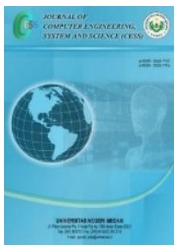
CESS (Journal of Computer Engineering, System and Science)
Vol. 10, No. 1, Januari 2025, pp.299-311

CESS
(Journal of Computer Engineering, System and Science)
Available online: https://jurnal.unimed.ac.id/2012/index.php/cess
ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Analisis Sentimen Menggunakan Metode Naïve Bayes Tentang Program Mudik Gratis Pemerintah Kota Medan 2024

Sentiment Analysis Using the Naïve Bayes Method on the 2024 Free Homecoming Program by the Medan City Government

Nurwana Nazla Saragih[1*], Rakhmat Kurniawan[2]

[1,2]Ilmu Komputer, Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia
Jl. Lap Golf No 120, Kp Tengah, Kec Pancur Batu, Kab Deli Serdang, Provinsi Sumatera Utara, Indonesia
Email: [1]nurwananazla@gmail.com, [2]rakhmat.kr@uinsu.ac.id
**Corresponding Author*

A B S T R A K

Pemerintah Kota Medan secara rutin mengadakan program mudik gratis bagi masyarakatnya, yang telah berlangsung selama tiga tahun berturut-turut (2022, 2023, dan 2024). Program ini mendapatkan respons beragam dari masyarakat, yang terekam melalui berbagai komentar di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program mudik gratis tahun 2024 dengan menggunakan metode Naïve Bayes. Data yang digunakan berupa 394 komentar dari akun Instagram resmi Wali Kota Medan, yang diklasifikasikan ke dalam tiga kategori sentimen positif, negatif, dan netral. Metode penelitian meliputi pengumpulan data, preprocessing (cleaning, case folding, tokenizing, filtering, stopword removal, stemming), serta penerapan algoritma Naïve Bayes dengan teknik TF-IDF sebagai fitur ekstraksi. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan F1-score. Hasil analisis menunjukkan bahwa dari total 394 komentar, terdapat 110 komentar positif, 107 komentar negatif, dan 177 komentar netral. Model Naïve Bayes yang digunakan mencapai akurasi sebesar 51,667%. Meskipun model mampu mengklasifikasikan data dengan tingkat keakuratan sedang, terdapat kelemahan dalam mendeteksi sentimen negatif, yang tercermin dari rendahnya recall pada kelas negatif. Faktor-faktor seperti ketidakseimbangan kelas, pemilihan fitur, dan kualitas data menjadi aspek yang perlu ditingkatkan untuk meningkatkan kinerja model di masa mendatang.

Kata Kunci: *Analisis Sentimen; Mudik Gratis; Naïve Bayes.*

A B S T R A C T

The Medan City Government has consistently organized a free homecoming program for its residents over the past three years (2022, 2023, and 2024). This initiative has received mixed

This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license

reactions from the public, as reflected in various comments on social media. This study aims to analyze public sentiment toward the 2024 free homecoming program using the Naïve Bayes method. The dataset consists of 394 comments collected from the official Instagram account of the Mayor of Medan, categorized into three sentiment classes positive, negative, and neutral. The research methodology includes data collection, preprocessing (cleaning, case folding, tokenizing, filtering, stopword removal, and stemming), and the implementation of the Naïve Bayes algorithm with Term Frequency-Inverse Document Frequency (TF-IDF) as the feature extraction technique. The model's performance is evaluated using a confusion matrix to measure accuracy, precision, recall, and F1-score. The results indicate that out of 394 comments, 110 were classified as positive, 107 as negative, and 177 as neutral. The Naïve Bayes model achieved an accuracy of 51.667%. While the model demonstrated moderate classification performance, it struggled to detect negative sentiments, as reflected by the low recall score for the negative class. Factors such as class imbalance, feature selection, and data quality need further refinement to improve the model's performance in future research.

Keywords: *Sentiment Analysis; Free Homecoming; Naïve Bayes*.

1. PENDAHULUAN

Pemerintah Kota Medan adalah salah satu pemerintah yang telah rutin mengadakan sistem mudik gratis bagi masyarakatnya. Diketahui selama masa kepemimpinan Bapak Bobby Nasution, Kota Medan telah mengadakan sistem mudik gratis ini selam 3 tahun, yakni pada tahun 2022, 2023, dan 2024. Tidak hanya itu, kuota yang disediakan juga semakin meningkat, yang dimana Terdapat kuota 3.500 penumpang pada tahun 2022, 5.000 penumpang pada tahun 2023, dan 6.000 penumpang pada tahun 2024. Dua belas rute kota tujuan di Sumut diluncurkan Pemko Medan sebagai bagian dari inisiatif Mudik Bareng 2024. Medan - Sidikalang, Medan - Tarutung, Medan - Sibolga, Medan - Rantau Prapat, Medan - Kota Pinang, Medan - Sibuhuan, Medan - Gunung Tua, Medan - Sipirok, Medan - Natal, Medan - Panyabungan, dan Medan - Padang Sidempuan adalah 12 jalan harus dipertimbangkan[1].

Namun berbeda dari tahun sebelumnya, tahun ini pendaftaran dilakukan hanya secara offline atau manual, tidak dilakukan secara online. Dilakukannya pendaftaran secara manual bukan berarti ingin mempersulit calon pemudik, namun mengingat di tahun sebelumnya terjadi pendaftaran ganda yang mencapai 1.200 pemudik. Untuk syarat pendaftaran tidak harus pemilik KTP Medan, namun tujuannya harus sesuai dengan yang telah ditetapkan. Selain itu para pemudik tak harus semua mendaftar namun bisa diwakili oleh satu orang atau orang lain, keluarga maupun orang sekampung atau satu tujuan untuk satu bus. Untuk armada Pemko Medan telah menyiapkan 179 bus yang terdiri dari 121 bus besar dan 58 bus berukuran sedang.

Sungguh luar biasa antusiasme masyarakat untuk mengikuti program Mudik Bareng Pemko Medan 2024. Pasalnya, kuota maksimal 6.000 penumpang langsung dianggap penuh pada hari kedua pendaftaran. Alhasil, pada hari kedua penyelenggaraan Mudik Bareng Pemko Medan 2024, pendaftaran resmi ditutup oleh Dinas Perhubungan pendaftarannya. Walau demikian, masih banyak masyarakat yang terlambat dan tidak berhasil mendaftar dan menjadi salah satu peserta mudik gratis tahun ini. Berbagai komentar terlihat di beberapa postingan Instagram Bapak Bobby Nasution. Ucapan rasa bangga, terima kasih, dan kekecawaan juga turut dilampirkan masyarakat di lama tersebut.

Analisis sentimen alias opinion mining adalah salah satu jenis natural languange dengan mengolah kata untuk melakukan pelacakan terhadap mood masyarakat atas suatu produk atau topik tertentu [2]. Metode naïve bayes merupakan teknik klasifikasi pembelajaran mesin Ini memberikan manfaat mengidentifikasi parameter proses klasifikasi secara cepat menggunakan data dan akurat [3].

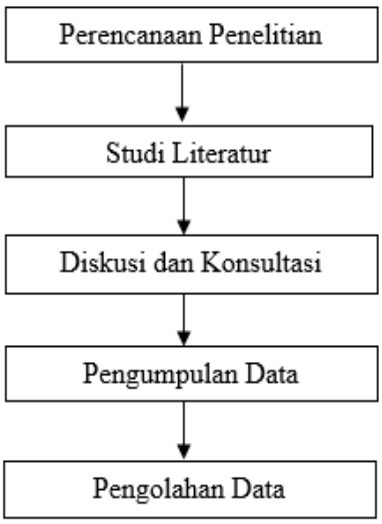
Penelitian tentang analisis sentimen pada layanan publik telah banyak dilakukan, khususnya dalam bidang transportasi dan kebijakan pemerintah. Beberapa studi sebelumnya menggunakan metode Naïve Bayes untuk menganalisis opini publik terhadap layanan pemerintah dan berbagai kebijakan transportasi. Namun, penelitian yang secara spesifik menganalisis sentimen masyarakat terhadap program mudik gratis di Kota Medan masih sangat terbatas. Selain itu, sebagian besar penelitian serupa menggunakan dataset yang lebih besar dari berbagai platform media sosial, sedangkan penelitian ini berfokus pada data dari satu sumber, yaitu Instagram.

Dari sisi metodologi, metode Naïve Bayes telah terbukti efektif dalam berbagai studi analisis sentimen, tetapi tingkat akurasinya sering dipengaruhi oleh kualitas data dan ketidakseimbangan kelas dalam dataset. Oleh karena itu, perlu dilakukan evaluasi terhadap performa model pada dataset yang lebih kecil dengan distribusi sentimen yang tidak merata.

2. METODE PENELITIAN

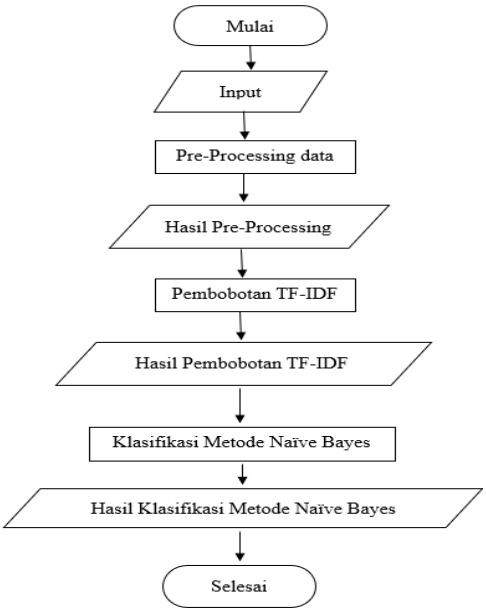
2.1 Tahapan Penelitian

Adapun kerangka penelitian adalah proses yang digunakan untuk memecahkan masalah yang diberikan dan digunakan untuk membantu peneliti dalam menawarkan kerangka kerja yang komprehensif dan transparan untuk fase-fase tersebut. Untuk melaksanakan penelitian ini, sejumlah langkah harus diselesaikan dalam bentuk *flowchart* Gambar.



Gambar 1. Kerangka Penelitian

Perancangan yaitu tahap dari perencanaan dari sejumlah komponen yang berbeda menjadi satu kesatuan yang utuh dan kohesif. Perencanaan model terdiri dari metode naïve bayes juga disajikan dalam *flowchart* pada Gambar 2.



Gambar 2. *Flowchart* Metode Naïve Bayes

2.2 TF-IDF

Term Frequency adalah sebuah istilah dalam penentuan tingkat frekuensi dalam dokumen, di mana kuantitas dokumen yang menggunakan istilah tersebut dikenal sebagai frekuensi dokumen. Pada rumus di atas, terlihat bahwa nilai term frequency dari document frequency[4]. N disebut sebagai jumlah seluruh dokumen, df adalah lambang dari document frequency dari term yang akan dicari inversnya. Maka, nilai invers document frequency (idf), yaitu:

$$\text{Idf} = \log\left(\frac{N}{df}\right) \quad (1)$$

$$W = tf * idf \quad (2)$$

Prosedur sortir akan selesai jika bobot (W) setiap dokumen diketahui. Semakin mirip dokumen dengan kata kunci yang digunakan, semakin tinggi nilai W, dan sebaliknya.

Penjelasan rumus:

d : dokumen ke-d
t : kata ke-t dari kata kunci
W : bobot dokumen ke-d dalam kaitannya dengan kata ke-t
Tf : jumlah kata dalam dokumen yang dicari Inverse Document Frequency, atau IDF
ft,d : frekuensi kata dalam d
df : jumlah dokumen yang memuat istilah pencarian

2.3 Naïve Bayes Classifier

Jika diuji pada kumpulan data yang sesuai, algoritma klasifikasi Naïve Bayes sangat efektif karena mudah dan sederhana untuk diterapkan. Hal ini terutama berlaku jika dikombinasikan dengan pemilihan fungsi, yang memungkinkan Naïve Bayes mengurangi redundansi dalam data. Selain itu, Naïve Bayes berkinerja baik jika dipasangkan dengan metode pengelompokan [5]. Algoritma ini tetap dapat menghasilkan hasil klasifikasi yang baik. Naïve bayes merupakan implentasi algoritma pembelajaran mesin probabilistik menggunakan

teknik perhitungan teorema bayes [6]. Berikut ini adalah bentuk umum dari teorema bayes :

$$P(H|X)^{\wedge} = =(p(H|X).P(H))/(P(X)) \quad (3)$$

Keterangan Rumus :
X : Data class belum validasi
H : Data class spesifik
P(H|X)^ : Probabilitas berdasarkan kondisi
P(H) : Probabilitas
P(X|H) : Probabilitas kondisi pada hipotesis
P(X) : Probabiliras H

2.4 Confusion Matrix

Confusion Matrix merupakan suatu instrumen yang digunakan untuk mengevaluasi performa dari model klasifikasi yang telah dihasilkan [7]. Pada *confusion matrix,* hasil kelas prediksi akan dibandingan dengan hasil kelas data yang sebenarnya, Hasil tersebut kemudian akan digunakan untuk menghitung nilai akurasi, *presion, recall, dan f-score* [8]. [8]. Dalam hal klasifikasi teks, kesalahan terkadang dapat terjadi. Oleh karena itu, penting untuk memahami tingkat presisi yang diperlukan untuk perbandingan dengan penelitian lain. Akurasi perhitungan merupakan salah satu metode untuk menentukan akurasi klasifikasi teks[9]. Tingkat kemiripan nilai aktual pada persamaan berikut dengan nilai prediksi dikenal sebagai akurasi:

$$\text{Akurasi} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \; x\; 100\% \quad (4)$$

Precision dijelaskan pada persamaan 2.6 sebagai berikut:

$$\textit{Precision} = \frac{\text{TP}}{\text{FP}+\text{TP}} \; x100\% \quad (5)$$

Recall dijelaskan pada persamaan 2.7 sebagai berikut :

$$\textit{Recall} = \frac{\text{TP}}{\text{FN}+\text{TP}} \; x\; 100\% \quad (6)$$

F1-score dijelaskan pada persamaan 2.8 sebagai berikut:

$$\textit{F1-score} = \frac{2 * (recall * precision)}{(recall + precision)} \; x\; 100\% \quad (7)$$

Keterangan:
a) TP adalah singkatan dari True Positive, yang menunjukkan bahwa total nilai dalam kategori positif mewakili nilai yang diantisipasi dan nilai aktual.
b) TN True Negative, yang menunjukkan bahwa jumlah kategori negatif berisi nilai yang diharapkan dan aktual.
c) False Positive (FP) adalah istilah yang digunakan untuk menggambarkan nilai yang memiliki prediksi kategori positif tetapi berisi jumlah nilai kategori negatif.
d) FN False Negative, yang menunjukkan bahwa nilai aktual memiliki kategori positif tetapi nilai yang diprediksi berada dalam jumlah kategori negatif.

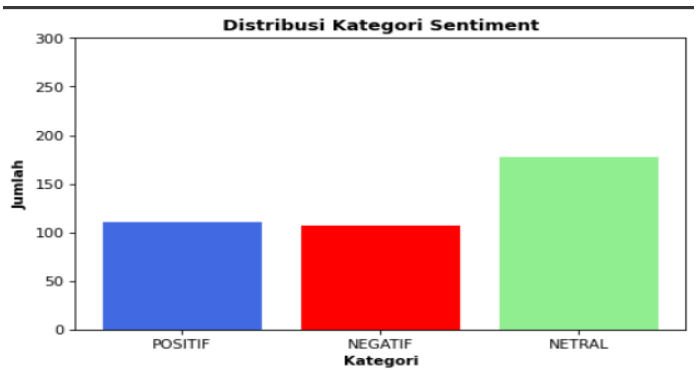
3. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk menguji sentimen. masyarakat terhadap program mudik gratis yang diselenggarakan di Kota Medan. Analisis ini dilakukan untuk

mengetahui tanggapan, opini, dan persepsi masyarakat, baik yang bersifat positif, negatif, maupun netral, terkait program tersebut[10]Dengan menggunakan metode analisis sentimen, penelitian ini diharapkan dapat memberikan gambaran mengenai efektivitas program mudik gratis dalam memenuhi harapan masyarakat serta menjadi bahan evaluasi bagi penyelenggara untuk meningkatkan kualitas layanan selanjutnya. Lebih jauh lagi, penelitian ini diharapkan dapat membantu terciptanya kebijakan transportasi yang lebih baik di Kota Medan.

3.1. Pengumpulan Data

Penelitian ini memanfaatkan data komentar yang diperoleh dari akun Instagram resmi Bobby Nasution, Wali Kota Medan, yang membahas program mudik gratis. Data tersebut terdiri dari 394 komentar yang mencerminkan berbagai tanggapan masyarakat terhadap program tersebut. Dari data komentar yang diperoleh, analisis sentimen akan dilakukan menggunakan metode Naive Bayes untuk mengklasifikasikan komentar-komentar tersebut[11]. Berdasarkan hasil klasifikasi, diperoleh jumlah sentimen sebagai berikut: 110 komentar dengan sentimen positif, 107 komentar dengan sentimen negatif, dan 177 komentar dengan sentimen netral. Distribusi sentimen ini dapat dilihat pada Gambar 3 berikut.



Gambar 3. Data Komentar

3.2. Preprocessing Data

Sebelum dianalisis, data yang diambil harus melewati sejumlah langkah persiapan. secara efektif Tujuan praproses adalah untuk menciptakan data yang bersih dan terstruktur yang dipersiapkan untuk pemrosesan tambahan dari kumpulan data mentah yang awalnya tidak terstruktur dan penuh dengan gangguan atau item yang tidak berguna[12]. Delapan set data pelatihan disediakan sebagai contoh perhitungan manual yang digunakan dalam penyelidikan ini. tersebut digunakan untuk memberikan ilustrasi proses perhitungan secara manual dalam analisis yang dilakukan. Berikut adalah rincian data yang digunakan sebagai contoh Tabel 1.

Tabel 1. Data Komentar

ID	Instagram Comment Text
1	Katanya benci dinasty tp progamnya....wkwkwk
2	Tanjungbalai dong pak
3	@stimsrgi 50 ribu ny rajawali hahahaha
4	Mantap Pak Wali 🔥🔥🔥
5	@ahmadalrasid_ sumbangsih nya dong bg

6	nanti pas libur raya natal ada ngga ya program kek gini?
7	Semoga ada @bobbynst
8	Ini Pp ga sih pak?

Dari data latih di atas, Untuk memastikan data siap untuk analisis, langkah praproses diselesaikan. Fase ini meliputi cleaning untuk membersihkan data, case folding untuk menyeragamkan huruf, tokenizing untuk memecah teks, filtering untuk menyaring data relevan, stopword removal untuk menghapus kata-kata tidak signifikan, dan stemming untuk mengubah kata menjadi bentuk dasar. Proses ini bertujuan menghasilkan data yang bersih dan siap digunakan dalam analisis.

a) Cleaning

Proses ini juga mencakup perbaikan data yang tidak konsisten, seperti mengganti ejaan yang salah atau menghapus spasi berlebih. *Cleaning* memastikan data menjadi lebih bersih, terstruktur, dan siap digunakan untuk analisis lebih lanjut, sehingga dapat meningkatkan kualitas dan akurasi hasil analisis.

Tabel 2. Data Cleaning

Id	Data *Cleaning*	Data *Cleaning*
1	Katanya benci dinasty tp progamnya wkwkwk	Katanya benci dinasty tp progamnya wkwkwk
2	Tanjungbalai dong pak	Tanjungbalai dong pak
3	stimsrgi ribu ny rajawali hahahaha	stimsrgi ribu ny rajawali hahahaha
4	Mantap Pak Wali	Mantap Pak Wali
5	ahmadalrasid sumbangsih nya dong bg	ahmadalrasid sumbangsih nya dong bg
6	nanti pas libur raya natal ada ngga ya program kek gini	nanti pas libur raya natal ada ngga ya program kek gini
7	Semoga ada bobbynst	Semoga ada bobbynst
8	Ini Pp ga sih pak	Ini Pp ga sih pak

b) Case Folding

Tujuan utama dari case folding adalah untuk menghindari perbedaan antara huruf kecil dan huruf besar mempengaruhi analisis. Misalnya, kata "Medan" dan "medan" akan diperlakukan sebagai dua entitas yang berbeda jika tidak diubah menjadi bentuk yang konsisten. Dengan melakukan case folding, data menjadi lebih seragam dan analisis menjadi lebih akurat, karena tidak ada perbedaan yang disebabkan oleh kapitalisasi huruf[13].

Tabel 3. Data Case folding

Id	Data Cleaning	Data Case folding
1	Katanya benci dinasty tp progamnya wkwkwk	katanya benci dinasty tp progamnya wkwkwk
2	Tanjungbalai dong pak	tanjungbalai dong pak
3	stimsrgi ribu ny rajawali hahahaha	ribu ny rajawali hahahaha
4	Mantap Pak Wali	mantap pak wali

5	ahmadalrasid sumbangsih nya dong bg	sumbangsih nya dong bg
6	nanti pas libur raya natal ada ngga ya program kek gini	nanti pas libur raya natal ada ngga ya program kek gini
7	Semoga ada bobbynst	semoga ada bobbynst
8	Ini Pp ga sih pak	ini pp ga sih pak

c) Tokenizing

Token ini bisa berupa kata, frasa, atau simbol, tergantung pada jenis analisis yang dilakukan. Misalnya, kalimat "Mudik gratis di Medan sangat menyenangkan" akan dipecah menjadi token-token seperti "Mudik", "gratis", "di", "Medan", "sangat", "menyenangkan". Proses ini penting untuk memudahkan pemrosesan data teks dalam tahap analisis berikutnya, seperti penghitungan frekuensi kata atau klasifikasi. Dengan tokenizing, teks yang panjang dan kompleks diubah menjadi unit-unit yang lebih mudah dianalisis.

Tabel 4. Data Tokenizing

Id	Data *Case folding*	Data *Tokenizing*
1	katanya benci dinasty tp progamnya wkwkwk	“katanya” , ”benci” , “dinasty” , “tp” , “progamnya” , “wkwkwk”
2	tanjungbalai dong pak	“Tanjungbalai” , “dong” , “pak”
3	ribu ny rajawali hahahaha	“ribu” , “ny” , “rajawali” , “hahahaha”
4	mantap pak wali	“mantap” , “pak” , “wali”
5	sumbangsih nya dong bg	“sumbangsih” , “nya” , “dong” , “bg”
6	nanti pas libur raya natal ada ngga ya program kek gini	“nanti” , “pas” , “libur” , “raya, “natal” , “ada” , “ngga” , “ya” , “program” , “kek” , “gini”
7	semoga ada bobbynst	“semoga” , “ada” , “bobbynst”
8	ini pp ga sih pak	“ini” , “pp” , “ga” , “sih” , “pak”

d) Filtering

Token yang dihapus biasanya meliputi kata-kata umum atau terlalu sering muncul, seperti "dan", "atau", "yang", yang disebut sebagai *stopwords*. Selain itu, *filtering* juga bisa melibatkan penghapusan kata-kata yang tidak sesuai dengan konteks analisis, seperti angka atau karakter khusus yang tidak mendukung tujuan analisis. Dengan *filtering*, data yang tersisa menjadi lebih fokus dan relevan untuk proses analisis selanjutnya.

Tabel 5. Data Filtering

Id	Data Tokenizing	Data Filtering
1	“katanya” , ”benci” , “dinasty” , “tp” , “progamnya” , “wkwkwk”	“katanya” , ”benci” , “dinasty” , “tp” , “progamnya” , “wkwkwk”
2	“Tanjungbalai” , “dong” , “pak”	“Tanjungbalai” , “dong” , “pak”
3	“ribu” , “ny” , “rajawali” , “hahahaha”	“ribu” , “ny” , “rajawali” , “hahahaha”
4	“mantap” , “pak” , “wali”	“mantap” , “pak” , “wali”

5	“sumbangsih” , “nya” , “dong” , “bg”	“sumbangsih” , “nya” , “dong” , “bg”
6	“nanti” , “pas” , “libur” , “raya, “natal” , “ada” , “ngga” , “ya” , “program” , “kek” , “gini”	“nanti” , “pas” , “libur” , “raya, “natal” , “ada” , “ngga” , “ya” , “program” , “kek” , “gini”
7	“semoga” , “ada” , “bobbynst”	“semoga” , “ada” , “bobbynst”
8	“ini” , “pp” , “ga” , “sih” , “pak”	“ini” , “pp” , “ga” , “sih” , “pak”

e) *Stopword*

kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna signifikan untuk analisis, seperti "dan", "yang", "atau", dan "di"[14]. Dalam proses *preprocessing*, *stopword* removal dilakukan untuk menghapus kata-kata tersebut, sehingga data yang tersisa lebih relevan dan fokus pada informasi penting. Penghapusan *stopword* membantu mengurangi noise dan meningkatkan akurasi dalam analisis teks, seperti analisis sentimen atau klasifikasi.

Tabel 6. Data Stopword

Id	Data Filtering	Data Stopword
1	“katanya” , ”benci” , “dinasty” , “tp” , “programnya” , “wkwkwk”	”benci” , “dinasty” , “programnya”
2	“Tanjungbalai” , “dong” , “pak”	“Tanjungbalai”
3	“ribu” , “ny” , “rajawali” , “hahahaha”	“ribu” , “rajawali” ,
4	“mantap” , “pak” , “wali”	“wali”
5	“sumbangsih” , “nya” , “dong” , “bg”	“sumbangsih”
6	“nanti” , “pas” , “libur” , “raya”, “natal” , “ada” , “ngga” , “ya” , “program” , “kek” , “gini”	“pas” , “libur” , “raya”, “natal” , “gini”
7	“semoga” , “ada” , “bobbynst”	“bobbynst”
8	“ini” , “pp” , “ga” , “sih” , “pak”	“pp” ,

f) Stemming

Tujuan stemming adalah untuk menyederhanakan variasi kata yang memiliki makna sama, seperti "berlari", "lari", dan "lari-lari", menjadi bentuk dasar "lari". Proses ini penting dalam analisis teks untuk mengurangi kompleksitas data dan memastikan bahwa kata-kata dengan makna serupa tidak dihitung sebagai entitas yang berbeda[15]. Dengan stemming, data menjadi lebih terstruktur, sehingga memudahkan proses analisis seperti klasifikasi atau penghitungan frekuensi kata.

Tabel 7. Stemming

Id	Data Stopword	Data Stemming
1	”benci” , “dinasty” , “programnya” ,	”benci” , “dinasty” , “program” ,
2	“Tanjungbalai”	“Tanjungbalai”
3	“ribu” , “rajawali”	“ribu” , “rajawali”
4	“wali”	“wali”

5	“sumbangsih”	“sumbang”
6	“pas” , “libur” , “raya”, “natal”, “gini”	“pas” , “libur” , “raya”, “natal” , “gini”
7	“bobbynst”	“bobbynst”
8	“ini” , “pp” , “ga” , “sih” , “pak”	“pp” ,

3.3 Penerapan Algoritma Naïve Bayes di Python

Klasifikasi teks di Python dimulai dengan mengubah data teks menjadi representasi numerik menggunakan CountVectorizer, yang menghitung frekuensi kata. Setelah itu, Train_test_split digunakan untuk memisahkan data menjadi data pelatihan dan pengujian. Untuk memisahkan data menjadi data pelatihan dan pengujian, gunakan train_test_split. Model MultinomialNB kemudian dibangun menggunakan data pelatihan, dan data pengujian digunakan untuk model dalam memprediksi label teks. Kode ini efektif dalam melakukan klasifikasi teks dengan cepat dan akurat, bahkan dengan asumsi independensi antar fitur.

	Dokumen	Kata	TF-IDF
0	0	benci	0.5634443855656935
1	0	dinasty	0.5634443855656935
2	0	progamnya	0.6042026553640724
3	1	rajawali	0.7400231759112181
4	1	ribu	0.6725813698834322
5	2	abang	0.5570125444742258
6	2	sumbangsih	0.8305040790377544
7	3	iya	0.2801676786987196
8	3	libur	0.48497990528258567
9	3	natal	0.4250062661564805

Gambar 4. Hasil TF-IDF

Gambar di atas menunjukkan TF-IDF (Term Frequency-Inverse Document Frequency), yang mengubah teks menjadi representasi numerik dengan menilai frekuensi kata dalam dokumen dan pentingnya kata tersebut dalam seluruh koleksi teks. Teknik ini membantu menyoroti kata-kata yang lebih relevan dan informatif untuk meningkatkan kualitas model dalam analisis teks.

```
X = data['teks']  # Kolom 'teks' yang berisi data teks
y = data['Sentiment']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=100)
```

Gambar 5. Pembagian Data Latih dan Uji

Gambar di bawah ini menunjukkan confusion matrix, yang digunakan untuk mengevaluasi kinerja model klasifikasi. Matriks ini memperlihatkan jumlah prediksi yang benar dan salah untuk setiap kelas yang ada dalam data pengujian.

```
print(confusion_matrix(y_test, y_pred))
```

```
[[ 2 15  0]
 [ 2 19  0]
 [ 1 11 10]]
```

Gambar 6. Confusion matrix

Matriks ini memperlihatkan jumlah prediksi yang benar dan salah untuk setiap kelas yang ada dalam data pengujian. Pada confusion matrix, terdapat empat komponen utama: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

```
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
Accuracy: 0.5166666666666667
              precision    recall  f1-score   support

     NEGATIF       0.40      0.12      0.18        17
      NETRAL       0.42      0.90      0.58        21
     POSITIF       1.00      0.45      0.62        22

    accuracy                           0.52        60
   macro avg       0.61      0.49      0.46        60
weighted avg       0.63      0.52      0.48        60
```

Gambar 7. Hasil Klasifikasi

Accuracy Model memiliki akurasi 51,67%, menunjukkan lebih dari setengah prediksi model sesuai dengan label sebenarnya. Precision Kelas POSITIF memiliki presisi tertinggi (100%), sedangkan NEGATIF rendah (40%). Recall Kelas NETRAL memiliki recall tertinggi (90%), sementara NEGATIF sangat rendah (12%). F1-Score Kelas NETRAL (0.58) dan POSITIF (0.62) cukup baik, tetapi NEGATIF (0.18) sangat buruk. Dapat dibilang bahwa model kurang seimbang dalam memprediksi, performa rendah terutama pada kelas NEGATIF.

4. KESIMPULAN

Analisis sentimen dengan Naïve Bayes membantu memahami persepsi masyarakat terhadap program mudik gratis Kota Medan 2024, mengidentifikasi sentimen dominan, dan memberikan wawasan untuk evaluasi serta perbaikan program. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program mudik gratis Kota Medan 2024 dengan menggunakan metode Naïve Bayes. Berdasarkan analisis terhadap 394 komentar yang dikumpulkan dari akun Instagram resmi Wali Kota Medan, diperoleh hasil klasifikasi sentimen sebagai berikut: 110 komentar (27.9%) termasuk dalam kategori positif, 107 komentar (27.2%) negatif, dan 177 komentar (44.9%) netral. Evaluasi kinerja model menunjukkan bahwa metode Naïve Bayes menghasilkan akurasi sebesar 51.667%. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sedang, tetapi masih memiliki kelemahan dalam mendeteksi sentimen negatif, yang ditunjukkan oleh rendahnya recall pada kelas negatif.

Dari analisis confusion matrix, diketahui bahwa kelas positif memiliki presisi tertinggi (100%), sementara kelas negatif memiliki recall yang rendah (12%), menunjukkan bahwa model kurang mampu mengenali komentar negatif secara efektif. F1-score untuk kelas netral (0.58) dan positif (0.62) cukup baik, tetapi kelas negatif (0.18) menunjukkan performa yang buruk. Hasil penelitian ini mengindikasikan bahwa ketidakseimbangan kelas dan kualitas data berpengaruh terhadap kinerja model. Oleh karena itu, untuk meningkatkan akurasi dan performa model di masa mendatang, dapat dilakukan peningkatan kualitas data, pemilihan fitur yang lebih relevan, serta eksplorasi metode klasifikasi lain sebagai perbandingan. Dengan perbaikan tersebut, diharapkan analisis sentimen dapat memberikan hasil yang lebih akurat dan dapat menjadi dasar bagi perbaikan kebijakan program mudik gratis di tahun-tahun berikutnya.

DAFTAR PUSTAKA

[1] "Pemanfaatan Media Kompas Kiblat RHI Dalam Pengakurasian Arah Kiblat Di Lingkungan Ranting Muhammadiyah Barus Mudik Tapanuli Tengah," *Ihsan: Jurnal Pengabdian Masyarakat*, Apr. 2022, doi: 10.30596/ihsan.v4i1.9802.

[2] Robie Fanreza, Munawir Pasaribu, and Arwin Juli Rakhmadi Butar-Butar, "Sentiment Analysis Of Public Comments On Quick Response Code Indonesian Standard (Qris) On Twitter Social Media Using The Naïve Bayes Classifier Method," vol. 5, no. 2, pp. 276–286, 2024.

[3] K. Dwi Ningtyas and R. Kurniawan, "Penerapan Natural Language Processing Pada Aplikasi Chatbot Info Layanan Kantor Menggunakan Naïve Baiyes Algorithm," *Januari*, 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index

[4] Muhammad Suhery, Gema Ramadhan, and Abdul Halim Hasugian, "Analisis Sentimen Masyarakat Kota Medan Terhadap Persiapan PON XXI Sumut-Aceh 2024 Menggunakan Metode Naive Bayes," *Polygon : Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam*, vol. 2, no. 5, pp. 10–18, Aug. 2024, doi: 10.62383/polygon.v2i5.207.

[5] T. H. Apandi and C. A. Sugianto, "Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP (Naive Bayes Algorithm for Satisfaction Prediction of e-ID Card Recording Service)," 2019.

[6] Z. Attoriq, F. Wahyudi, and A. H. Hasugian, "Science Midwifery Analysis of side effects from using influenza drugs using the Naïve Bayes method," Online, 2024. [Online]. Available: www.midwifery.iocspublisher.orgJournalhomepage:www.midwifery.iocspublisher.org

[7] Yeni Safitri, Rakhmat Kurniawan, and Suhardi, "Analisis Sentimen Mengenai Childfree Menggunakan Metode Naïve Bayes," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, Jul. 2024, doi: 10.33022/ijcs.v13i4.4136.

[8] V. S. Ginting, K. Kusrini, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, Jun. 2020, doi: 10.35585/inspir.v10i1.2535.

[9] A. H. Hasugian, R. A. Putri, and M. Khairani, "Analysis of Public Sentiment towards Beverage Franchise Reviews Using the Naïve Bayes Algorithm", doi: 10.54209/jurnalkomputer.v15i01.102.

[10] E. Darwisah Harahap and R. Kurniawan, "Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes," 2024.

[11] irwan, *Analisis Sentimen Terhadap Tindakan Pemerintah Indonesia Untuk Menampung Sementara Pengungsi Etnis Rohingya Menggunakan Naïve Bayes Classifier*. Hogarth Press ; Institute of Psycho-analysis, 2024.

[12] R. R. Harahap and Mhd. Furqan, "Sentiment Analysis towards the 2024 Vice Presidential Candidate Debate Using the Support Vector Machine Algorithm," *sinkron*, vol. 8, no. 3, pp. 1783–1794, Jul. 2024, doi: 10.33395/sinkron.v8i3.13903.

[13] G. Septania Al-Husna, D. Asmarajati, I. Ahmad Ihsanuddin, R. Mahmudati, U. Sains Al-Qur, and R. Artikel, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Linkedin 1) Info Artikel Abstrak," vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123.

[14] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: 10.58602/jics.v2i2.24.

[15] Hawraa Fadhil Khelil, Mohammed Fadhil Ibrahim, Hafsa Ataallah Hussein, and Raed Kamil Naser, "Evaluation of Different Stemming Techniques on Arabic Customer Reviews," *Journal of Techniques*, vol. 6, no. 2, pp. 1–8, Feb. 2024, doi: 10.51173/jt.v6i2.2313.