



Analisis Sentimen Opini Publik Terhadap Kebijakan Pemerintah Tentang Penetapan Bencana Nasional Sumatera Aceh Menggunakan Algoritma Naïve Bayes

Ghivarel¹, Joni Karman^{2*}, Satrianansyah²

¹ Fakultas Ilmu Teknik, Program Studi Informatika, Universitas Bina Insan, Lubuklinggau, Indonesia

² Fakultas Ilmu Teknik, Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau, Indonesia

Email: ¹erelkole@gmail.com, ^{2*}joni_karman@univbinainsan.ac.id, ³satrianansyah@univbinainsan.ac.id

Email Penulis Korespondensi: joni_karman@univbinainsan.ac.id

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap kebijakan pemerintah mengenai penetapan Bencana Nasional Aceh dengan memanfaatkan teknik machine learning pada data ulasan yang diperoleh dari media daring. Proses pelabelan sentimen dilakukan secara otomatis berdasarkan skor penilaian pengguna, di mana skor 1–2 dikategorikan sebagai sentimen negatif, skor 3 sebagai sentimen netral, dan skor 4–5 sebagai sentimen positif. Hasil pelabelan menunjukkan bahwa distribusi data didominasi oleh sentimen negatif sebanyak 5.593 ulasan, diikuti sentimen netral sebanyak 1.956 ulasan dan sentimen positif sebanyak 1.460 ulasan, yang mengindikasikan dominasi respons kritis masyarakat terhadap kebijakan tersebut. Evaluasi kinerja model dilakukan menggunakan cross-validation, confusion matrix, dan kurva ROC (Receiver Operating Characteristic). Hasil cross-validation menunjukkan akurasi rata-rata sebesar 80,62% dengan standar deviasi yang rendah, menandakan performa model yang stabil dan konsisten. Analisis confusion matrix memperlihatkan bahwa model mampu mengklasifikasikan sentimen negatif dengan sangat baik, namun masih mengalami kesulitan dalam membedakan sentimen netral dan positif akibat ambiguitas bahasa dan tumpang tindih makna. Selanjutnya, evaluasi menggunakan kurva ROC menunjukkan nilai AUC sebesar 0,84 untuk kelas negatif serta 0,82 untuk kelas netral dan positif, yang mengindikasikan kemampuan diskriminasi model yang baik pada klasifikasi multi-kelas. Secara keseluruhan, hasil penelitian menunjukkan bahwa model analisis sentimen yang digunakan cukup efektif dalam mengidentifikasi pola opini publik terhadap kebijakan penetapan Bencana Nasional Aceh, meskipun masih diperlukan pengembangan lebih lanjut untuk meningkatkan pemahaman konteks dan nuansa bahasa pada sentimen netral dan positif.

Kata Kunci: Sentiment Analysis; Public Opinion; Government Policy; National Disaster in Aceh; Machine Learning, Confusion Matrix; ROC Curve

Abstract—This study aims to analyze public opinion sentiment toward the Indonesian government's policy on the designation of Aceh as a National Disaster by utilizing machine learning techniques on online review data. Sentiment labeling was conducted automatically based on user rating scores, where ratings of 1–2 were classified as negative, rating 3 as neutral, and ratings of 4–5 as positive. The labeling results indicate that the dataset is dominated by negative sentiment with 5,593 reviews, followed by 1,956 neutral reviews and 1,460 positive reviews, reflecting the prevalence of critical public responses to the policy. Model performance was evaluated using cross-validation, confusion matrix analysis, and the Receiver Operating Characteristic (ROC) curve. The cross-validation results show an average accuracy of 80.62% with a low standard deviation, indicating stable and consistent model performance. The confusion matrix analysis reveals that the model performs very well in classifying negative sentiment, but still faces challenges in distinguishing neutral and positive sentiments due to linguistic ambiguity and overlapping contextual expressions. Furthermore, the ROC curve evaluation demonstrates good discriminative ability, with AUC values of 0.84 for the negative class and 0.82 for both neutral and positive classes. Overall, the results indicate that the proposed sentiment analysis model is effective in capturing public opinion patterns regarding the National Disaster designation policy in Aceh, although further improvements are required to enhance contextual understanding, particularly for neutral and positive sentiments.

Keywords: Sentiment Analysis; Public Opinion; Government Policy; Aceh National Disaster; Machine Learning; Confusion Matrix; ROC Curve

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi di era digital telah mendorong masyarakat untuk semakin aktif menyampaikan opini dan tanggapan terhadap berbagai kebijakan pemerintah melalui platform digital. Salah satu media yang memiliki pengaruh besar dalam pembentukan opini publik adalah YouTube. Platform ini tidak hanya digunakan sebagai media berbagi video, tetapi juga sebagai ruang diskusi publik melalui kolom komentar, di mana masyarakat dapat mengekspresikan pandangan, dukungan, maupun kritik terhadap isu-isu nasional, termasuk kebijakan pemerintah terkait penetapan bencana nasional [1].

Aceh merupakan wilayah yang memiliki tingkat kerentanan tinggi terhadap bencana alam. Ketika pemerintah menetapkan suatu peristiwa bencana di Aceh sebagai bencana nasional, kebijakan tersebut memicu berbagai respons dari masyarakat yang terefleksi dalam komentar pada video berita, siaran pers, dan konten diskusi terkait di YouTube. Opini publik yang muncul mencerminkan persepsi masyarakat terhadap kinerja pemerintah, mulai dari kecepatan penanganan, transparansi informasi, hingga efektivitas distribusi bantuan kepada masyarakat terdampak. Namun, data komentar YouTube memiliki karakteristik yang tidak terstruktur, bersifat bebas, serta mengandung bahasa informal, singkatan, emotikon, dan ekspresi emosional yang beragam. Selain itu, jumlah komentar yang besar dan terus bertambah menyebabkan proses analisis secara manual menjadi tidak efektif, memakan waktu, dan rentan terhadap subjektivitas. Oleh karena itu, diperlukan metode analisis otomatis yang mampu mengolah data teks komentar YouTube secara cepat, akurat, dan objektif. Machine Learning menawarkan solusi untuk menganalisis opini publik dalam jumlah besar melalui teknik analisis sentimen. Analisis sentimen bertujuan untuk mengklasifikasikan opini masyarakat ke dalam kategori



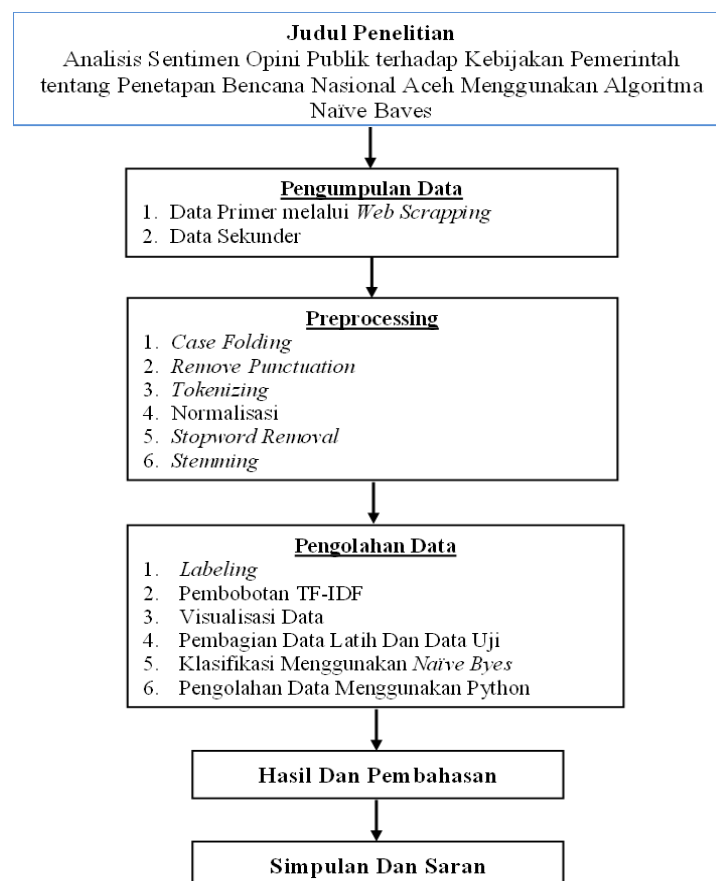
sentimen positif, negatif, atau netral. Salah satu algoritma yang banyak digunakan dalam klasifikasi sentimen teks adalah Naïve Bayes, yaitu metode pembelajaran mesin berbasis probabilistik yang sederhana namun efektif. Algoritma Naïve Bayes memiliki keunggulan berupa proses komputasi yang cepat, mudah diimplementasikan, serta mampu memberikan performa yang baik pada data teks tidak terstruktur [2] [3] [4], termasuk komentar YouTube. Pengambilan data dalam penelitian ini dilakukan dengan cara melakukan crawling komentar YouTube menggunakan YouTube Data API. Data yang dikumpulkan berasal dari komentar pada video-video yang membahas kebijakan pemerintah terkait penetapan bencana nasional Aceh, seperti video berita dari kanal resmi media, pernyataan pemerintah, serta konten diskusi publik.

Data komentar yang diperoleh selanjutnya melalui tahap praproses teks, meliputi pembersihan data, normalisasi kata, tokenisasi, stopword removal, dan stemming, sebelum dilakukan proses klasifikasi sentimen menggunakan algoritma Naïve Bayes. Metode ini dipilih karena bersifat mudah, kuat serta berdiri sendiri [5], [6], [7]. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen opini publik pada platform YouTube terhadap kebijakan pemerintah mengenai penetapan bencana nasional Aceh menggunakan algoritma Naïve Bayes. Penelitian terkait yang pernah dilakukan oleh Farook dkk [8], tentang sentimen masyarakat terhadap calon presiden Indonesia 2024 berdasarkan opini dari Twitter, metodenya dengan Metode Naive Bayes, hasil dari pengamatan polaritas sentiment masyarakat terhadap pasangan calon presiden dan wakilpresiden menunjukkan pasangan Prabowo Subianto – Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 6,4% sentimen negatif dan 25,9% sentimen netral. Sedangkan pasangan Joko Widodo – Jusuf Kalla mendapatkan total 37,6% sentimen positif, 34,4% sentimen negatif, dan 27,9 sentimen netral [9]. Penelitian berikutnya oleh Dedi Darwis, dkk tentang analisis sentimen *review* data *twitter* bmkg nasional, metodenya Naïve Bayes, hasil penelitian tingkat akurasi berdasarkan pengujian yang dilakukan adalah 68,97%. Sentimen analisis yang dilakukan adalah membuat tiga kelas yaitu positif, negatif dan netral menggunakan algoritma klasifikasi Naive Bayes. Hasil penelitian diharapkan mampu memberikan gambaran yang objektif mengenai kecenderungan sentimen masyarakat serta menjadi bahan evaluasi bagi pemerintah dalam meningkatkan efektivitas komunikasi kebijakan dan strategi penanganan bencana. Selain itu, penelitian ini juga diharapkan dapat membuktikan efektivitas algoritma Naïve Bayes dalam klasifikasi sentimen komentar YouTube pada konteks kebijakan publik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam melakukan penelitian ini, terdapat pola dalam berpikir untuk menentukan bagaimana alur penelitian yang akan dilakukan oleh penulis, adapun beberapa tahapan yaitu dapat dilihat pada Gambar 1 di bawah ini:



Gambar 1. Kerangka Berfikir



2.2 Algoritma Naive Bayes

Algoritma naive bayes merupakan salah satu algoritma dalam teknologi klasifikasi yang mudah diimplementasikan dan cepat dalam kecepatan pemrosesan. Algoritma naive bayes didasarkan pada probabilitas bersyarat [10]. ini menggunakan teorema bayes, rumus itu menghitung probabilitas dengan menghitung frekuensi nilai dan kombinasi nilai dalam sejarah data. teorema bayes menemukan probabilitas suatu peristiwa terjadi mengingat probabilitas peristiwa lain yang telah terjadi. metode naive bayes menggunakan model statistik untuk melakukan proses klasifikasi data. metode ini menghitung nilai *probabilitas* data uji berdasarkan data kasus yang sudah pernah terjadi [11].

Algoritma naive bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema bayes (atau aturan bayes) dengan asumsi *independensi* (ketidak tergantungan) yang kuat (naif) [12]. dengan kata lain, dalam *naive bayes*, model yang digunakan adalah “*model fitur independen*”. dalam bayes (terutama *naive bayes*), maksud *independensi* yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama metode naive bayes merupakan metode yang menggunakan probabilitas dan statistik yang dikemukakan oleh seorang ilmuwan inggris thomas bayes yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya [13], [14], [15].

2.3 Metode Pengujian

Penulis menggunakan metode pengujian library *Confusion Matrix* yang terdapat pada metode Naive Bayes. Metode pengujian ini digunakan untuk mengukur performa pada metode klasifikasi. Pada umumnya terdapat istilah dalam penggunaan metode *Confusion Matrix* yaitu sebagai berikut:

a. Akurasi adalah tingkat kedekatan nilai sebenarnya yang didapatkan. Berikut adalah rumus untuk menentukan akurasi:

$$Akurasi = \frac{TN+TP}{TN+FN+TP+FP} \quad (1)$$

Keterangan: TN = *True Negative*, TP = *True Positive*, FN = *False Negativ*, TN = *True Negative*

b. Presisi merupakan sebuah tingkat kecocokan dari data yang diambil dengan informasi yang ada. Berikut adalah rumus untuk menentukan presisi:

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall* adalah tingkat kedekatan dari nilai sebenarnya dengan nilai yang didapatkan. Berikut rumus untuk menghitung *Recall*:

$$Recall (N) = \frac{TN}{TN+FN} \quad (3)$$

$$Recall (P) = \frac{TP}{TP+FP} \quad (4)$$

d. *F1 Score* merupakan suatu perbandingan rata-rata presisi dan *recall* yang telah dibobotkan. Berikut rumus untuk menghitung *F1 Score*:

$$F1 \text{ score}) = \frac{2 \times Recall \times Presisi}{Recall+Presisi} \quad (5)$$

Istilah yang merepresentasikan *Confussion Matrix* dapat dilihat pada tabel 1 berikut ini :

Tabel 1. *Confussion Matrix*

Aktual	Predicted Condition	
	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Matrik evaluasi yang digunakan untuk mengukur kinerja suatu model klasifikasi (terutama klasifikasi biner) dengan membandingkan nilai prediksi model (*Predicted Condition*) terhadap nilai sebenarnya (*Aktual*). *True Positive (TP)*: Kondisi ketika data aktual bernilai Positif dan model berhasil memprediksinya secara tepat sebagai Positif. Ini menandakan prediksi positif yang benar. *False Negative (FN)*: Kondisi ketika data aktual bernilai Positif, namun model salah memprediksinya sebagai Negatif. Ini biasa disebut sebagai kesalahan Tipe II (*Type II Error*) atau miss. *False Positive (FP)*: Kondisi ketika data aktual bernilai Negatif, tetapi model salah memprediksinya sebagai Positif. Ini dikenal sebagai kesalahan Tipe I (*Type I Error*) atau false alarm. *True Negative (TN)*: Kondisi ketika data aktual bernilai Negatif dan model dengan benar memprediksinya sebagai Negatif. Ini menandakan prediksi negatif yang tepat.

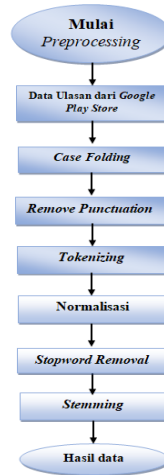
2.4 Pengelolaan Data

Dalam melakukan penelitian ini dilakukan pengolahan pada data ulasan pengguna dengan tahapan sebagai berikut:

a. Tahapan *Preprocessing*

Pada saat melakukan pengolahan data sebelumnya dilakukan tahapan *Praprocessing* menggunakan bahasa pemrograman Python yang terdiri dari beberapa tahapan seperti pengambilan data ulasan pengguna data ulasan berupa

komentar pengguna YouTube yang berkaitan dengan sentimen opini publik terhadap kebijakan pemerintah mengenai penetapan Bencana Nasional yang dilakukan dengan cara *Web Scraping* menggunakan Python. Setelah dilakukan pengambilan data, kemudian data ulasan tersebut akan diolah menggunakan menggunakan *library* Python. Lalu dilakukan proses *Case Folding*, *Tokenizing*, *Remove Punctuation*, *Normalisasi*, *Stopword Remove* dan *Stemming* yang digambarkan pada Gambar 2.



Gambar 2. Flowchart Tahap *Preprocessing*

b. Tahapan *Labelling*

Labelling atau pelabelan pada data dilakukan untuk menganalisa sebuah kata atau kalimat apakah termasuk kedalam kategori Positif, Netral atau Negatif. Pelabelan data dapat dilihat dari Tabel 2.

Tabel 2. *Labelling*

Rating	Sentiment
5	Positif
4	Positif
3	Netral
2	Negatif
1	Negatif

Selanjutnya dilakukan pembagian ke dalam tiga kelas sentimen yaitu jika data mendapat skor = 1 or skor = 2, maka termasuk kelas negatif, jika mendapat skor = 3, termasuk kelas netral dan jika mendapat skor = 4 or skor = 5, maka masuk pada kelas positif. Hasil *labeling* dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil *Labeling*

Score	Stemming	Label
0 5	['akses', 'cepat']	positif
1 3	['salah', 'masuk', 'dokumen', 'tolong', 'admin']	netral
2 1	['susah', 'registrasi']	negatif
3 2	['tolong', 'upgrade', 'verifikasi']	negatif
4 4	['keren', 'aplikasi']	Positif

Rating terhadap kelas sentimen adalah proses yang digunakan untuk menilai atau mengklasifikasikan seberapa positif, negatif, atau netral suatu teks atau komentar. Dalam konteks analisis sentimen, ini adalah cara untuk mengukur atau menggambarkan sejauh mana suatu teks atau pendapat menyatakan perasaan positif, negatif, atau tidak memiliki perasaan khusus (netral). Ini dapat diterjemahkan ke dalam skor numerik untuk mewakili sentimen secara lebih kuantitatif. Bagaimana rating terhadap kelas sentimen berkaitan dengan skor. Proses penilaian sentimen adalah subyektif, yang berarti itu bergantung pada penilaian manusia. Orang yang memberikan rating atau label positif, negatif, atau netral berdasarkan penafsiran mereka terhadap teks atau komentar. Biasanya, terdapat tiga kelas sentimen utama: "positif," "negatif," dan "netral." Teks diberi label sesuai dengan kelas sentimen yang sesuai dengan penelitian ini. Dari tabel hasil *labeling* didapatkan melalui fungsi yang dijalankan seperti Fungsi *analyze* mengambil satu parameter, yaitu *score*, yang mewakili skor numerik. Pertama, fungsi ini menguji apakah *score* sama dengan 1 atau 2. Jika benar, itu berarti skor tersebut dianggap "negatif," dan fungsi akan mengembalikan string 'negatif'. Jika *score* sama dengan 3, maka itu dianggap "netral," dan fungsi akan mengembalikan string 'netral'. Terakhir, jika *score* sama dengan 4 atau 5, itu dianggap "positif," dan fungsi akan mengembalikan string 'positif'. Sedangkan hubungan antara rating terhadap kelas sentimen dan skor adalah bahwa rating adalah representasi kualitatif dari sentimen, sedangkan skor adalah representasi kuantitatif. Rating memutuskan ke arah mana sentimen tersebut condong (positif, negatif,



atau netral), sedangkan skor memberikan nilai lebih spesifik tentang sejauh mana sentimen tersebut condong. Hubungan ini memungkinkan kita untuk mengukur dan memahami sentimen dalam teks atau data komentar dengan lebih terstruktur dan kuantitatif. Dengan konversi dari rating ke skor, analisis sentimen dapat menjadi lebih objektif dan dapat digunakan dalam berbagai aplikasi, salah satunya dalam melakukan analisis sentimen terhadap aplikasi Identitas Kependudukan Digital.

c. Tahap Pembobotan TF-IDF

Selanjutnya menggunakan algoritma *Naïve Bayes* sebagai metode klasifikasi data latih dan data uji dengan melakukan pembobotan menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Metode TF-IDF adalah metode untuk menentukan frekuensi relatif setiap kata atau karakter dimana setiap kata diberi bobot menurut kepentingannya dalam bentuk nilai apakah kata tersebut penting atau tidak, berdasarkan banyaknya kemunculan kata tersebut dalam suatu dokumen dan membandingkan kata-kata tersebut dengan keseluruhan dokumen yang ada [16].

d. Tahap *Cosine Similarity*

Cosine similarity adalah salah satu metode yang dapat digunakan untuk mengukur kemiripan antara dua dokumen. Metode ini bekerja dengan cara mengubah dokumen tersebut menjadi bentuk vektor dan kemudian menghitung sudut antara kedua vektor tersebut. Nilai cosine similarity berkisar antara 0 dan 1, di mana nilai 0 menunjukkan bahwa kedua dokumen tidak memiliki kemiripan sama sekali, dan nilai 1 menunjukkan bahwa kedua dokumen memiliki kemiripan yang sempurna [17].

e. Tahap Visualisasi Data

Setelah melakukan pelabelan, data akan diubah kedalam bentuk *chart* untuk visualisasi data yang lebih menarik dengan menggunakan *library* Matplotlib. Matplotlib adalah pustaka Python lintas platform untuk membuat plot 2D berkualitas tinggi. Matplotlib memudahkan pembuatan grafik, histogram, spektrum, bar chart dengan mudah dilakukan [18].

f. Tahap Pembagian Data Latih Dan Data Uji

Data yang telah dilakukan *Preprocessing* dan pelabelan kemudian dilakukan pembagian antara data latih dan data uji. Data latih merupakan data untuk membuat model klasifikasi, sedangkan Data uji digunakan untuk pengujian model yang ada. Selanjutnya dibagi menjadi data latih dan data uji untuk pembentukan model klasifikasi yang digunakan yaitu menggunakan dua nilai perbandingan sebagai berikut: data latih 70% dan 30% data uji, data latih 90% dan 10% data uji. Hasil split data dapat dilihat pada Tabel 4.

Tabel 4. Hasil Split Data

Rasio	Hasil Split Data	
	Data Training	Data Testing
70 : 30	2.800	1200
90 : 10	3.600	400

g. Penerapan Algoritma *Naïve Bayes*

Proses klasifikasi ini menggunakan perpustakaan *K-Nearest Neighbors* dengan python. Dataset yang terdiri dari data ulasan berupa komentar pengguna YouTube yang berkaitan dengan sentimen opini publik terhadap kebijakan pemerintah mengenai penetapan Bencana Nasional Aceh yang didapatkan melalui proses *web scrapping* melalui *video youtube* yang terdapat pada python menggunakan *google colab*, setelah itu dilakukan *preprocessing* untuk menyiapkan data yang akan dilakukan proses selanjutnya dan diberi bobot akan digunakan dalam proses klasifikasi. Data yang sudah dilakukan pembobotan kata sebanyak data kemudian dibagi menjadi data latih dan data uji dengan perbandingan 90% : 10%, dan 70% : 30%.

h. Penerapan *K-Fold Cross Validation*

Metode *K-Folds Cross-Validation* membagi data yang telah dibobot berdasarkan term menjadi K bagian, dengan K adalah ukuran dari term. K buah percobaan dilakukan dalam satu set percobaan, dengan menggunakan satu bagian sebagai data uji [19] [20].

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Pengumpulan Data Ulasan

Langkah awal penelitian adalah pengumpulan data ulasan komentar Sentimen Opini Publik terhadap Kebijakan Pemerintah tentang Penetapan Bencana Nasional Aceh. Proses ini dilakukan dengan memanfaatkan pustaka komentar youtube, yang berhasil mengumpulkan sebanyak 9003 ulasan komentar berbahasa Indonesia. Data yang terkumpul kemudian disimpan dalam format CSV untuk memfasilitasi tahapan analisis selanjutnya.

3.1.2 Pelabelan Sentimen

Setelah data ulasan berhasil dikumpulkan, tahap selanjutnya adalah pelabelan sentimen. Proses pelabelan ini dilakukan secara otomatis berdasarkan skor (rating) yang diberikan oleh pengguna. Ulasan dengan skor 4 atau 5 dikategorikan sebagai 'positif', ulasan dengan skor 1 atau 2 sebagai 'negatif', dan ulasan dengan skor 3 sebagai 'netral'. Berdasarkan

proses pelabelan ini, didapatkan distribusi sentimen sebagai berikut: terdapat 5593 ulasan berlabel 'negatif', 1956 ulasan berlabel 'netral', dan 1460 ulasan berlabel 'positif'. Distribusi ini mengindikasikan bahwa ulasan negatif memiliki proporsi yang paling dominan dalam dataset yang terkumpul.

3.1.3 Pra-pemrosesan Teks

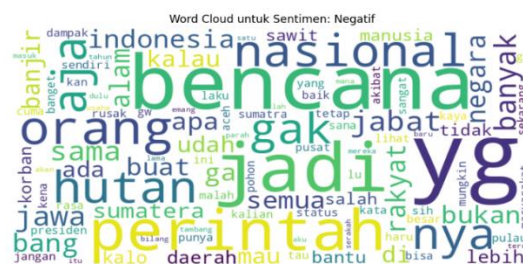
Untuk mempersiapkan data teks ulasan agar siap untuk analisis lebih lanjut, serangkaian tahapan pra-pemrosesan telah dilakukan. Tahapan ini meliputi: case folding untuk mengubah seluruh teks menjadi huruf kecil; penghapusan angka dan tanda baca untuk membersihkan teks dari karakter non-alfabetik; penghapusan stopwords menggunakan daftar kata henti bahasa Indonesia dari NLTK untuk menghilangkan kata-kata umum yang tidak relevan dengan sentimen; dan stemming dengan Sastrawi Stemmer untuk mengubah kata menjadi bentuk dasarnya. Proses pra-pemrosesan ini menghasilkan teks yang bersih dan distandarisasi pada kolom 'clean_text', yang kemudian digunakan untuk ekstraksi fitur.

3.1.4 Ekstraksi Fitur dengan TF-IDF

Setelah pra-pemrosesan, teks ulasan diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Tfidf Vectorizer digunakan untuk menghasilkan vektor fitur dari kolom 'clean_text'. Dalam implementasi ini, max_features diatur sebesar 5000, yang berarti hanya 5000 fitur (kata) dengan bobot TF-IDF tertinggi yang akan dipertimbangkan. Pemilihan fitur ini bertujuan untuk mengurangi dimensi data sambil tetap mempertahankan informasi yang paling relevan untuk klasifikasi sentimen.

3.1.5 Visualisasi Word Cloud

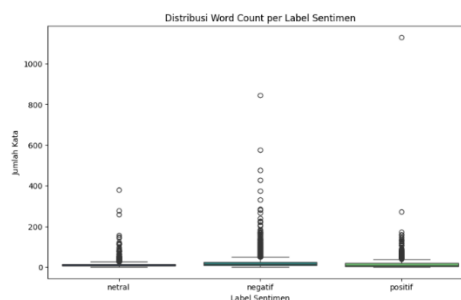
Untuk mendapatkan gambaran visual mengenai kata-kata yang paling sering muncul dalam setiap kategori sentimen setelah pra-pemrosesan, word cloud telah dibuat. Gambar 4.1 menampilkan tiga word cloud terpisah untuk sentimen positif, netral, dan negatif. Pada word cloud sentimen positif, kata-kata seperti "bagus", "aplikasi", "masuk", dan "terima kasih" menonjol, merefleksikan aspek-aspek yang diapresiasi pengguna. Untuk sentimen netral, kata-kata seperti "ngelag", "bug", "suka", "update", dan "error" sering terlihat, mengindikasikan komentar mengenai pengalaman teknis atau performa. Sementara itu, word cloud sentimen negatif didominasi oleh kata-kata seperti "gak", "login", "akun", "susah", "aplikasi", "masalah", dan "uang", secara jelas menggambarkan keluhan dan frustrasi pengguna terhadap berbagai isu. Visualisasi Word Cloud dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Word Cloud

3.1.6 Distribusi Panjang Komentar per Sentimen

Analisis terhadap distribusi panjang komentar (jumlah kata) untuk setiap kategori sentimen divisualisasikan menggunakan boxplot. Gambar 4.2 menyajikan Distribusi Panjang Komentar per Sentimen, menunjukkan perbandingan distribusi panjang komentar antara sentimen negatif, positif, dan netral. Dari boxplot tersebut, terlihat bahwa ulasan dengan sentimen negatif cenderung memiliki rentang panjang komentar yang lebih luas, dengan median sekitar 7 kata dan beberapa outlier yang menunjukkan ulasan sangat panjang (di atas 25 kata). Sebaliknya, ulasan sentimen positif memiliki median panjang komentar yang lebih pendek, sekitar 4-5 kata, dengan rentang yang lebih sempit. Ulasan netral menunjukkan median panjang komentar yang mirip dengan negatif, sekitar 7 kata, namun dengan outlier yang lebih sedikit dibandingkan sentimen negatif. Distribusi Panjang Komentar per Sentimen dapat dilihat pada Gambar 4 berikut ini :



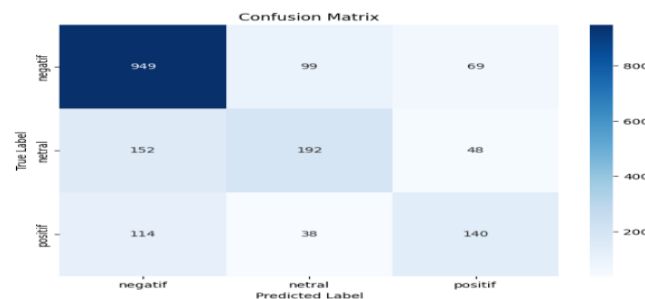
Gambar 4. Distribusi Panjang Komentar Per Sentimen

3.1.7 Pelatihan dan Evaluasi Model Klasifikasi

Data yang telah diekstraksi fiturnya dibagi menjadi 80% data latih dan 20% data uji menggunakan `train_test_split` dengan stratifikasi untuk memastikan proporsi kelas sentimen yang sama di kedua set data. klasifikasi sentimen kemudian dilatih Naive Bayes (MultinomialNB), Gambar 5 menampilkan Classification Report untuk keempat model, dan Gambar 6 menampilkan Confusion Matrix untuk model Nive Bayes. Classification Report pada masing-masing classifier terdapat pada Gambar 5 berikut.

Classification Report:				
	precision	recall	f1-score	support
negatif	0.78	0.85	0.81	1117
netral	0.58	0.49	0.53	392
positif	0.54	0.48	0.51	292
accuracy			0.71	1801
macro avg	0.64	0.61	0.62	1801
weighted avg	0.70	0.71	0.70	1801

Gambar 5. Classification Report Pada Masing-Masing Classifier



Gambar 6. Confusion Matrix Pada Masing-Masing Classifier

Hasil evaluasi menunjukkan kinerja yang sangat baik dari sebagian besar model. Berikut penjabaran hasil grafik Confusion Matrix pada analisis Sentimen Opini Publik terhadap Kebijakan Pemerintah tentang Penetapan Bencana Nasional Aceh: Confusion Matrix tersebut menunjukkan kemampuan model dalam mengklasifikasikan opini publik ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Pada sentimen negatif, model berhasil mengklasifikasikan dengan sangat baik, terlihat dari nilai 949 data negatif yang diprediksi dengan benar sebagai negatif. Namun, masih terdapat kesalahan prediksi, yaitu 99 data negatif yang diprediksi sebagai netral dan 69 data negatif yang diprediksi sebagai positif. Hal ini menunjukkan bahwa meskipun dominan akurat, sebagian opini negatif memiliki kemiripan pola bahasa dengan sentimen netral dan positif. Untuk sentimen netral, performa model relatif lebih rendah. Dari total data netral, hanya 192 data yang berhasil diprediksi dengan benar. Sebanyak 152 data netral keliru diprediksi sebagai negatif dan 48 data sebagai positif. Kondisi ini mengindikasikan bahwa sentimen netral cenderung ambigu dan sulit dibedakan, terutama dengan sentimen negatif dalam konteks kebijakan pemerintah dan isu bencana nasional. Pada sentimen positif, model mampu mengklasifikasikan 140 data dengan benar. Namun, terdapat 114 data positif yang salah diprediksi sebagai negatif dan 38 data sebagai netral. Hal ini menunjukkan bahwa sebagian opini positif memiliki ekspresi bahasa yang kurang tegas atau bercampur dengan kritik, sehingga sering disalahartikan sebagai sentimen negatif.

Secara keseluruhan, Confusion Matrix memperlihatkan bahwa sentimen negatif merupakan kelas yang paling dominan dan paling akurat dikenali oleh model, yang mencerminkan kuatnya respons kritis masyarakat terhadap kebijakan penetapan Bencana Nasional Aceh. Sementara itu, tantangan utama model terletak pada membedakan sentimen netral dan positif, yang memiliki karakteristik bahasa lebih halus dan tumpang tindih. Hasil ini dapat menjadi bahan evaluasi untuk peningkatan model analisis sentimen, khususnya dalam memahami konteks kebijakan publik dan isu kebencanaan.

3.1.8 Validasi Silang (Cross-Validation)

Untuk mendapatkan estimasi kinerja model yang lebih robust dan mengurangi potensi bias dari satu kali pembagian data, dilakukan 5-Fold Cross Validation untuk setiap model. Hasilnya menunjukkan konsistensi kinerja yang luar biasa pada model-model tertentu. Tabel 5 merupakan ringkasan hasil 5-Fold Cross Validation untuk setiap model yang tersaji:

Tabel 5. Hasil 1-Fold Cross Validation Model Klasifikasi Sentimen

Model	Akurasi per Fold			Rata-rata Akurasi	Standar Deviasi
Naive	0.81347729	0.79523455	0.80379747		
Bayes	0.81340782	0.80484171		0.8062	0.0068

Berikut penjelasan dari hasil Cross-Validation Model Klasifikasi Sentimen yang ditampilkan: Hasil *cross-validation scores* menunjukkan bahwa akurasi model pada setiap fold berada pada rentang 79,52% hingga 81,35%. Nilai ini relatif konsisten di semua fold, yang menandakan bahwa model memiliki kinerja yang stabil saat diuji pada berbagai pembagian data latih dan data uji. Tidak terdapat perbedaan akurasi yang ekstrem antar fold, sehingga model tidak terlalu bergantung pada satu subset data tertentu. Nilai mean accuracy sebesar 0,8062 (80,62%) mengindikasikan bahwa secara rata-rata model mampu mengklasifikasikan sentimen dengan tingkat ketepatan yang cukup baik. Angka ini menunjukkan bahwa lebih dari 80% data sentimen berhasil diprediksi dengan benar, sehingga model sudah layak digunakan untuk analisis sentimen opini publik. Sementara itu, standard deviation sebesar 0,0068 tergolong sangat kecil. Hal ini menunjukkan bahwa variasi akurasi antar fold sangat rendah, sehingga performa model dapat dikatakan konsisten dan robust. Dengan kata lain, model tidak mengalami *overfitting* maupun *underfitting* yang signifikan dan mampu melakukan generalisasi dengan baik terhadap data baru.

3.1.9 Kurva ROC (Receiver Operating Characteristic)

a. ROC Curve Kelas Negatif (AUC = 0,84)

Kurva ROC untuk kelas negatif menunjukkan performa terbaik dibandingkan kelas lainnya. Nilai AUC sebesar 0,84 menandakan bahwa model memiliki kemampuan yang baik dalam membedakan data sentimen negatif dari kelas lain. Kurva yang relatif lebih dekat ke sudut kiri atas grafik menunjukkan bahwa model mampu mencapai True Positive Rate (TPR) yang tinggi dengan False Positive Rate (FPR) yang rendah, sehingga prediksi sentimen negatif tergolong akurat dan konsisten.

b. ROC Curve Kelas Netral (AUC = 0,82)

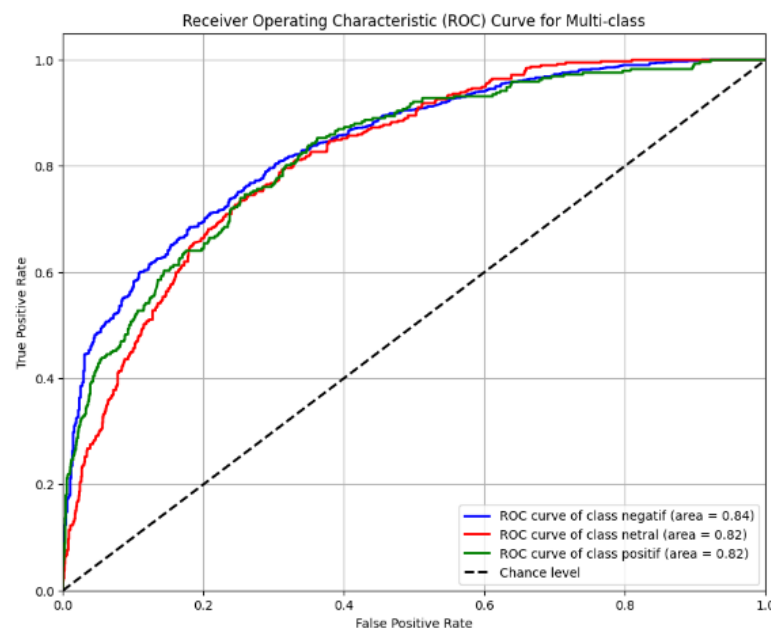
Untuk kelas netral, nilai AUC sebesar 0,82 mengindikasikan bahwa model juga memiliki kinerja cukup baik dalam mengklasifikasikan sentimen netral. Kurva ROC berada jauh di atas garis diagonal (chance level), yang berarti performa model jauh lebih baik dibandingkan klasifikasi acak. Namun, jika dibandingkan dengan kelas negatif, kemampuan pemisahan kelas netral sedikit lebih rendah, kemungkinan karena karakteristik data netral yang lebih ambigu.

c. ROC Curve Kelas Positif (AUC = 0,82)

Kurva ROC untuk kelas positif juga menghasilkan AUC sebesar 0,82, yang menunjukkan bahwa model cukup efektif dalam mengenali sentimen positif. Pola kurva hampir serupa dengan kelas netral, menandakan tingkat kesulitan klasifikasi yang relatif sama. Model mampu menyeimbangkan antara TPR dan FPR dengan baik, meskipun masih terdapat potensi kesalahan pada area ambang tertentu.

3.1.10 Perbandingan dengan Garis Chance Level

Garis diagonal putus-putus merepresentasikan klasifikasi acak (AUC = 0,5). Seluruh kurva ROC untuk ketiga kelas berada jauh di atas garis ini, yang berarti model klasifikasi sentimen yang digunakan memiliki kemampuan diskriminasi yang signifikan dan andal. Gambar 7 menampilkan ROC Curve pada masing-masing classifier:



Gambar 7. ROC Curve Pada Masing-Masing Classifier

3.1.11 Pengujian Interaktif Model

Untuk mendemonstrasikan kapabilitas model yang telah dilatih dalam mengklasifikasikan sentimen secara real-time atau berdasarkan masukan manual, dilakukan pengujian interaktif. Dalam pengujian ini, pengguna dapat memasukkan

komentar secara langsung, dan model akan memprediksi sentimen komentar tersebut (positif, negatif, atau netral). Gambar 4.6 menunjukkan contoh hasil pengujian interaktif model. Misalnya, ketika dimasukkan komentar "aplikasi yang sangat bagus, berinteraksi dengan sesama menjadi sangat mudah", model berhasil memprediksi sentimennya sebagai POSITIF. Komentar "kenapa twitter sekarang lebih banyak ujaran kebenciannya" diprediksi NEGATIF. Menariknya, komentar "aplikasi yang biasa saja" dan "biasa saja" juga diprediksi NEGATIF dan POSITIF secara berturut-turut, menunjukkan bahwa model mampu menangkap nuansa sentimen meskipun kalimatnya singkat. Pengujian ini memberikan validasi visual tentang bagaimana model dapat beroperasi pada data baru, di luar dataset pelatihan dan pengujian formal, dan menunjukkan bahwa model yang telah dibangun memiliki kemampuan praktis dalam klasifikasi sentimen ulasan. Gambar 8 merupakan pengujian interaktif model:

```

while True:
    comment = input("Masukkan komentar Anda (ketik 'exit' untuk keluar): ")
    if comment.lower() == 'exit':
        break

    # Preprocessing komentar
    clean_comment = clean_text(comment)

    # Vectorize komentar
    comment_vectorized = tfidf_vectorizer.transform([clean_comment])

    # Prediksi
    predicted_label_encoded = naive_bayes_model.predict(comment_vectorized)
    predicted_label = label_encoder.inverse_transform(predicted_label_encoded)

    print(f"Sentimen diprediksi: {predicted_label[0]}\n")

print("Program selesai.")

```

... Masukkan komentar Anda (ketik 'exit' untuk keluar): exit
Program selesai.

Gambar 8. Pengujian Interaktif Model

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode Naive Bayes mampu digunakan untuk melakukan analisis sentimen terhadap opini publik mengenai kebijakan pemerintah terkait penetapan Bencana Nasional Aceh pada platform YouTube. Proses penelitian dimulai dari pengumpulan data sebanyak 9.003 komentar, kemudian dilakukan tahapan pra-pemrosesan teks yang meliputi case folding, cleaning, stopword removal, dan stemming sehingga menghasilkan data yang lebih bersih dan siap untuk dianalisis. Selanjutnya, fitur teks diekstraksi menggunakan metode TF-IDF dan diklasifikasikan menggunakan algoritma Naive Bayes untuk menentukan kategori sentimen negatif, netral, dan positif. Hasil penelitian menunjukkan bahwa sentimen negatif merupakan kategori yang paling dominan dalam opini publik yang dianalisis. Berdasarkan evaluasi model, algoritma Naive Bayes memperoleh tingkat akurasi sebesar 80,62% dengan nilai standar deviasi 0,0068 pada pengujian 5-Fold Cross Validation, yang menunjukkan bahwa model memiliki performa yang stabil dan mampu melakukan generalisasi dengan baik terhadap data baru. Selain itu, hasil pengujian menggunakan Confusion Matrix dan kurva ROC memperlihatkan bahwa model memiliki kemampuan klasifikasi yang cukup baik, terutama dalam mengenali sentimen negatif dengan nilai AUC mencapai 0,84, sedangkan sentimen netral dan positif masing-masing memperoleh nilai AUC sebesar 0,82. Hasil tersebut menunjukkan bahwa model mampu membedakan sentimen publik secara efektif meskipun masih terdapat kesalahan klasifikasi pada sentimen netral dan positif yang memiliki karakteristik bahasa lebih ambigu. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, seperti ketidakseimbangan jumlah data pada setiap kelas sentimen serta penggunaan satu algoritma klasifikasi saja. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan teknik penyeimbangan data, menambah jumlah dataset, serta membandingkan performa beberapa algoritma klasifikasi lain seperti Support Vector Machine, Random Forest, atau metode berbasis Deep Learning guna memperoleh hasil analisis sentimen yang lebih akurat dan optimal. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi dalam memahami respons masyarakat terhadap kebijakan pemerintah melalui media sosial serta mendukung pengambilan keputusan yang lebih tepat berdasarkan opini publik.

REFERENCES

- [1] Y. Yadi, "Analisa Usability Pada Website Traveloka," *Jurnal Ilmiah Betrik*, vol. 9, no. 03, pp. 172–180, 2018, doi: 10.36050/betrik.v9i03.43.
- [2] M. Y. Hidayat, "Analisis Sentimen Terhadap Pidato Presiden 2026 Pada Aplikasi Tiktok dan X Menggunakan Metode Naive Bayes Classifier," *SAINTEKES: Jurnal Sains, Teknologi dan Kesehatan*, vol. 5, no. 1, pp. 15–23, 2026, doi: 10.55681/saintekes.v1i1.xxx.
- [3] I. W. A. Agetia, N. L. E. Armoni, and I. P. A. U. Irawan, "Evaluasi Kinerja Algoritma Naive Bayes, SVM, dan IndoBERT pada Analisis Sentimen Ulasan Pengguna Gojek Berbasis Text Mining," *TIN: Terapan Informatika Nusantara*, vol. 7, no. 1, pp. 89–98, 2026, doi: 10.47065/tin.v7i1.9617.



- [4] R. P. Sundawa, M. R. Zahran, and B. Irfansyah, "Perbandingan Metode Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi Roblox," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 6, no. 1, pp. 100–115, 2026, doi: 10.55606/jutiti.v6i1.6569.
- [5] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *Jurnal Penelitian Komunikasi Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.
- [6] I. Elly, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *Jurnal Khatulistiwa Informatika*, vol. VII, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.5740.
- [7] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *Jurnal Bumigora Information Technology (BITE)*, vol. 2, no. 2, pp. 105–112, 2020, doi: 10.30812/bite.v2i2.904.
- [8] D. A. Farook, M. Yusuf, and G. Syatau, "Analisis Sentimen Terhadap Popularitas Partai Pendukung Capres 2024 di Twitter Menggunakan Algoritma Naive Bayes Classifier," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 17, no. 2, pp. 216–226, 2024, doi: 10.35457/antivirus.v17i2.3261.216.
- [9] F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Teknologi & Informasi ITSmart*, vol. 2, no. 2, p. 35, 2016, doi: 10.20961/its.v2i2.630.
- [10] L. Lingga and F. N. Hasan, "Penerapan Naïve Bayes untuk Mengklasifikasikan Sentimen Tidak Seimbang pada Ulasan Aplikasi Berbasis Etika Konsumen," *Journal of Information System Research (JOSH)*, vol. 6, no. 4, pp. 2294–2306, 2025, doi: 10.47065/josh.v6i4.7867.
- [11] M. Iqbal, S. Andryana, and R. T. Komala Sari, "Visualisasi dan Analisa Data Penyebaran Covid-19 dengan Metode Klasifikasi Naïve Bayes," *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 5, no. 4, pp. 389–394, 2021, doi: 10.35870/jtik.v5i4.233.
- [12] F. M. Nizam, D. I. Sudahri, and E. Irawan, "Penerapan Metode Naive Bayes Dalam Menentukan Tingkat Kenyamanan Pada Rumah Sakit Terhadap Pasien," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 3, pp. 117–122, 2021, doi: 10.30865/klik.v2i3.297.
- [13] F. A. Nugroho, A. F. Solikin, M. D. Anggraini, and K. Kusri, "Sistem Pakar Diagnosa Virus Corona Dengan Metode Naïve Bayes," *Jurnal Teknologi Informasi dan Komunikasi (TIKomsin)*, vol. 9, no. 1, p. 81, 2021, doi: 10.30646/tikomsin.v9i1.553.
- [14] A. F. Watratan, A. P. B, and D. Moeis, "Implementation of the Naive Bayes Algorithm to Predict the Spread of Covid-19 in Indonesia," *Journal of Applied Computer Science and Technology*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.55606/jurritek.v1i1.127.
- [15] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [16] S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [17] R. T. Atmi, "Evaluation of The Library and Archives of East Java's Workshop as A Knowledge Sharing Form of Jawatimuran Local Wisdom Preservation," *Record and Library Journal*, vol. 1, no. 2, p. 172, 2016, doi: 10.20473/rlj.v1i2.1171.
- [18] F. David, "Visualisasi Data Dalam Bentuk 3 Dimensi Dengan Abstrak Seminar Nasional PIMIMD-5 , ITP , Padang," pp. 1–6, 2019, doi: 10.21063/PIMIMD5.2019.1.
- [19] M. Yoka Fathoni, S. Wijayanto, "Forecasting Penjualan Gas LPG di Toko Sembako Menggunakan Metode Fuzzy Time Series," *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)*, vol. 13, no. 2, pp. 87–96, 2021, doi: 10.5281/3541.jupiter.2021.10.
- [20] R. N. Irawan, K. M. Hindrayani, "Penerapan Cross Validation sebagai Analisis Sentimen Pelayanan Publik Kereta Api Lokal Daop 8 Menggunakan Metode Multinomial Naïve Bayes," *G-Tech : Jurnal Teknologi Terapan*, vol. 6, no. 2, pp. 295–305, 2022, doi: 10.33379/gtech.v8i2.4117.