervised data*), sehingga agar dapat diproses dengan menggunakan teknik *supervised learning*, maka data *twitter* yang telah dikumpulkan sebelumnya perlu untuk diberi label terlebih dahulu.

Proses pelabelan data dilakukan dengan memberikan nilai prediksi kelas sentiment yang telah ditentukan sebelumnya yaitu Positif, Netral dan Negatif. Proses pelabelan dalam penelitian ini dilakukan secara manual untuk memberikan nilai klasifikasi dengan cara menghitung banyak *text* yang terkandung dalam pada setiap dokumen, proses pelabelan dilakukan sebagai berikut.

1. Tentukan kata-kata yang memiliki arti positif, negatif dan netral
2. Hitung jumlah kata positif, negatif dan juga netral pada dokumen
3. Jika jumlah kata positif > jumlah kata negatif, maka label sentimennya adalah positif
4. Jika jumlah kata negatif > jumlah kata positif, maka label sentimennya adalah negatif
5. Jika jumlah kata netral > jumlah kata positif dan negatif, maka label sentimennya adalah netral.

e. Model Naïve Bayes Clasifikasi (NBC)

Pada tahap ini data hasil yang telah diberi label, selanjutnya akan dilakukan proses klasifikasi sentimen negatif, positif dan netral terhadap data *tweet* menggunakan metode *Naïve Bayes Classification* (NBC) yaitu *Gaussian Naive Bayes*, *Multinomial Naive Bayes*, *Bernoulli Naive Bayes*. Pada implementasi dengan metode NBC, dataset yang telah ditransformasi dipisahkan, sehingga dalam pembagian data tes dan data uji dengan pembagian 80 : 20.

Adapun langkah-langkah yang dilakukan dalam klasifikasi opini publik terhadap pemindahan Ibu Kota Nusantara menggunakan metode Naïve Bayes ini adalah sebagai berikut:

1. Masukkan dataset sebanyak 2932 *record*.
2. Pembagian dataset menjadi 2 bagian yaitu data latih dan data uji dengan komposisi 80:20.
3. Hitung $P(C_i)$ untuk setiap kelas
4. Hitung $P(X|C_i)$, i =positif, netral, negatif untuk setiap kelas
5. Hasil klasifikasi.

f. Analisis Model

Pada bagian ini hasil klasifikasi Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes akan dievaluasi menggunakan confusion matrix untuk melihat akurasi dari ketiga metode tersebut berdasarkan nilai yang diperoleh dari nilai True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN).

4 HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pada penelitian ini, dataset diperoleh melalui proses *Scrapping* data *Twitter* atau *X* dengan menggunakan *Google Colab* dan bahasa pemrograman *Python*. Proses pengumpulan data dilakukan dari tanggal 18 Januari 2022 hingga 3 Maret 2024 dengan menggunakan kata kunci yang spesifik. Kata kunci yang digunakan antara lain *'Ibu Kota Baru'* dan *'IKN'*. Hasil keseluruhan dari penggunaan kata kunci tersebut menghasilkan 3.755 *tweet*. Berikut adalah tabel 1 Hasil *scrapping* data *tweet*.

Tabel 1. Hasil Scrapping Data

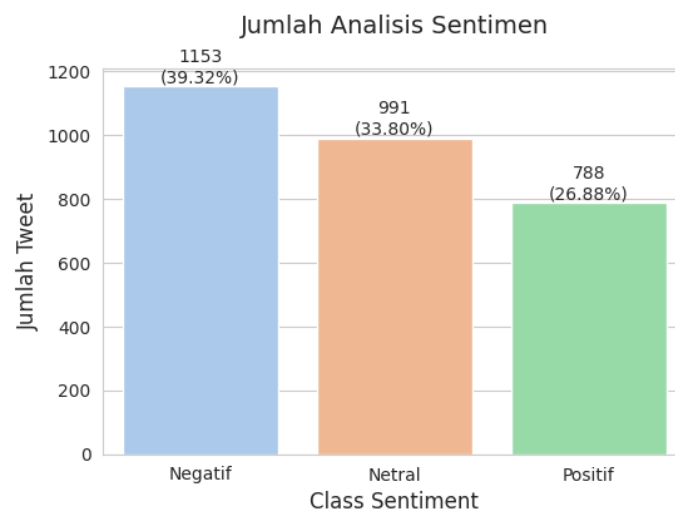
username	full_text	cleaning	case_folding	normalisasi
rahman_surya7	@dekor_haus Benar benar mendoakan agar IKN dap...	Benar benar mendoakan agar IKN dapat dilaksan...	benar benar mendoakan agar ikn dapat dilaksan...	benar benar mendoakan agar ikn dapat dilaksana...
haebaragi	@tanyarlfe Jangan dong, yg di jkt aja sulit a...	Jangan dong yg di jkt aja sulit apalagi klo d...	jangan dong yg di jkt aja sulit apalagi klo d...	jangan dong yang di jakarta saja sulit apalagi...
ikn0ss	デザストと蓮く んは、ずっと好 き。			
ibrahim_karai	@AgusYudhoyono Ente dulu Habis Habisan mengkri...	Ente dulu Habis Habisan mengkritik IKN begitu...	ente dulu habis habisan mengkritik ikn begitu...	ente dulu habis habisan mengkritik ikn begitu ...
Ikn210	@Gabriel_gunner2 mzee leo nimeregister 4 of my...	mzee leo nimeregister of my friendsNipaste n...	mzee leo nimeregister of my friendsnipaste n...	mzee leo nimeregister of my friendsnipaste nip...

Setelah *Scrapping* data *Twitter* atau *X* dan sudah dilakukan preprocessing, Pelabelan data menggunakan kamus lexicon based adalah proses pemberian label sentimen pada data teks berdasarkan kamus kata-kata yang memiliki konotasi positif atau negatif. Dengan menggunakan kamus leksikon, setiap kata dalam teks dibandingkan dengan daftar kata positif dan negatif. Jika jumlah kata positif lebih banyak, teks diberi label "Positif"; jika jumlah kata negatif lebih banyak, teks diberi label "Negatif"; dan jika jumlahnya sama atau tidak ada, teks diberi label "Netral". Berikut Tabel 2 Hasil Labelling data.

Table 2. Hasil Labelling Data

	steming_data	Sentiment
0	doa ikn laksana lancar tuntas kendala indonesi...	Positif
1	jakarta sulit ikn sulit	Netral
2	NaN	Netral
3	ente habis habis kritik ikn kasih kue kuasa la...	Positif
4	mzee leo nimeregister of my friendsnipaste nip...	Netral

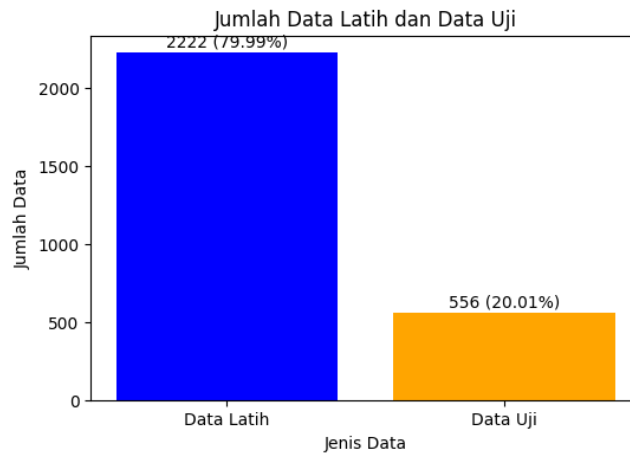
Setelah melakukan klasifikasi, kemudian akan dilakukan visualisai sentiment, berikut adalah visualisasi setimen menggunakan diagram batang, seperti pada Gambar 2 berikut ini.



Gambar 2. Hasil Analisis Tweet terkait IKN

Berdasarkan hasil analisis seperti pada gambar diatas, menunjukkan bahwa sentimen negatif terbanyak yaitu sekitar 1153 tweet (39.32%), kemudian netral sebanyak 991 tweet (33.80%) dan positif sebanyak 788 tweet (26.88%). Total tweet 2932.

Splitting Dataset adalah proses membagi dataset menjadi beberapa bagian untuk tujuan pelatihan dan pengujian model. Proses ini penting untuk mengevaluasi performa model secara objektif. Dataset dibagi menjadi dua bagian yaitu 80% data pelatihan yang digunakan untuk melatih model, dan data pengujian 20% yang digunakan untuk menguji model, berikut ini Gambar 3 Diagram Data uji dan data latih.



Gambar 3. Pembagian tweet data uji dan data latih

4.2 Penerapan Metode Naïve Bayes

Probabilitas posterior adalah probabilitas suatu kelas atau kategori setelah mempertimbangkan fitur. Dalam konteks klasifikasi menggunakan algoritma Naive Bayes, probabilitas posterior digunakan untuk menentukan kelas terbaik untuk sebuah data berdasarkan fitur-fitur yang dimilikinya.

Dalam Probabilitas Posterior dianalisis memiliki kata-kata word1, word2, dan word3. Probabilitas posterior untuk setiap kelas dihitung dengan rumus:

$$P(C|\text{Features}) \propto P(C) \cdot P(\text{word1}|C) \cdot P(\text{word2}|C) \cdot P(\text{word3}|C)$$

Dimana:

$P(C)$ adalah probabilitas prior untuk kelas C.

$P(\text{wordi}|C)$ adalah probabilitas kondisional token wordi dalam kelas C.

Probabilitas Posterior untuk Positif:

$$P(\text{Positif}|\text{document}) \propto P(\text{Positif}) \cdot P(\text{word1}|\text{Positif}) \cdot P(\text{word2}|\text{Positif}) \cdot P(\text{word3}|\text{Positif})$$

Dalam contoh, kita mendapatkan:

$$P(\text{Positif}|\text{document}) \approx 1.497218483974853 \times 10^{(-31)}$$

Probabilitas Posterior untuk Negatif:

$$P(\text{Negatif}|\text{document}) \propto P(\text{Negatif}) \cdot P(\text{word1}|\text{Negatif}) \cdot P(\text{word2}|\text{Negatif}) \cdot P(\text{word3}|\text{Negatif})$$

Dalam contoh, kita mendapatkan:

$$P(\text{Negatif}|\text{document}) \approx 2.1907270457144735 \times 10^{(-31)}$$

Probabilitas Posterior untuk Netral:

$$P(\text{Netral}|\text{document}) \propto P(\text{Netral}) \cdot P(\text{word1}|\text{Netral}) \cdot P(\text{word2}|\text{Netral}) \cdot P(\text{word3}|\text{Netral})$$

Dalam contoh, kita mendapatkan:

$$P(\text{Netral}|\text{document}) \approx 1.590319633359076 \times 10^{(-31)}$$

Hasil hitung probabilitas posterir adalah seperti pada tabel 3 berikut ini.

Tabel 3. Hasil hitung probabilitas posterior

	steming_data	Sentiment	posterior_probabilities
0	doa ikn laksana lancar tuntas kendala indonesi...	Positif	{'Positif': 1.497218483974853e-31, 'Negatif': ...

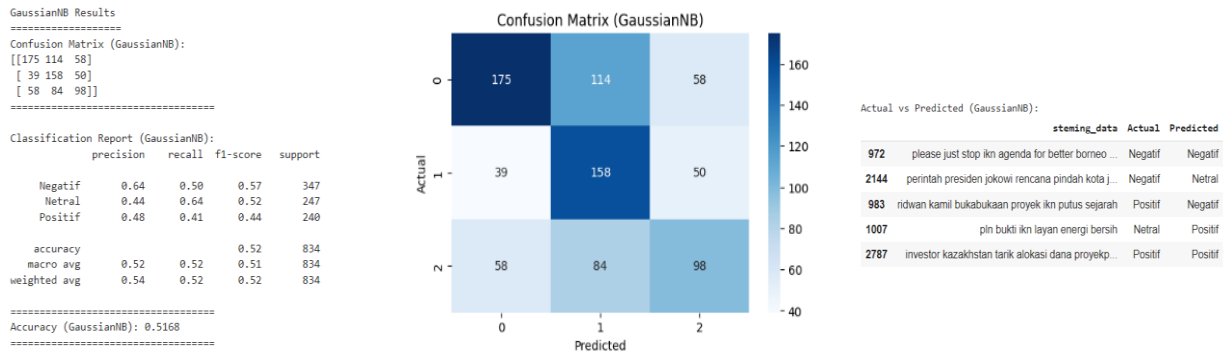
1	jakarta sulit ikn sulit	Netral	{'Positif': 2.060817710950884e-16, 'Negatif': ...
3	ente habis habis kritik ikn kasih kue kuasa la...	Positif	{'Positif': 4.813544416367439e-58, 'Negatif': ...
4	mzee leo nimeregister of my friendsnipaste nip...	Netral	{'Positif': 2.458083211254068e-35, 'Negatif': ...
5	jakarta ajeeee ikn kejauhannn anjg ongkos gue	Negatif	{'Positif': 9.11955778589083e-28, 'Negatif': 1...

4.2.1. Naive Bayes Gaussian (GaussianNB)

Naive Bayes Gaussian (GaussianNB) adalah salah satu varian dari algoritma Naive Bayes yang mengasumsikan bahwa fitur-fitur dalam dataset mengikuti distribusi normal (Gaussian). Ini adalah teknik klasifikasi yang sederhana namun efektif, khususnya dalam kasus di mana fitur-fitur dianggap saling bebas satu sama lain. Persamaannya adalah sebagai berikut.

$$P(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Berikut adalah hasil dari kinerja GaussianNB, sebagaimana dalam Gambar 4 dibawah ini.



Gambar 4. Akurasi Kinerja Model GaussianNB

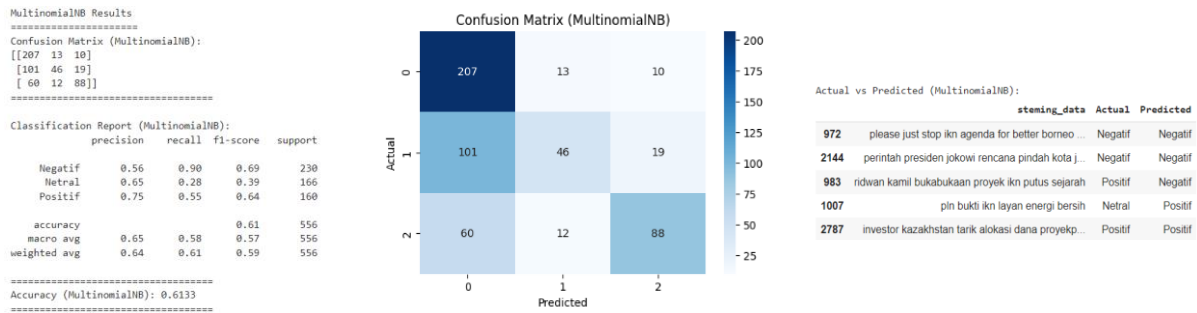
4.2.2. Naive Bayes Multinomial (MultinomialNB)

Naive Bayes Multinomial (MultinomialNB) adalah salah satu varian dari algoritma Naive Bayes yang digunakan untuk klasifikasi teks dan masalah klasifikasi lainnya yang melibatkan data diskret. Naive Bayes Multinomial menggunakan teorema Bayes, probabilitas dokumen d termasuk dalam kelas C. Berikut adalah persamaan Naive Bayes Multinomial. Berikut adalah persamaannya.

$$P(c|d) \propto P(c) \prod_{i=1}^N P(x_i|c)^{k_i}$$

Hasil kinerja dari model MultinomialNB adalah seperti pada Gambar 4 Berikut ini.

Analysis of the Effectiveness of the Naïve Bayes Method on Public Opinion Sentiment in X Regarding the Relocation of the Indonesian Capital City (IKN)



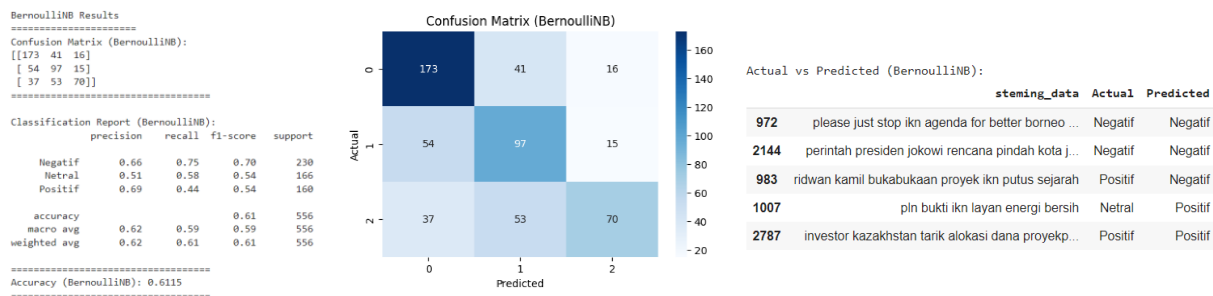
Gambar 5. Akurasi Kinerja Model MultinomialNB

4.2.3. Naive Bayes Bernoulli (BernoulliNB)

Naive Bayes Bernoulli (BernoulliNB) adalah salah satu varian dari algoritma Naive Bayes yang dirancang untuk data biner atau data yang dapat dipetakan ke dalam bentuk biner. Probabilitas Posterior dalam Bernoulli menggunakan teorema Bayes, probabilitas dokumen d termasuk dalam kelas c. Berikut adalah persamaannya.

$$P(c|d) \propto P(c) \prod_{i=1}^N P(x_i = 1|c)^{x_i} \cdot P(x_i = 0|c)^{(1-x_i)}$$

Hasil kinerja dari model BernoulliNB adalah seperti pada Gambar 6 Berikut ini.



Gambar 6. Akurasi Kinerja Model BernoulliNB

4.2.4. Word Cloud

Wordcloud adalah visualisasi grafis yang digunakan untuk menampilkan frekuensi atau pentingnya kata-kata dalam sebuah teks. Kata-kata yang lebih sering muncul dalam teks ditampilkan dengan ukuran yang lebih besar dan lebih menonjol dibandingkan dengan kata-kata yang kurang sering muncul. Berikut Gambar 7 adalah visualisasi word cloud sentimen.



Gambar 7. Visualisasi WordCloud

5 KESIMPULAN

Penelitian ini menggunakan algoritma Naive Bayes untuk menganalisis sentimen *tweet* terkait Ibu Kota Negara (IKN) baru, mengklasifikasikannya ke dalam sentimen positif, negatif, dan netral. Hasil analisis menunjukkan bahwa sentimen negatif lebih dominan dibandingkan sentimen netral dan positif. Evaluasi model dilakukan menggunakan metrik precision, recall, F1-score, dan accuracy untuk tiga jenis Naive Bayes: GaussianNB, MultinomialNB, dan BernoulliNB. GaussianNB menunjukkan precision 0.5043 (negatif), 0.6393 (netral), 0.4756 (positif), recall 0.6434 (negatif), 0.4433 (netral), 0.4083 (positif), F1-score 0.5645 (negatif), 0.5223 (netral), 0.4399 (positif), dan akurasi 51.68%. MultinomialNB memiliki precision 0.562 (negatif), 0.648 (netral), 0.752 (positif), recall 0.900 (negatif), 0.277 (netral), 0.550 (positif), F1-score 0.692 (negatif), 0.389 (netral), 0.637 (positif), dan akurasi 61.3%. BernoulliNB menunjukkan precision 0.808 (negatif), 0.648 (netral), 0.752 (positif), recall 0.807 (negatif), 0.277 (netral), 0.550 (positif), F1-score 0.808 (negatif), 0.389 (netral), 0.637 (positif), dan akurasi 70.0%. Visualisasi data dilakukan melalui wordcloud untuk setiap kategori sentimen, yang menunjukkan kata-kata paling sering muncul dalam teks berdasarkan sentimen yang berbeda. Pada sentimen negatif, kata yang menonjol adalah "korupsi", "biaya", dan "masalah". Pada sentimen netral, kata yang menonjol adalah "pemindahan", "proyek", dan "proses". Pada sentimen positif, kata yang menonjol adalah "peluang", "pertumbuhan", dan "investasi". Berikut adalah tiga poin saran yang lebih singkat untuk penelitian ini: Eksplorasi algoritma lain seperti Random Forest, SVM, LSTM, atau BERT, serta penggunaan teknik ensemble untuk meningkatkan performa klasifikasi sentimen. Perluas kata kunci dan periode pengumpulan data, serta pertimbangkan sumber data lain seperti berita, forum online, dan platform media sosial lainnya untuk representasi sentimen yang lebih komprehensif. Lakukan analisis sentimen dalam berbagai bahasa lokal selain bahasa Indonesia untuk mencerminkan keberagaman bahasa dan budaya, serta gunakan model bahasa yang mendukung analisis teks multibahasa.

DAFTAR REFERENSI

- [1] H. Dhery, A. Assyam, and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.
- [2] D. Rusdian and D. Rosiyadi, "METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE," vol. 4, no. 2, pp. 230–235, 2019.
- [3] R. Safitri, N. Alfira, and D. Tamitiadini, *Analisis Sentimen : Metode Alternatif Penelitian Big Data*. Malang: UB Press, 2021. [Online]. Available: <https://play.google.com/books/reader?id=TVdmEAAAQBAJ&pg=GBS.PR4&hl=id>
- [4] M. R. D. Fachreza, S. Suhartono, and M. A. Yaqin, "Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 3, pp. 243–251, 2023, doi: 10.14421/jiska.2023.8.3.243-251.
- [5] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [6] Apif Supriadi and Fatmasari, "Implementasi Metode Klasifikasi Naive Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web," *J. Sist. Inf.*, vol. 10, no. 1, pp. 46–54, 2021, doi: 10.51998/jsi.v10i1.356.
- [7] M. R. Fais Sya' bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [8] S. A. Dainamang, N. Hayatin, and D. R. Chandranegara, "Analisis Sentimen Media Sosial Twiiter terhadap RUU Omnibus Law dengan Metode Naive Bayes dan Particle Swarm Optimization," *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 211–218, 2022, doi: 10.34010/komputika.v11i2.6037.
- [9] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available:

- <http://jim.teknokrat.ac.id/index.php/informatika>
- [10] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
 - [11] I. Zulfa, E. Winarko, P. Studi, S. Ilmu, and F. Ugm, “Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network,” vol. 11, no. 2, 2017.
 - [12] S. Hikmawan, A. Pardamean, and S. N. Khasanah, “Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning,” *J. Kaji. Ilm.*, vol. 20, no. 2, pp. 167–176, May 2020, doi: 10.31599/jki.v20i2.117.
 - [13] M. Yusran, S. Siswanto, and A. Islamiyati, “Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking,” *Sistemasi*, vol. 13, no. 1, p. 96, 2024, doi: 10.32520/stmsi.v13i1.3187.
 - [14] A. O. Salau, T. A. Assegie, A. T. Akindadelo, and J. N. Eneh, “Evaluation of Bernoulli Naive Bayes model for detection of distributed denial of service attacks,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 1203–1208, 2023, doi: 10.11591/eei.v12i2.4020.
 - [15] W. B. Zufikar, A. R. Atmadja, and S. F. Pratama, “Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes,” *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.
 - [16] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, “Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC),” *Sistemasi*, vol. 10, no. 2, p. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
 - [17] M. Qamal, I. Sahputra, N. Nurdin, M. Maryana, and M. Mukarramah, “Sistem Pendukung Keputusan Penentuan Penerimaan Bantuan PKH Menggunakan Metode Naïve Bayes,” *TECHSI - J. Tek. Inform.*, vol. 14, no. 1, p. 21, 2023, doi: 10.29103/techsi.v14i1.6960.
 - [18] N. Nurdin, E. Susanti, H. A.-K. Aidilof, and D. Priyanto, “Comparison of Naive Bayes and Dempster Shafer Methods in Expert System for Early Diagnosis of COVID-19,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 1, pp. 215–228, 2022, doi: 10.30812/matrik.v22i1.2280.
 - [19] Suhaili Sahibul Muna, N. Nurdin, and T. Taufiq, “Comparative Analysis of State Universities on Website Performance in Aceh Using the PIECES Method,” *J. Informatics Telecommun. Eng.*, vol. 7, no. 1, pp. 71–83, 2023, doi: 10.31289/jite.v7i1.9167.
 - [20] K. Wabang, “No Title,” *Implementasi Algoritma Naïve Bayes Classif. (Bernoulli NB, Multinomial NB, Dan Gaussian NB) Untuk Mengklasifikasi Pengaduan Masy.*, 2023.
 - [21] F. F. Mailoa, “Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia,” *J. Inf. Syst. Public Heal.*, vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.