

PENERAPAN TEXT MINING MENGGUNAKAN METODE MULTINOMIAL NAIVE BAYES PROGRAM MAKAN SIANG GRATIS PADA X

Muhammad Jibran Abdurrahman¹, Arief Wibowo^{2*}

^{1,2} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹2011501539@student.budiluhur.ac.id, ^{2*}arief.wibowo@budiluhur.ac.id

(* : corresponding author)

Abstrak- Makan siang gratis adalah salah satu program unggulan dari capres dan cawapres prabowo-gibran yang bertujuan untuk meningkatkan gizi anak dan kecukupan ibu hamil. Makan siang gratis memberikan kepada siswa sekolah dasar(SD), sekolah menengah pertama (SMP), sekolah menengah akhir (SMA), pesantren, dan ibu hamil. Program makan siang gratis menuai banyak tanggapan dari masyarakat, dengan sentimen yang bervariasi seperti positif, negatif, dan netral. Sehingga dilakukan penelitian menggunakan teknik *text mining* untuk mengetahui sentimen masyarakat berdasarkan komentar yang bersumber dari salah satu platform media sosial yaitu X. dalam *text mining* menggunakan *multinomial naive bayes* sebagai metode, untuk klasifikasi sentimen positif, negatif, atau netral pada komentar. Data penelitian terdapat 2004 komentar yang melalui tahap *preprocessing* terpankas menjadi 1007 serta menjadikan dataset lebih terstruktur dan terfokus terhadap program makan siang gratis. Setelah itu, dilakukan pelabelan menggunakan kamus leksikon yang kemudian divalidasi oleh pakar yang menghasilkan sentimen positif 680, netral 75, dan negatif 252. penggunaan *tf-idf* untuk mentransformasi kata menjadi angka serta menggunakan 2 rasio yaitu 70:30 dan 80:20. Metode *multinomial naive bayes* pada rasio 70:30 berhasil mendapatkan tingkat akurasi sebesar 63,58% dan 80:20 memperoleh 57,21%. Dari hasil pelabelan validasi pakar terhadap program unggulan capres dan cawapres prabowo-gibran, tanggapan masyarakat berdasarkan hasil penelitian lebih unggul sentimen positif yang menunjukkan masyarakat mendukung program unggulan makan siang gratis.

Kata Kunci: *multinomial naive bayes, text mining, program makan siang gratis, X*

APPLICATION OF TEXT MINING USING MULTINOMIAL NAIVE BAYES METHOD FREE LUNCHES PROGRAM AT X

Abstract- *Free lunch is one of the flagship programs of the presidential and vice presidential candidate, prabowo-gibran, which aim to improve child nutrition and the adequacy of pregnant women. Free lunches are provided to elementary school student (SD), junior high school (SMA), pesantren, and pregnant women. The free lunch program garnered a lot of responses from the community, with varying sentiments such as positive, negative, and neutral. So research was conducted using text mining techniques to find out the assesment of the community based on comments sources from one of the social media platforms, namely X. in text mining using multinomial naive bayes as a method, for the classification of positive, negative, and neutral sentiments on comments. The research data contained 2004 comments which through the preprocessing stage trimmed to 1007 and make the dataset more structured and focused on the free lunch program. After that, labelling is done using a lexicon dictionary which is then be validated by experts who produced 680 positive, 75 neutral, and 252 negative sentiments. Tf-idf is used to transform words into numbers and use 2 ratios, namely 70:30 and 80:20. The multinomial naive bayes method at a ratio of 70:30 managed to get an accuracy rate of 63,58% and 80:20 obtained 57,21%. From the results of experts validation labeling of the presidential and vice presidential candidate prabowo-gibran's flagship program, community responses based on research results are superior to positive sentiments that show the community support the free lunches flagship program.*

Keywords: *multinomial naive bayes, text mining, free lunch program, X*

1. PENDAHULUAN

Tingkat *stunting* indonesia adalah salah satu yang tertinggi di dunia, terutama pada anak dan balita. Data yang dikumpulkan oleh kementerian kesehatan republik indonesia 2024, menunjukkan bahwa sekitar 24% orang menderita *stunting* pada tahun 2021. Data ini turun sekitar 3% menjadi 21% pada tahun 2024. Namun, meskipun penurunan, indonesia tetap menjadi negara tertinggi di dunia dalam kasus *stunting*, terutama pada anak dan balita, menurut organisasi kesehatan dunia (WHO). Salah satu inisiatif yang dilakukan oleh pasangan capres-cawapres 02 adalah menyediakan makan siang gratis untuk anak yang masih sekolah [1].

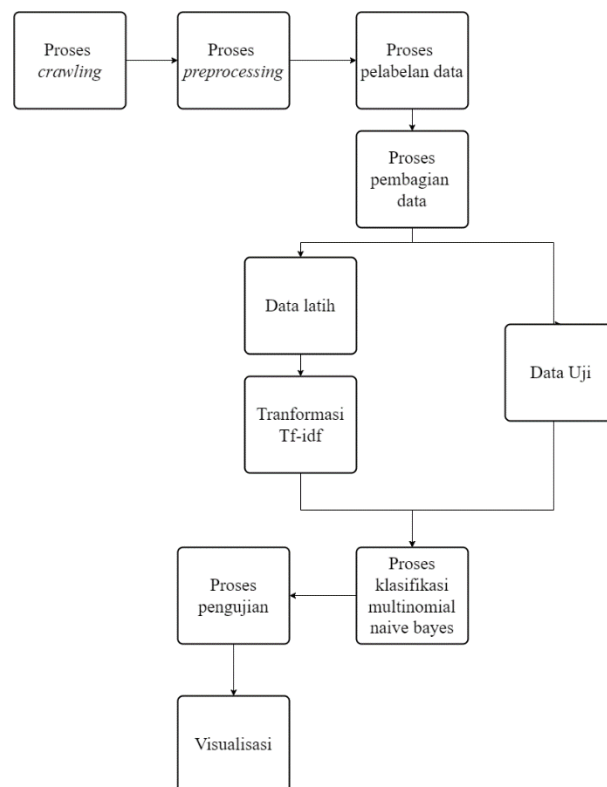
Dengan menawarkan makan siang gratis, Prabowo-Gibran berusaha memberikan perlindungan sosial kepada mereka yang rentan terhadap kemiskinan dan ketidaksetaraan, serta menjamin bahwa semua orang dapat berpartisipasi dalam kemajuan bangsa [2]. Salah satu program bantuan gizi yang memberikan makan dan susu gratis kepada anak balita dan ibu hamil disekolah dan pesantren. Anak di prasekolah, sekolah dasar, sekolah menengah pertama, sekolah menengah atas, dan pesantren mendapatkan makan siang gratis setiap hari. Program makan siang gratis memiliki tujuan yaitu dikategorikan ke dalam indikator pencapaian. Berbagi susu gratis kepada ibu hamil lebih relevan sebagai penanganan *stunting* daripada upaya untuk mengatasi melalui program ini. Program makan siang gratis tampaknya memiliki dampak positif di banyak negara. Makan siang gratis ini meningkatkan nutrisi anak, partisipasi sekolah, karakter siswa, dan ekonomi keluarga di seluruh negeri [3].

Program makan siang gratis dari capres dan cawapres Prabowo-Gibran menuai banyak respon dari masyarakat. Dalam penelitian ini melakukan proses pembuatan program dengan *text mining* untuk melakukan klasifikasi. Sentimen analisis atau opini mining adalah istilah untuk *text mining* yang mencakup pengenalan sentimen atau polaritas pada komentar publik [4]. Beberapa proses penelitian sebelumnya topik yang diangkat yaitu analisis sentimen produk skincare *Somecinc* *Niacinamide* di *Female Daily* menggunakan rasio 80:20 dan 3080 *review*. Metode yang digunakan penelitian adalah *Multinomial Naive Bayes* berhasil mendapatkan hasil akurasi sebesar 61% [5]. Lalu dalam penelitian analisis sentimen berfokuskan terhadap sentimen pengguna aplikasi peduli lindungi menggunakan 3 rasio yaitu 70:30, 80:20, dan 90:20 dan total 7.500 ulasan. Metode yang digunakan pada penelitian ini adalah *k-nearest neighbor* dengan hasil akurasi sebesar 60,66% pada rasio 90:10 dan nilai $K=2$ [6].

Sehingga penelitian bertujuan untuk mengetahui penilaian sentimen masyarakat terhadap program unggulan capres dan cawapres Prabowo-Gibran. Peneliti mengimplementasikan *Multinomial Naive Bayes* sebagai metode dalam penelitian klasifikasi komentar masyarakat pada sosial media X, serta mengetahui seberapa baik metode digunakan pada penelitian ini.

2. METODE PENELITIAN

Proses program yang dibangun untuk melakukan klasifikasi kasus program makan siang gratis menggunakan metode *Multinomial Naive Bayes*. Proses dari pembangunan program memiliki tahapan, Gambar 1 merupakan diagram alur dari metode penelitian.



Gambar 1. Metode penelitian

2.1 Text Mining

Sentimen analisis atau opini mining adalah istilah untuk *text mining* yang mencakup pengenalan sentimen atau polaritas pada komentar publik [4]. *text mining* merupakan algoritma dari penambangan teks agar dapat melengkapkan permasalahan kebutuhan informasi menggunakan metode *data mining*, *machine learning*, *natural language processing*, pencarian informasi, dan manajemen pengetahuan [7].

2.2 Pengumpulan data

Metode pengambilan data melalui crawling, dimana data berasal dari tweet yang terkait dengan platform media sosial X. proses ini dilakukan berdasarkan bahasa pemrograman python, untuk mendapatkan akses ke token API X menggunakan kata kunci dan disimpan kedalam file format .csv [8]. Setelah proses mendapatkan token, selanjutnya proses crawling dilakukan pada platform google collab dengan Bahasa python. proses pengumpulan komentar dibutuhkan keyword agar memperoleh dataset sesuai dengan relevansi kasus seperti “makan siang gratis”, “program makan gratis”, dan “susu dan makan siang gratis”. Dataset yang berhasil diperoleh sebanyak 2004 komentar dalam rentang waktu 29 maret sampai dengan 9 mei 2024.

2.3 Text Preprocessing

Dataset yang telah diperoleh dari hasil crawling masih belum dapat diproses, karena masih memiliki unsur simbol dan kata yang tidak digunakan pada data. Proses *preprocessing* dilakukan untuk membuat dataset tersebut menjadi lebih terstruktur dan terorganisir untuk dapat dilakukan klasifikasi [9]. Tahap *preprocessing* memiliki beberapa tahapan yaitu *cleaning*, *casefolding*, *tokenizing*, *normalization*, *stopword*, dan *stemming*.

a. Cleaning

Cleaning adalah tahap pemrosesan pembersihan teks dengan menghapus simbol, angka, tanda baca, spasi berlebihan, kata terdiri satu sampai dua, username, hashtags, dan links [10].

b. Casefolding

Data yang diperoleh dari hasil pengumpulan data masih memiliki huruf kapital yang tidak konsisten. pemrosesan casefolding menjadikan huruf pada seluruh dataset menjadi kecil [11].

c. Tokenizing

proses pemisahan kalimat menjadi kata kecil yang bermakna atau token seperti makan siang gratis menjadi “makan”, “siang”, “gratis” [12].

d. Normalization

Proses normalisasi memiliki tujuan mengurangi variasi dan meningkatkan konsistensi, dalam hal ini membantu mempermudah perbandingan, penggunaan data, atau analisis dalam berbagai konteks [3].

e. Stopword

Stopword memiliki kata umum seperti “dan”, “atau”, “yang”, “di”, “dari”, dan lainnya, tergantung dengan konteks serta bahasa [3].

f. Stemming

Stemming adalah tahapan penghapusan imbuhan pada awal teks seperti kata “makanannya” menjadi “makanan” [13].

2.4 Pelabelan

Proses pelabelan menggunakan leksikon dengan kamus inset terdapat kata dan bobot dalam menentukan kata tersebut. Perhitungan polaritas dilakukan dengan menjumlahkan bobot dari kata dalam setiap komentar, hasil digunakan untuk menandakan positif atau negatif. Komentar dikatakan positif jika polaritas positif, dan negatif apabila polaritas negatif [14].

2.5 Multinomial Naïve Bayes

Multinomial naive bayes adalah metode klasifikasi menggunakan pembelajaran secara *supervised* dengan model probabilistik. Proses perhitungan peluang kemunculan suatu kelas dimulai apabila data telah dimasukkan untuk digunakan pembelajaran pada data latih. Proses perhitungan berasal dari probabilitas maka kata i termasuk dari kategori atau kelas tertentu [15].

$$P(c) = \frac{N_c}{N_{doc}} \quad (1)$$

Keterangan:

C : Kelas atau kategori
 doc : Dokumen
 N_c : Dokumen latih mengandung banyak kategori C
 N_{doc} : Jumlah besar dokumen latih yang digunakan

$$P(w_i, c) = \frac{\text{count}(w_i, c) + 1}{\sum_w \text{count}(w, c) + |V|} \quad (2)$$

Keterangan:

w_i : kata ke-i di seluruh dokumen yang berkategori
 $\text{Count}(w_i, c)$: Jumlah kata tertentu yang ditemukan dalam kelas atau kategori tertentu
 $\sum_w \text{count}(w, c)$: Jumlah total kata dalam kelas
 $|V|$: merupakan total jumlah kelas unik

2.6 X

Proses perhitungan berasal dari probabilitas maka kata i termasuk dari kategori atau kelas tertentu X . Pendiri X adalah Jack Dorsey, Biz Stone, Evan Williams, dan Noah Glass pada tahun 2006, yang kemudian dibeli oleh Elon Musk pada tahun 2022. Pengguna media sosial X di Indonesia terbukti sangat pesat setiap hari dengan mengirimkan 500 juta komentar atau kicauan, yang digunakan untuk berbagi informasi, termasuk fakta dan opini, serta mengunggah informasi pribadi [16].

2.7 TF-IDF

TF-IDF adalah *term frequency – inverse document frequency* proses transformasi teks menjadi bentuk vektor yang mudah dipahami oleh komputer. Kata yang diproses dapat dipahami apakah informatif atau tidak dengan tf-idf. Tf merupakan frekuensi dari kemunculan kata terhadap sebuah dokumen. Idf adalah *Invers frekuensi dokumen* [15].

$$tf_{t,d} = \frac{n_{t,d}}{\text{jumlah total terms di dokumen}} \quad (3)$$

$$idf_d = \log_{10}\left(\frac{\text{jumlah dokumen}}{\text{jumlah dokumen dengan term}}\right) \quad (4)$$

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (5)$$

2.8 Confusion matrix

Confusion matrix adalah tahap evaluasi model untuk memberikan informasi mengenai seberapa baik mode klasifikasi data, dengan melakukan perbandingan. Perbandingan yang dilakukan pada tahap evaluasi adalah nilai hasil sebenarnya dengan klasifikasi [17]. Evaluasi dengan confusion matrix dengan empat jenis nilai aktual dan prediksi dapat mencari nilai parameter. Terdapat empat parameter dari confusion matrix seperti accuracy, recall, precision, dan f1-score [18].

3. HASIL DAN PEMBAHASAN

Pada bagian ini, metodologi penelitian dapat dibuat terlebih dahulu dengan membahas topik penelitian, hasil implementasi dan pengujian.

3.1 Tahapan *Crawling*

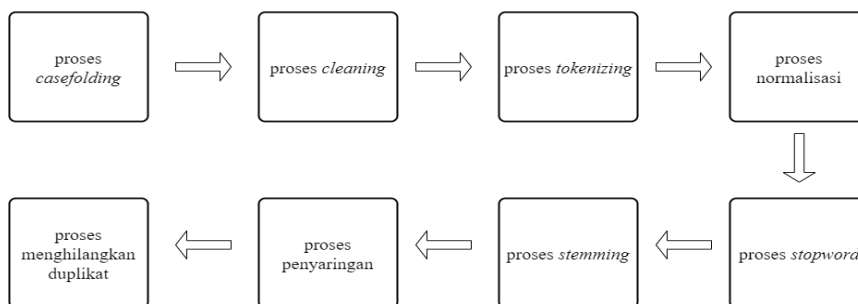
Data yang diperoleh pada media sosial X, menggunakan teknik bernama *crawling*, proses ini merupakan salah satu dari langkah *text mining* dalam pengumpulan dataset. Proses *crawling* ini menggunakan bahasa *python* dan platform *google collab* untuk memperoleh dataset. Jarak waktu perolehan dataset dari tanggal 29 maret sampai 9 mei 2024, berhasil mendapatkan 2004 komentar menggunakan kata kunci “makan siang gratis”, “program makan siang gratis”, dan “makan dan susu gratis”. Attribute yang digunakan pada program ini terdapat 3 yaitu *Id*, tanggal, dan komentar. Hasil *crawling* dapat dilihat pada tabel 1.

Tabel 1. Hasil *Crawling*

No	Tanggal	Komentar
1	Thu May 09 23:49:49 2024	@alisyarief Anaknya seneng tuh dijahati sama bapaknya
2	Wed May 08 04:46:13 2024	@tempodotco Kementerian makan siang gratis enaknya di singkat jadi apa ya ?
3	Sat Mar 30 15:02:48 2024	@8angSaid Jogedin aja. Kan dapet makan siang dan susu gratis. Dungu tetap dipelihara. Jangan dibuang. Hehehe
...
2004	Sun Mar 31 00:12:21 2024	@koreahole @tanyarlifes Programnya pak Anies ada lho buat guru honorer tapi kebanyakan orang lebih milih program makan gratis

3.2 Tahapan *preprocessing*

Tahapan *preprocessing* merupakan Langkah untuk membersihkan dan menyaring dataset. proses ini terdiri dari beberapa tahapan dalam mengubah dataset menjadi terstruktur seperti *casefolding*, *cleaning*, *tokenizing*, *normalization*, *normalisasi*, *stopword*, dan *stemming*. Tahapan *preprocessing* dapat dilihat pada gambar 2.



Gambar 2. Tahapan *preprocessing*

3.3 Pelabelan

Pemberian label pada komentar adalah untuk memberikan emosi kepada komentar agar dapat diketahui pesan dan kesan dari komentar. Sentimen dari positif yaitu sifat yang menunjukkan dukungan atau persetujuan, sedangkan negatif menggambarkan ketidakpuasan dari pandangan komentar. Netral merupakan sentimen yang tidak menunjukkan sisi positif dan negatif. Proses pelabelan ini menggunakan metode leksikon untuk mengetahui sentimen dari komentar. Setelah melalui proses pelabelan dengan leksikon, hasil diberikan kepada pakar untuk dilakukan validasi sentimen. Hasil pelabelan validasi pakar dapat dilihat pada tabel 2.

Tabel 2. Hasil Pelabelan

No	Komentar	Sentimen
1	selektif tentu anggaran belanja negara program makan siang gratis ibu kota nusantara makan biaya jalan	Negatif
2	manja subsidi selesai ibu kota nusantara makan siang gratis tunjuk bakti negara bangsa	Positif

No	Komentar	Sentimen
3	masyarakat akses informasi makan siang gratis atas tengkes masyarakat informasi	Netral

Sentimen yang digunakan adalah hasil yang telah divalidasi oleh pakar. Pada proses ini juga dilakukan pembagian data yang dimana untuk membuat data menjadi 2 bagian yaitu Data latih dan uji. Setiap dataset dipisah dengan 2 rasio yaitu 70:30 dan 80:20. Hasil dari pembagian data latih dan uji dapat dilihat pada tabel 3.

Tabel 3. Hasil Pembagian Dataset

Data	Rasio
Data latih	705(70%)
Data Uji	806(80%)
Total Data	1007

3.4 Term *Frequency-Inverse Document Frequency*

Proses transformasi teks yang dipecah menjadi kata, untuk memberikan bobot. proses teknik tf-idf dilakukan pada dataset latih dan uji agar kata pada kedua data memiliki bobot. Tahap pertama adalah menghitung tf pada kata, setelah itu idf, dan terakhir tf-idf.

- a. Perhitungan TF adalah mencari nilai frekuensi berdasarkan tingkat kemunculan pada sebuah kalimat. Perhitungan dapat dilihat pada tabel 4.

Tabel 4. Perhitungan TF

Kata	TF	Nilai TF
Masuk	1/8	0,125
Program	1/8	0,125
Kerja	1/8	0,125
Prabowo	1/8	0,125
Gibran	1/8	0,125
Makan	1/8	0,125
Siang	1/8	0,125
Gratis	1/8	0,125

- b. Perhitungan IDF adalah menghitung tingkat kemunculan kata pada seluruh dokumen. Hasil perhitungan dapat dilihat pada tabel 5.

Tabel 5. Hasil Perhitungan IDF

Kata	IDF	Nilai IDF
Masuk	1	1,301
Program	1	1,301
Kerja	1	1,301
Prabowo	1	1,301
Gibran	1	1,301
Makan	4	0,903
Siang	4	0,903
Gratis	4	0,903

- c. Perhitungan TF-IDF adalah mencari frekuensi kata dengan mengkalikan hasil nilai TF dan IDF. Hasil perhitungan dapat dilihat pada tabel 6.

Tabel 6. Hasil Perhitungan TF-IDF

Kata	TF	IDF	Nilai TF-IDF
Masuk	0,125	1,301	0,163
Program	0,125	1,301	0,163
Kerja	0,125	1,301	0,163
Prabowo	0,125	1,301	0,163
Gibran	0,125	1,301	0,163
Makan	0,125	0,903	0,113
Siang	0,125	0,903	0,113
Gratis	0,125	0,903	0,113

3.5 Multinomial Naive Bayes

Proses klasifikasi *multinomial naive bayes* adalah menentukan prediksi label. Setelah melalui proses transformasi, hasil *tfidf* digunakan untuk menentukan sentimen prediksi. Proses pada klasifikasi terdapat perhitungan *prior probability*, penjumlahan *tfidf*, *likelihood*, dan menghitung prediksi pada sebuah dokumen.

a. Prior Probability

Perhitungan ini menghitung *prior probability* dengan seluruh sentimen dan total dokumen. perhitungan *prior probability* dapat dilihat pada tabel 7.

Tabel 7. Hasil Perhitungan Prior Probability

Sentimen	Menghitung Prior	Nilai Prior
Positif	$\text{Log}_{10}(2/4)$	-0,30103
Negatif	$\text{Log}_{10}(1/4)$	-0,60206
Netral	$\text{Log}_{10}(1/4)$	-0,60206

b. Likelihood

Dalam perhitungan *likelihood* adalah menghitung berdasarkan *prior probability*, kata unik, dan bobot *tf-idf* berdasarkan dari setiap sentimen.

- Menghitung nilai *likelihood* pada sentimen positif. Perhitungan dapat dilihat pada tabel 8.

Tabel 8. Hasil Perhitungan Likelihood Positif

Kata sentimen	Menghitung <i>likelihood</i>	<i>likelihood</i>
inspiratif positif	$\text{Log}_{10}(0+1/(0,748+20))$	-1,317
moga positif	$\text{Log}_{10}(0+1/(0,748+20))$	-1,317
program positif	$\text{Log}_{10}(0,163+1/0,748+20))$	-0,675
makan positif	$\text{Log}_{10}(0,195+1/0,748+20))$	-0,614
siang positif	$\text{Log}_{10}(0,195+1/0,748+20))$	-0,614
gratis positif	$\text{Log}_{10}(0,195+1/0,748+20))$	-0,614
adopsi positif	$\text{Log}_{10}(0+1/(0,748+20))$	-1,317
indonesia positif	$\text{Log}_{10}(0+1/(0,748+20))$	-1,317

- Menghitung nilai *likelihood* pada sentimen negatif. Perhitungan dapat dilihat pada tabel 9.

Tabel 9. Hasil Perhitungan Likelihood Negatif

Kata sentimen	Menghitung <i>likelihood</i>	<i>likelihood</i>
inspiratif negatif	$\text{Log}_{10}(0+1/(0,387+20))$	-1,309
moga negatif	$\text{Log}_{10}(0+1/(0,387+20))$	-1,309
program negatif	$\text{Log}_{10}(0+1/(0,387+20))$	-1,309
makan negatif	$\text{Log}_{10}(0,129+1/(0,387+20))$	-0,749
siang negatif	$\text{Log}_{10}(0,129+1/(0,387+20))$	-0,749
gratis negatif	$\text{Log}_{10}(0,129+1/(0,387+20))$	-0,749
adopsi negatif	$\text{Log}_{10}(0+1/(0,387+20))$	-1,309
indonesia negatif	$\text{Log}_{10}(0+1/(0,387+20))$	-1,309

- Menghitung nilai *likelihood* pada sentimen netral. Perhitungan dapat dilihat pada tabel 10.

Tabel 10. Hasil Perhitungan Likelihood Netral

Kata sentimen	Menghitung <i>likelihood</i>	<i>likelihood</i>
inspiratif netral	$\text{Log}_{10}(0+1/(0,678+20))$	-1,316
moga netral	$\text{Log}_{10}(0+1/(0,678+20))$	-1,316
program netral	$\text{Log}_{10}(0+1/(0,678+20))$	-1,316
makan netral	$\text{Log}_{10}(0,226+1/0,678+20))$	-0,562
siang netral	$\text{Log}_{10}(0,226+1/0,678+20))$	-0,562
gratis netral	$\text{Log}_{10}(0,226+1/0,678+20))$	-0,562
adopsi netral	$\text{Log}_{10}(0+1/0,678+20))$	-1,316

Kata sentimen	Menghitung <i>likelihood</i>	<i>likelihood</i>
indonesia netral	$\text{Log}10(0+10,678+20)$	-1,316

c. Menghitung Prediksi

Selanjutnya dalam menghitung prediksi menggunakan nilai dari *likelihood* yang sudah dilakukan. Perhitungan *likelihood* ini telah menghitung berdasarkan dari ke- 3 sentimen positif, negatif, dan netral. Berdasarkan hasil perhitungan prediksi dari sentimen komentar uji adalah positif. Proses dari perhitungan prediksi dapat dilihat pada tabel 11.

Tabel 11. Hasil Perhitungan Prediksi

Sentimen	Perhitungan Prediksi	Prediksi
Positif	$-0,301+(-1,317)+(-1,317)+(-0,675)+(-0,614)+(-0,614)+(-0,614)+(-1,317)+(-1,317)$	-8,086
Negatif	$-0,602+(-1,309)+(-1,309)+(-1,309)+(-0,749)+(-0,749)+(-0,749)+(-1,309)+(-1,309)$	-9,397
Netral	$-0,602+(-1,316)+(-1,316)+(-1,316)+(-0,562)+(-0,562)+(-0,562)+(-1,316)+(-1,316)$	-8,865

3.6 Pengujian

Pada tahap pengujian klasifikasi *multinomial naive bayes* menggunakan *confusion matrix*, untuk mengetahui nilai akurasi, presisi, *recall*, dan *f1-score*. dari penelitian. Dari kedua pengujian menggunakan rasio 70:30 dan 80:20. Pada tahap pengujian klasifikasi *multinomial naive bayes* menggunakan *confusion matrix*, untuk mengetahui nilai akurasi, presisi, *recall*, dan *f1-score*. dari penelitian. Dari kedua pengujian menggunakan rasio 70:30 dan 80:20. *Confusion matrix* terhadap metode *multinomial naive bayes* dari ke-2 rasio dapat dilihat pada tabel 12.

Tabel 12. *Confusion Matrix*

Aktual/predik 70:30	Positif	Negatif	Netral
Positif	187	67	24
Negatif	1	5	1
Netral	15	2	0
Aktual/predik 80:20	Positif	Negatif	Netral
Positif	95	36	11
Negatif	12	17	2
Netral	19	6	3

Berdasarkan pengujian *confusion matrix* menggunakan pelabelan yang dilakukan oleh pakar dan metode klasifikasi *multinomial naive bayes* diperoleh nilai akurasi rasio 70:30 sebesar 63.58%, presisi 46.23%, *recall* 32.96%, dan *f1-score* 30.03%. untuk rasio 80:20 mendapatkan sebesar 57.21%, presisi 44.15%, *recall* 40.99%, dan *f1-score* 40.77%. Hasil pengujian *confusion matrix* keakuratan tertinggi dihasilkan rasio 70:30 mendapatkan nilai terbesar yaitu 63.58%. hasil pengujian *confusion matrix* metode *multinomial naive bayes* dari ke-2 rasio dapat dilihat pada tabel 13.

Tabel 13. Hasil Pengujian

Latih	Uji	Akurasi	Presisi	<i>Recall</i>	<i>F1score</i>
70(705)	30(302)	63.58%	46.23%	32.96%	30.03%
80(806)	20(201)	57.21%	44.15%	40.99%	40.77%

4. KESIMPULAN

Hasil pengujian sistem *text mining* yang dibuat menggunakan data dan metode *multinomial naive bayes* Dataset yang diperoleh sebesar 2004 komentar didapatkan dari media sosial X mengenai kasus program makan siang gratis, dengan pengujian klasifikasi *multinomial naive bayes* dan menggunakan dua rasio pembagian data berbeda. Pelabelan dilakukan dengan kamus leksikon yang dilakukan validasi oleh pakar, menghasilkan

sentimen positif 680, netral 75, dan negatif 252. Rasio yang digunakan pada penelitian ini adalah 70:30 dan 80:20 dengan tingkat akurasi yang berbeda. Pengujian *multinomial naive bayes* menghasilkan rasio 70:30 mendapatkan tingkat akurasi terbesar yaitu 63,58% dan 80:20 memperoleh hanya 57,21%. Berdasarkan hasil pelabelan pakar sentimen terbesar diperoleh positif sebesar 680, Maka tanggapan dan persepsi masyarakat terhadap program makan siang gratis sangat positif. Adapun penelitian masih dibutuhkan pengembangan seperti peningkatan akurasi dengan memperbanyak data yang digunakan, memaksimalkan proses preprocessing, dan metode klasifikasi yang lebih maju.

DAFTAR PUSTAKA

- [1] F. Ika Septiani, N. Rosiana, and A. Azzahra, "Dampak Makan Siang Gratis Pada Kondisi Keuangan Negara Dan Peningkatan Mutu Pendidikan," vol. 1, pp. 192–196, 2024.
- [2] P. Ardella Maharani, A. Riyani Namira, and T. Viony Chairunnisa, "Peran Makan Siang Gratis Dalam Janji Kampanye Prabowo Gibran Dan Realisasinya," 2024.
- [3] F. N. Zaman, M. A. Fadhilah, M. A. Ulinuha, and K. Umam, "Menganalisis Respons Netizen Twitter Terhadap Program Makan Siang Gratis Menerapkan NLP Metode Naïve Bayes," 2024. [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [4] A. S. Aribowo and S. Khomsah, "Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19) Implementasi Text Mining Untuk Deteksi Emosi Menggunakan Metode Leksikon (Studi Kasus: Twit Tentang Covid-19)," *J. Inform. dan Teknol. Inf.*, vol. 18, no. 1, pp. 49–60, 2021, doi: 10.31515/telematika.v18i1.4341.
- [5] D. A. WP, J. D. Firizqi, and Z. A. Amalia, "Analisis Sentimen Produk Skincare Somethinc Niacinamide di Female Daily dengan Naïve Bayes Classifier," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 2, p. 946, Apr. 2024, doi: 10.30865/mib.v8i2.7571.
- [6] Sulfayanti, I. Fadhilah Nurul, and N. Nur, "Analisis Performa K-Nearest Neighbor Dalam Mengklasifikasi Sentimen Pengguna Aplikasi PeduliLindungi," *J. Comput. Inf. Syst. (J-CIS)*, vol. 7, no. 1, 2024, doi: 10.31605/jcis.v7i1.
- [7] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *Media Online*, vol. 2, no. 5, pp. 178–185, 2022, [Online]. Available: <https://djournals.com/klik>
- [8] Faradian Havadz, Rubhasy Albar, and Yunan Fauzi Wijaya, "Analisis Sentimen Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naïve Bayes Classifier Pada Media Sosial X," *J. Ilm. Sain dan Teknol.*, pp. 150–163, 2024.
- [9] H. Muzaki and H. Marcos, "Analisis Sentimen RUU KUHP Di Medsos Twitter Dengan Metode Naïve Bayes," *Jurnal Teknologi Informasi*, vol. 7, no. 1, pp. 47–50, 2023.
- [10] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [11] Z. N. Aulia, G. K. Jati, and I. Santoso, "Analisis Sentimen Tanggapanpublic Mengenai E-Tilang Melalui Media Sosial Youtube Menggunakan Algoritma Naive Bayes," *Jurnal IKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 150–156, 2023. [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2824/2077>
- [12] S. Wulandari and F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," 2024, doi: 10.30865/mib.v8i2.7520.
- [13] F. Nurrachmat Hidayat, "Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i1.1359.
- [14] W. Trisnawati and A. Wibowo, "Sentiment Analysis Of ICT Service User Using Naive Bayes Classifier And SVM Methods With TF-IDF Text Weighting," *Jurnal Teknologi Informasi*, vol. 5, no. 3, pp. 709–719, 2024, doi: 10.52436/1.jutif.2024.5.3.1784.
- [15] M. A. Java, M. Syafrullah, and F. Teknologi, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine," *Jurnal Ticom: Technology of Information and Communication*, vol. 12, no. 2, 2024.
- [16] F. M. Carina, Admi Salma, Dony Permana, and Zamahsary Martha, "Sentiment Analysis of X Application Users on the Conflict Between Israel and Palestine Using Support Vector Machine Algorithm," *UNP J. Stat. Data Sci.*, vol. 2, no. 2, pp. 204–212, 2024, doi: 10.24036/ujsds/vol2-iss2/170.
- [17] Nafisyah Siti and Sulistiyowati Rini, "Analisis Sentimen Ulasan Produk Toko Online Esrocte Untuk Peningkatan Pelayanan Menggunakan Algoritma Naïve Bayer," 2024. [Online]. Available: <https://blantika.publikasiku.id/>
- [18] F. R. Irawan, A. Jazuli, and T. Khotimah, "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbor," *JIKO: Jurnal Informatika dan Komputer*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.