

Analisis Sentimen Masyarakat Terhadap Program Gratis Pol Di TikTok Menggunakan Algoritma *Naive bayes*

Nandhita Helda Widayani^{1*}, Eka Arriyanti², Yulindawati³

¹ Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda Indonesia

² Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda Indonesia

³ Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda Indonesia

Email: ¹12243006@wicida.ac.id, ²ekaarry@wicida.ac.id, ³Yulindawati@wicida.ac.id

(* : 2243006@wicida.ac.id)

Abstrak - Program Gratis Pol merupakan salah satu program pendidikan di Kalimantan Timur yang memperoleh perhatian masyarakat dan menimbulkan beragam tanggapan pada media sosial, khususnya TikTok. Karakteristik komentar TikTok yang banyak menggunakan bahasa informal, singkatan, dan variasi bahasa sehari-hari menjadi tantangan dalam proses identifikasi sentimen sehingga diperlukan metode analisis yang mampu mengklasifikasikan opini masyarakat secara sistematis. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Program Gratis Pol berdasarkan komentar pengguna TikTok menggunakan algoritma *Naive Bayes*. Penelitian dilakukan melalui tahapan *text preprocessing*, pelabelan sentimen menggunakan pendekatan lexicon-based, representasi fitur menggunakan TF-IDF, serta proses klasifikasi menggunakan algoritma *Naive Bayes*. Data penelitian diperoleh dari 13 video TikTok terpilih dengan total 1.528 komentar yang dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi dengan jumlah 722 komentar, diikuti 496 komentar netral dan 310 komentar negatif. Model klasifikasi menghasilkan akurasi sebesar 56% dengan nilai macro average F1-score sebesar 0,44 dan weighted average F1-score sebesar 0,48. Penelitian ini memberikan kontribusi dalam menggambarkan persepsi masyarakat terhadap Program Gratis Pol serta menunjukkan penerapan algoritma *Naive Bayes* pada analisis sentimen komentar media sosial yang memiliki karakteristik bahasa tidak terstruktur

Kata Kunci: Analisis Sentimen, TikTok, *Naive bayes*, TF-IDF, Text Mining.

Abstract - The Gratis Pol Program is an educational initiative in East Kalimantan that has garnered public attention and sparked a wide range of reactions on social media, particularly TikTok. The characteristic use of informal language, abbreviations, and colloquial expressions in TikTok comments poses a challenge for sentiment analysis, necessitating a method capable of systematically classifying public opinion. This study aims to analyze public sentiment toward the Gratis Pol Program based on TikTok user comments using the *Naive Bayes* algorithm. The research was conducted through the stages of text preprocessing, sentiment labeling using a lexicon-based approach, feature representation using TF-IDF, and the classification process using the *Naive Bayes* algorithm. The research data was obtained from 13 selected TikTok videos with a total of 1,528 comments, divided into 80% training data and 20% test data. The results show that positive sentiment dominates with 722 comments, followed by 496 neutral comments and 310 negative comments. The classification model achieved an accuracy of 56%, with a macro average F1-score of 0.44 and a weighted average F1-score of 0.48. This study contributes to understanding public perception of the Gratis Pol Program and demonstrates the application of the *Naive Bayes* algorithm in analyzing the sentiment of social media comments that possess certain characteristics.

Keywords: Sentiment Analysis, TikTok, *Naive bayes*, TF-IDF, Text Mining.

1. PENDAHULUAN

Kebijakan publik merupakan instrumen strategis yang digunakan pemerintah untuk mengatasi berbagai permasalahan sosial, ekonomi, dan politik yang dihadapi masyarakat. Definisi kebijakan publik telah berkembang seiring waktu, mencerminkan kompleksitas dan dinamika dalam proses pembuatan keputusan pemerintahan[1]. Dalam implementasinya, kebijakan pemerintah diwujudkan melalui berbagai program pembangunan yang dirancang untuk mencapai tujuan jangka panjang. Salah satu program yang menjadi bagian dari kebijakan tersebut adalah program Gratis Poll di Provinsi Kalimantan Timur. Gratis Poll merupakan program utama pembangunan berbagai sarana pendidikan dapat disebut starting point dari sebuah perencanaan untuk tujuan jangka panjang. Begitu pula di aspek penataan, telah jelas pembagian kerja masing-masing di tingkat organisasi perangkat daerah (OPD) dan satuan tugas (*task force*) dalam kesatuan kendali oleh Gubernur Kalimantan Timur[2].

Program Gratis Pol merupakan program pendidikan yang dicanangkan Pemerintah Provinsi Kalimantan Timur untuk meningkatkan akses dan pemerataan pendidikan melalui pemberian fasilitas pendidikan bagi masyarakat. Program ini memperoleh perhatian masyarakat sehingga menimbulkan berbagai tanggapan dan opini pada media sosial, khususnya TikTok. Beragam tanggapan yang muncul, baik berupa sentimen positif, netral, maupun negatif, dapat mencerminkan

persepsi masyarakat terhadap implementasi program tersebut. Informasi mengenai kecenderungan persepsi masyarakat menjadi penting untuk diketahui karena dapat memberikan gambaran mengenai penerimaan masyarakat terhadap program yang dijalankan. Selain itu, komentar pada media sosial memiliki karakteristik bahasa yang beragam, seperti penggunaan bahasa informal, singkatan, serta variasi bahasa sehari-hari sehingga diperlukan metode analisis yang mampu mengidentifikasi sentimen secara lebih sistematis dan objektif. Untuk mengatasi permasalahan tersebut, diperlukan suatu metode yang mampu mengidentifikasi dan mengklasifikasikan opini masyarakat secara sistematis. Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu metode yang digunakan untuk mengidentifikasi opini, emosi, dan persepsi masyarakat terhadap suatu isu berdasarkan data teks. Dalam proses analisis sentimen, terdapat berbagai algoritma yang dapat digunakan untuk melakukan klasifikasi data teks, salah satunya adalah *Naive Bayes*. Algoritma ini dikenal memiliki kinerja yang baik serta efisien dalam mengolah data, sehingga banyak digunakan dalam penelitian analisis sentimen.

Naive Bayes adalah salah satu metode yang populer dalam data mining karena kesederhanaannya dalam penggunaan serta efisiensi dalam proses pengolahan data[3]. Metode ini juga banyak digunakan dalam machine learning dan data mining karena tingkat efektivitas dan efisiensinya. Walaupun didasarkan pada asumsi independensi antar atribut, metode ini tetap mampu menghasilkan performa klasifikasi yang tinggi[4].

Analisis sentimen merupakan proses untuk mengolah suatu data teks yang akan dikategorikan ke dalam sentimen positif dan sentimen negatif [5]. Analisis sentimen bertujuan untuk mengklasifikasikan teks ulasan pelanggan ke dalam kategori positif, negatif, dan netral, serta mendeteksi ketidaksesuaian antara isi ulasan dengan rating yang diberikan[6]. Penelitian sebelumnya menunjukkan bahwa algoritma *Naive Bayes* efektif digunakan dalam analisis sentimen, Nilai akurasi pada *Naive Bayes* memberikan nilai yang tinggi sebesar 96%[7]. Selanjutnya, penelitian lain menunjukkan bahwa penggunaan algoritma *Naive Bayes* pada ulasan aplikasi BRImo menghasilkan akurasi sebesar 84,52% dengan dominasi sentimen positif sebesar 53,09% [8]. Selanjutnya, metode ini juga diterapkan pada analisis sentimen aplikasi Sentuh Tanahku dan menunjukkan bahwa mayoritas pengguna memberikan respon positif sebanyak 407 data dibandingkan sentimen negatif sebanyak 237 data, dengan tingkat akurasi sebesar 89%, precision 83%, dan recall 87%[9]. Selain itu, penelitian lain mengenai analisis sentimen komentar video mobil listrik di platform YouTube menggunakan metode *Naive Bayes* menunjukkan bahwa model mampu menghasilkan akurasi sebesar 70,69%, dengan precision 43,64% dan recall 39,48%[10]. Penelitian lain mengenai analisis sentimen komentar YouTube terhadap prediksi resesi ekonomi tahun 2023 menggunakan algoritma *Naive Bayes* menunjukkan bahwa mayoritas sentimen bersifat positif sebesar 40%, diikuti sentimen netral sebesar 34% dan negatif sebesar 26%[11].

Penelitian terdahulu telah banyak menerapkan algoritma *Naive Bayes* dalam analisis sentimen pada berbagai platform digital seperti YouTube, Google Play, maupun aplikasi layanan publik. Namun, penelitian mengenai analisis sentimen masyarakat terhadap program Gratis Pol pada platform TikTok masih terbatas. Selain itu, TikTok sebagai media sosial banyak digunakan masyarakat untuk menyampaikan opini sehingga menghasilkan berbagai tanggapan terkait program Gratis Pol.

Beragam opini yang muncul, baik positif, netral, maupun negatif, menjadikan analisis sentimen penting dilakukan untuk mengetahui kecenderungan persepsi masyarakat secara lebih sistematis. Selain itu, karakteristik komentar TikTok yang banyak menggunakan bahasa informal, singkatan, serta variasi bahasa sehari-hari menjadi tantangan dalam proses klasifikasi sentimen. Kondisi tersebut dapat mempengaruhi kemampuan model dalam mengidentifikasi sentimen secara tepat sehingga diperlukan metode yang mampu mengolah data teks secara efektif. Analisis sentimen juga dapat membantu memberikan gambaran mengenai tanggapan masyarakat terhadap implementasi program yang dijalankan. Oleh karena itu, penelitian ini dilakukan untuk menganalisis sentimen masyarakat terhadap program Gratis Pol berdasarkan komentar pengguna TikTok menggunakan algoritma *Naive Bayes*.

Penelitian mengenai opini masyarakat terhadap suatu kebijakan melalui media sosial menjadi penting karena dapat memberikan informasi mengenai respons publik secara lebih cepat dan luas. Data komentar yang diperoleh dari media sosial dapat dimanfaatkan sebagai sumber informasi untuk memahami pola tanggapan masyarakat terhadap suatu program yang sedang dijalankan. Dengan memanfaatkan data tersebut, analisis sentimen dapat digunakan sebagai pendekatan untuk memperoleh informasi yang lebih terukur berdasarkan opini yang disampaikan pengguna.

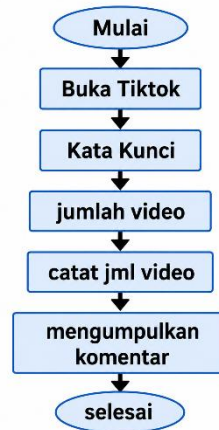
Pemanfaatan data berbasis teks juga memungkinkan proses identifikasi kecenderungan pandangan publik dilakukan secara lebih efisien dibandingkan pengamatan secara manual. Informasi yang dihasilkan dari proses pengolahan data tersebut dapat digunakan untuk melihat pola respons yang muncul terhadap suatu kebijakan atau program yang diterapkan. Selain itu, hasil pengelompokan sentimen dapat menjadi bahan evaluasi untuk memahami bagaimana penerimaan publik terhadap pelaksanaan suatu program berdasarkan informasi yang disampaikan pengguna pada platform digital.

Perkembangan teknologi pengolahan bahasa alami (*Natural Language Processing*) turut mendukung proses pengelompokan data berbasis teks secara lebih terstruktur. Penerapan teknik klasifikasi pada kumpulan komentar digital memungkinkan identifikasi kecenderungan suatu pembahasan dilakukan secara lebih cepat dan efisien. Selain itu, hasil pengolahan tersebut dapat dimanfaatkan untuk melihat pola komunikasi publik terhadap suatu kebijakan yang sedang berlangsung. Pemanfaatan pendekatan komputasional dalam pengolahan data teks juga memberikan peluang untuk

menghasilkan informasi yang lebih terukur sehingga dapat membantu proses interpretasi terhadap berbagai bentuk respons yang muncul pada ruang digital.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian



Gambar 1 Diagram pengumpulan data

Proses pengumpulan data diawali dengan membuka platform TikTok dan melakukan pencarian menggunakan kata kunci “Gratis Pol” untuk memperoleh video yang berkaitan dengan topik penelitian. Berdasarkan hasil pencarian, diperoleh sebanyak 185 video yang relevan dengan kata kunci tersebut. Selanjutnya dilakukan proses pemilihan sampel video menggunakan teknik purposive sampling dengan mempertimbangkan beberapa kriteria, yaitu keterbaruan waktu publikasi video, jumlah tanda suka (like), jumlah komentar, serta kesesuaian isi pembahasan video terhadap program Gratis Pol. Berdasarkan proses seleksi tersebut, diperoleh sebanyak 13 video yang memenuhi kriteria dan digunakan sebagai sumber data penelitian. Tahap berikutnya dilakukan pengumpulan komentar pengguna dari video terpilih secara manual melalui teknik penyalinan (*copy-paste*) komentar yang relevan dengan topik penelitian untuk digunakan pada proses analisis sentimen. Selanjutnya dilakukan tahap preprocessing sebagai berikut:

2.1.1 Text Preprocessing

Tahap pertama dalam text mining adalah *text preprocessing*, yaitu proses penting yang dilakukan sebelum data digunakan dalam pembangunan model agar data lebih siap untuk diolah[12]. Data komentar yang diperoleh kemudian melalui tahap pra-pemrosesan untuk membersihkan dan menyiapkan data agar siap dianalisis, yang meliputi proses *cleaning* (penghapusan karakter tidak penting), *case folding* (mengubah huruf menjadi kecil), *tokenizing* (memecah kalimat menjadi kata), *stopword removal* (menghapus kata yang tidak penting), serta *stemming* (mengubah kata ke bentuk dasar).

2.1.2 Pelabelan Data

Sebagai alternatif, telah dikembangkan berbagai metode pelabelan otomatis dalam proses analisis sentimen. Salah satu metode yang banyak digunakan adalah pendekatan berbasis leksikon (*lexicon-based*), yaitu metode yang memanfaatkan kamus kata yang telah memiliki nilai atau skor polaritas sentimen tertentu untuk menentukan kecenderungan sentimen suatu teks[13].

2.1.3. Representasi Fitur (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk memberikan bobot pada kata dalam dokumen berdasarkan tingkat kemunculan kata tersebut serta tingkat kepentingannya dalam keseluruhan dokumen[14]. Berikut adalah rumus TF-IDF
$$TF-IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{df(t)} \right) \quad (1)$$

TF(t,d) adalah frekuensi kemunculan kata t dalam dokumen d, N adalah jumlah seluruh dokumen, dan df(t) adalah jumlah dokumen yang mengandung kata t.

TF-IDF digunakan untuk merepresentasikan teks ke dalam bentuk numerik sehingga dapat diproses oleh algoritma machine learning. Hasil pembobotan menggunakan TF-IDF kemudian digunakan sebagai fitur masukan dalam proses klasifikasi menggunakan algoritma *Naïve Bayes*.

2.1.4. Klasifikasi *Naive Bayes*

Naive Bayes merupakan metode klasifikasi probabilistik yang didasarkan pada teorema Bayes, dengan asumsi bahwa setiap fitur bersifat independen dan tidak saling memengaruhi satu sama lain dalam menentukan kelas target. Setiap fitur dianggap memberikan kontribusi yang sama terhadap kemungkinan suatu data termasuk ke dalam kelas tertentu. Metode ini relatif mudah diimplementasikan, memiliki efisiensi komputasi yang tinggi, serta mampu memberikan kinerja yang baik pada dataset berukuran besar dengan dimensi yang tinggi [15]. Model klasifikasi dibangun menggunakan algoritma *Naive Bayes* yang dirumuskan berdasarkan persamaan berikut.

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)} \quad (2)$$

Pada persamaan tersebut, $P(C|X)$ merupakan probabilitas suatu data X termasuk ke dalam kelas C , $P(X|C)$ merupakan probabilitas kemunculan data pada kelas tertentu, $P(C)$ merupakan probabilitas awal kelas, dan $P(X)$ merupakan probabilitas keseluruhan data. Algoritma *Naive Bayes* mengasumsikan bahwa setiap fitur bersifat independen, sehingga proses perhitungan menjadi lebih sederhana dan efisien. Probabilitas kemudian dihitung untuk setiap kelas, yaitu positif, negatif, dan netral. Kelas dengan nilai probabilitas terbesar dipilih sebagai hasil klasifikasi.

Tabel 1 Sampel video tiktok

No	Video	Akun	Kata Kunci	Jml_Komentar
1		tvri_Kalimantan timur	Gratis pol	21
...	...			
12		Kaltim.populer	Kaltim. populer gratis pol	114
13		Selasardotco	Selasardotco gratis pol	99

HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna TikTok yang berkaitan dengan program Gratis Pol. Proses pengumpulan data diawali dengan melakukan pencarian video pada platform TikTok menggunakan kata kunci “Gratis Pol”. Berdasarkan hasil pencarian diperoleh sebanyak 185 video yang berkaitan dengan topik penelitian. Selanjutnya dilakukan proses pemilihan video menggunakan teknik purposive sampling, yaitu teknik pengambilan sampel berdasarkan kriteria tertentu yang disesuaikan dengan kebutuhan penelitian. Kriteria yang digunakan dalam pemilihan video meliputi keterbaruan waktu publikasi video, jumlah tanda suka (*like*), jumlah komentar, serta kesesuaian isi video dengan pembahasan mengenai program Gratis Pol. Berdasarkan proses seleksi tersebut diperoleh sebanyak 13 video yang memenuhi kriteria dan digunakan sebagai sumber data penelitian.

Tahap berikutnya dilakukan proses pengumpulan komentar pengguna secara manual melalui teknik *copy-paste* komentar dari 13 video yang telah dipilih. Berdasarkan proses pengumpulan data diperoleh sebanyak 1.528 komentar yang selanjutnya digunakan sebagai dataset penelitian.

Sebelum dilakukan proses analisis sentimen, data terlebih dahulu melalui tahap preprocessing untuk meningkatkan kualitas data teks agar lebih terstruktur dan siap digunakan pada proses klasifikasi. Tahapan preprocessing yang dilakukan meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Tahap *cleaning* dilakukan untuk menghapus karakter yang tidak diperlukan seperti simbol, angka, tanda baca, serta karakter khusus lainnya. Tahap *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Selanjutnya dilakukan *tokenizing* untuk memecah kalimat menjadi kata-kata tunggal. Tahap *stopword removal* dilakukan untuk menghapus kata-kata yang tidak memiliki makna penting dalam proses analisis, sedangkan tahap *stemming* dilakukan untuk mengubah kata berimbuhan menjadi bentuk kata dasar.

Melalui tahapan tersebut diperoleh data komentar yang lebih bersih dan terstruktur sehingga dapat digunakan pada proses analisis sentimen menggunakan algoritma *Naive Bayes*. Berikut merupakan sampel data komentar yang telah melalui tahapan preprocessing.

Tabel 2 Hasil Preprocessing

Vid	NO	Komentar	clean_text	sentimen
Vid1	1	kesah nya gratis pol, sekali nya kd semua anak yg dpt tetap bayar, domisili di kaltim jar minimal 3 thn anak ku dari bayi sampai ganal dah kd dapat 🤔🤔🤔	kesah gratis pol sekali nya kd semua anak yg dpt tetap bayar domisili kaltim jar minimal thn anak ku bayi ganal dah kd	negatif
Vid1	2	byk pander gk ush di percaya cukup kita di tipu ny	byk pander gk ush percaya cukup tipu ny	negatif
...
Vid13	1526	jngn terkecoh dengan omongn manis nya kita sendiri yng merasakn d lapngn 6ng benr enda nya ksta nya grtis 😊😊😊	jngn kecoh omongn manis sendiri yng merasakn d lapngn ng benr enda ksta grtis	netral
Vid13	1527	daftarnya langsung dikampus	daftar langsung kampus	positif
Vid13	1528	Okee ma	okee ma	netral

Selanjutnya, data yang telah diproses dilakukan pelabelan sentimen menggunakan pendekatan berbasis leksikon (*lexicon-based approach*), di mana setiap kata memiliki nilai sentimen tertentu yang digunakan untuk menentukan kategori sentimen komentar menjadi positif, netral, atau negatif. Setelah melalui tahap preprocessing dan pelabelan, diperoleh sebanyak 137 data komentar yang siap digunakan pada tahap analisis selanjutnya. Hasil preprocessing dan pelabelan dapat dilihat pada tabel diatas yang menunjukkan perbandingan antara teks sebelum dan sesudah diproses beserta label sentimennya.

3.2 Labeling Sentimen

Berdasarkan hasil pelabelan sentimen menggunakan pendekatan lexicon-based, diperoleh total sebanyak 1.528 data komentar yang telah dikategorikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Hasil pelabelan menunjukkan bahwa sentimen positif mendominasi dengan jumlah sebanyak 722 komentar, diikuti sentimen netral sebanyak 496 komentar, serta sentimen negatif sebanyak 310 komentar.

Hasil tersebut menunjukkan bahwa sebagian besar pengguna TikTok memberikan tanggapan yang cenderung positif terhadap program Gratis Pol. Meskipun demikian, masih terdapat komentar dengan sentimen netral dan negatif yang menunjukkan adanya variasi persepsi masyarakat terhadap implementasi program tersebut. Perbedaan distribusi jumlah sentimen pada masing-masing kelas juga dapat mempengaruhi proses klasifikasi model, karena distribusi data yang tidak seimbang berpotensi membuat model lebih dominan dalam mengenali kelas dengan jumlah data yang lebih besar. Hasil distribusi pelabelan sentimen dapat dilihat pada Tabel 3.

Tabel 3 Hasil Labeling

sentimen	jumlah
positif	722

netral	496
negatif	310
total	1528

3.3 TF-IDF

Pada tahap ini, data teks hasil pra-pemrosesan diubah menjadi bentuk numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Proses ini dilakukan dengan menggunakan `TfidfVectorizer` dengan batas maksimal 5000 fitur, yang dipilih berdasarkan tingkat kemunculan dan kepentingannya dalam data.

```
tfidf = TfidfVectorizer(max_features=5000)

X = tfidf.fit_transform(df['clean_text'])
y = df['sentimen']
```

Gambar 2 Pembobotan TF-IDFss

Data yang digunakan berasal dari kolom `clean_text`, kemudian diproses menggunakan fungsi `fit_transform()` untuk menghasilkan matriks fitur (X) yang berisi bobot setiap kata. Sementara itu, variabel `y` digunakan sebagai label sentimen. Hasil representasi ini selanjutnya digunakan sebagai input dalam proses klasifikasi menggunakan algoritma *Naive bayes*.

3.4 Split Data

Dalam pengembangan model machine learning, pembagian data merupakan proses memisahkan dataset menjadi beberapa bagian, yaitu data latih (training) dan data uji (testing). Berikut merupakan proses pembagian data yang dilakukan.

```
x_train, x_test = train_test_split(
    X, y, test_size=0.2, random_state=42)
```

Gambar 3 Split Data

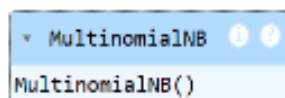
Berdasarkan gambar tersebut, pembagian data dilakukan dengan komposisi 80% sebagai data latih dan 20% sebagai data uji [16]. Berdasarkan pembagian tersebut, diperoleh sebanyak 1528 data, 1222 data sebagai data latih dan 306 data sebagai data uji. Data latih digunakan untuk melatih model klasifikasi *Naive Bayes*, sedangkan data uji digunakan untuk mengevaluasi kinerja model dalam memprediksi sentimen.

3.5 Implementasi *Naive bayes*

Penelitian ini menerapkan algoritma *Naive bayes* dalam proses klasifikasi sentimen. Implementasi algoritma ditunjukkan melalui kode program pada Gambar 4 yang menggambarkan proses pembentukan dan pelatihan model.

```
# Membuat model Naive Bayes
model = MultinomialNB()

# Melatih model dengan data training
model.fit(X_train, y_train)
```



Gambar 4 Algoritma *Naive bayes*

Pada tahap ini, model klasifikasi dibangun menggunakan algoritma *Naive Bayes* dengan tipe `MultinomialNB` yang sesuai untuk data teks berbasis frekuensi seperti TF-IDF. Model kemudian dilatih menggunakan data latih (training data) yang

telah melalui proses pembobotan. Proses pelatihan dilakukan dengan fungsi $\text{fit}()$ untuk mempelajari pola dari data, sehingga model dapat digunakan untuk melakukan prediksi terhadap data uji.

3.6 Evaluasi Model

Berdasarkan hasil evaluasi model menggunakan classification report, diperoleh nilai akurasi sebesar 56%. Hasil tersebut menunjukkan bahwa model *Naive Bayes* memiliki kemampuan yang cukup dalam mengklasifikasikan sentimen komentar masyarakat terhadap program Gratis Pol pada platform TikTok.

Pada kelas sentimen negatif, model memperoleh nilai precision sebesar 0,86, recall sebesar 0,25, dan f1-score sebesar 0,39. Nilai precision yang tinggi menunjukkan bahwa sebagian besar data yang diprediksi negatif memang benar termasuk sentimen negatif. Namun, nilai recall yang rendah menunjukkan bahwa model masih kesulitan mendeteksi seluruh data negatif yang tersedia.

Pada kelas sentimen netral, model memperoleh nilai precision sebesar 0,76, recall sebesar 0,14, dan f1-score sebesar 0,24. Hasil ini menunjukkan bahwa model belum mampu mengenali sentimen netral dengan baik. Rendahnya nilai recall pada kelas netral menunjukkan bahwa sebagian besar data netral masih sering salah diklasifikasikan ke kelas lain, terutama ke kelas positif.

Sementara itu, pada kelas sentimen positif, model menunjukkan performa terbaik dengan nilai precision sebesar 0,53, recall sebesar 0,99, dan f1-score sebesar 0,69. Nilai recall yang sangat tinggi menunjukkan bahwa hampir seluruh data sentimen positif berhasil dikenali oleh model. Hal ini kemungkinan disebabkan oleh jumlah data sentimen positif yang lebih dominan dibandingkan kelas lainnya, sehingga model lebih banyak mempelajari pola kata pada kelas positif. Untuk evaluasi yang lebih mendalam, classification report yang mencakup precision, recall, dan F1-score untuk setiap kelas disajikan pada gambar di bawah[17].

Accuracy: 0.5620915032679739

Classification Report:

	precision	recall	f1-score	support
negatif	0.86	0.25	0.39	48
netral	0.76	0.14	0.24	113
positif	0.53	0.99	0.69	145
accuracy			0.56	306
macro avg	0.72	0.46	0.44	306
weighted avg	0.67	0.56	0.48	306

Gambar 5 Classification Report

Secara keseluruhan, nilai macro average f1-score sebesar 0,44 dan weighted average f1-score sebesar 0,48 menunjukkan bahwa performa model masih belum merata pada setiap kelas sentimen. Kondisi ini dipengaruhi oleh ketidakseimbangan jumlah data serta karakteristik komentar TikTok yang banyak menggunakan bahasa informal, singkatan, dan sarkasme sehingga menyulitkan proses klasifikasi menggunakan metode *Naive Bayes* dan pendekatan lexicon-based.

Jika dibandingkan dengan penelitian terdahulu yang menggunakan algoritma *Naive Bayes*, hasil penelitian ini masih tergolong lebih rendah. Hal tersebut disebabkan oleh perbedaan karakteristik data yang digunakan. Penelitian ini menggunakan data komentar TikTok yang cenderung tidak terstruktur dan memiliki banyak variasi bahasa informal, sehingga proses klasifikasi sentimen menjadi lebih kompleks dibandingkan data ulasan formal pada platform lain.

3.7 Visualisasi *WordCloud*

Wordcloud merupakan visualisasi frekuensi kata dalam suatu teks dalam bentuk daftar berbobot[18]. *WordCloud* salah satu teknik visualisasi yang digunakan untuk menampilkan frekuensi kemunculan kata dalam bentuk visual. Metode ini banyak digunakan dalam text mining karena mampu menyajikan informasi kata-kata yang dianalisis secara visual, sehingga memudahkan pemahaman meskipun tetap bersifat informatif[19].

Selanjutnya dilakukan proses visualisasi data menggunakan *WordCloud* berdasarkan hasil klasifikasi data sentiment. Hasil visualisasi data menggunakan *WordCloud* dapat dilihat pada gambar 6 berikut [20].

Rendahnya performa pada kelas negatif dan netral kemungkinan disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah data positif lebih dominan dibandingkan kelas lainnya. Selain itu, penggunaan pendekatan lexicon-based dalam proses pelabelan juga memiliki keterbatasan dalam memahami konteks bahasa informal, singkatan, bahasa daerah, serta sarkasme yang banyak ditemukan pada komentar TikTok. Karakteristik komentar media sosial yang cenderung ambigu dan tidak terstruktur menyebabkan beberapa komentar netral maupun negatif memiliki pola kata yang mirip dengan sentimen positif. Hal ini mempengaruhi kemampuan model dalam membedakan setiap kelas sentimen secara optimal.

KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap Program Gratis Pol pada platform TikTok menggunakan pendekatan *lexicon-based*, pembobotan TF-IDF, dan algoritma *Naive Bayes*. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi tanggapan masyarakat dengan jumlah 722 komentar, diikuti 496 komentar netral dan 310 komentar negatif dari total 1.528 data komentar yang dianalisis. Model klasifikasi menghasilkan tingkat akurasi sebesar 56%, dengan performa terbaik pada kelas sentimen positif, namun masih mengalami keterbatasan dalam mengenali sentimen netral dan negatif akibat ketidakseimbangan distribusi data serta karakteristik komentar TikTok yang banyak mengandung bahasa informal, singkatan, dan variasi bahasa sehari-hari. Penelitian ini memberikan kontribusi dalam menggambarkan persepsi masyarakat terhadap Program Gratis Pol serta menunjukkan penerapan metode *lexicon-based* dan algoritma *Naive Bayes* dalam analisis sentimen media sosial. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, melakukan penyeimbangan data, serta membandingkan algoritma lain agar performa klasifikasi sentimen dapat ditingkatkan.

REFERENCES

- [1] A. Permana, A. Maulida, F. Husaeni, H. M. A, S. Ruhimasari, and P. P. Surya, "Definisi Kebijakan Publik: Pengertian, Unsur-unsur Dan Perkembangan," *J. Penelit. Nusant.*, vol. 1, no. 5, pp. 453–457, 2025, doi: <https://doi.org/10.59435/menulis.v1i5.292>.
- [2] A. Galluci, Irwanto, Muhlis, M. A. Rahman, A. Rachmadani, and S. Hardyanti, "PERAN DAN KEBIJAKAN OTONOMI DAERAH PEMERINTAH KALIMANTAN TIMUR DALAM PEMERATAAN AKSES PENDIDIKAN: ANALISIS HUKUM ADMINISTRASI NEGARA," *J. Ilm. PGSD STKIP Subang*, vol. 12, no. 1, pp. 196–211, 2026, [Online]. Available: <https://www.preventionweb.net/news/preliminary-report-february-6-2023-earthquakes-turkiye>
- [3] Riskawati, Fatihanursari, Lin, and A. R. Rinaldi, "Penerapan Metode Naive Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 346–353, 2024, doi: 10.36040/jati.v8i1.8699.
- [4] H. A. R. Harpizon, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafira, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naive Bayes," *JNKTI (Jurnal Nas. Komputasi dan Teknol. Informasi)*, vol. 5, no. 1, pp. 131–140, 2022.
- [5] R. Isnain, Auliya, H. Sulistiani, M. Hurohman, Bagus, A. Nurkholis, and Styawati, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 8, no. 2, pp. 299–303, 2022.
- [6] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] A. Safira and F. N. Hasan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023.
- [8] M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [9] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [10] A. Karimah, G. Dwilestari, and Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 767–773, 2024.
- [11] A. A. Ningtyas, A. Solichin, and R. Pradana, "Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritma Naive Bayes," *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 20, no. 1, pp. 9–16, 2023, doi: 10.36080/bit.v20i1.2317.
- [12] K. C. Astuti, A. Firmansyah, and A. Riyadi, "Implementasi Text Mining untuk Analisis Sentimen Masyarakat terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 8, no. 1, pp. 383–394, 2024, doi: doi.org/10.33395/remik.v8i1.13421.
- [13] A. S. Rizkia, Wufron, and F. F. Roji, "Sentiment Analysis of Coretax : A Comparison of Manual , Transformers- Based , and Lexicon-Based Data Labeling on IndoBERT Performance Analisis Sentimen Coretax : Perbandingan Pelabelan Data Manual ," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 1037–1048, 2025.
- [14] V. W. D. Thomas and F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218.



- [15] B. Hirwono, A. Hermawan, and D. Avianto, "Implementasi Metode Naïve Bayes untuk Klasifikasi Penderita Penyakit Jantung," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 3, pp. 450–457, 2023, doi: 10.35870/jtik.v7i3.910.
- [16] D. Prasetia, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, "ANALISIS SENTIMEN PENGGUNA APLIKASI MYBLUEBIRD DENGAN ALGORITMA NAÏVE BAYES DI PLAYSTORE," *JITET(Jurnal Inform. dan Tek. Elektro Tepian)*, vol. 13, no. 1, pp. 602–608, 2025, doi: doi.org/10.23960/jitet.v13i1.5687.
- [17] R. Syahri and D. Puspita, "Classification Of Outstanding Students Using Support Vector Machine (SVM)," *JITE (Journal Informatics Telecommun. Eng. Available)*, vol. 9, no. 1, pp. 13–23, 2025, doi: 10.31289/jite.v9i1.13191.
- [18] I. Zulfahmi, "Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decission Tree," *J. Penelit. Rumpun Ilmu Tek.*, vol. 3, no. 1, pp. 11–21, 2024, doi: <https://doi.org/10.55606/juprit.v3i1.3096>.
- [19] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek Dengan Support Vector Machine," *J. Serina Sains, Tek. dan Kedokt.*, vol. 2, no. 1, pp. 61–70, 2024, doi: 10.24912/jsstk.v2i1.32058.
- [20] A. Hanafiah *et al.*, "Sentimen analisis terhadap customer review produk shopee berbasis wordcloud dengan algoritma naïve bayes classifier sentiment analysis of customer reviews of shopee products based on wordcloud using naïve bayes classifier algorithm," *J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 230–236, 2023.

