

## Analisis Sentimen Program Makan Siang Gratis Menggunakan Multinomial Naive Bayes

Muhammad Lutfhi Akmal<sup>1)</sup>, Dody Pernadi<sup>2\*)</sup>, Jennie kusumaningrum<sup>3)</sup>

<sup>1,2\*)</sup> Program Studi Sistem Informasi, Universitas Gunadarma

<sup>3)</sup> Program Studi Teknik Sipil, Universitas Gunadarma

<sup>1)</sup>mluthfiakmal1933@gmail.com; <sup>2)</sup>dodypernadi@gmail.com; <sup>3)</sup>kusumaningrumjennie@gmail.com

### ABSTRACT

*Twitter is one of the social media networks used by the public to express opinions, criticisms and points of view. A topic that has been widely discussed is the free lunch work program promoted by the presidential candidate pair Prabowo and Gibran. The existence of pros and cons in assessing a policy or work program is very high, an approach is needed to analyze public sentiment towards the program. Analysis of public sentiment towards this program is important to provide an overview of how well the program is received by the public and how public opinion affects the program. This research aims to determine the sentiment category towards the free lunch program using the Naive Bayes method. This research involves collecting and analyzing tweets related to "Free Lunch Program" from Twitter(X), using authentication tokens. The data was processed through pre-processing, then classified with Multinomial Naive Bayes. A total of 1902 data were obtained from labeling with the Lexicon Based method. The results obtained 84.06% accuracy, 83.9% precision, 98.9% recall, and 90.70% F1-score calculated using confusion matrix. The sentiment analysis results show that the majority of community responses tend to be positive, in other words, the community supports and is optimistic about the free lunch work program which can directly provide social benefits to the community in fulfilling basic needs*

**Keywords:** *Sentiment Analysis, Lexicon Based, Multinomial Naive Bayes, Makan Siang Gratis, Twitter(X)*

### I. PENDAHULUAN

Media sosial telah menjadi *platform* utama bagi masyarakat umum untuk mengekspresikan opini dan sudut pandang di era digital saat ini. Twitter sebagai salah satu platform media sosial yang paling populer. Twitter(X) di Indonesia digunakan secara luas untuk komunikasi dan informasi cepat. Pengguna memanfaatkan *platform* ini untuk menyampaikan ide kepada pemimpin dan mempromosikan peristiwa serta budaya. Bahasa ofensif di Twitter(X) bisa mempengaruhi opini publik terhadap kinerja pemimpin (Kahi et al., 2024). Twitter(X) juga menyediakan data yang sangat berguna dalam bentuk tweet yang dapat dianalisis untuk berbagai tujuan, termasuk analisis konten Analisis sentimen merupakan proses otomatis yang digunakan untuk memahami dan mengolah data tekstual guna memperoleh informasi (Hasri & Alita, 2022), terutama untuk menentukan apakah pola pikir pengguna positif atau negatif terhadap topik yang sedang dibahas.

Salah satu topik yang menarik perhatian di media sosial baru-baru ini adalah Program Makan Siang Gratis, yang diprakarsai oleh pasangan calon presiden Prabowo Subianto dan Gibran Rakabuming Raka. Tujuan dari program ini adalah untuk menyediakan makanan siang gratis bagi anak-anak Indonesia termasuk anak dalam kandungan dan anak usia sekolah sejak bayi hingga dewasa (Sitanggang et al., 2024), yang merupakan bagian dari kampanye mereka untuk meningkatkan kohesi sosial dan membentuk komite-komite untuk menangani isu-isu yang berkaitan dengan kohesi publik. Tanggapan masyarakat terhadap program ini sangat luas dan sering diungkapkan melalui situs media sosial seperti Twitter(X).

Analisis sentimen merupakan Teknik dalam pengolahan bahasa alami untuk mengetahui opini terhadap suatu topik tertentu berdasarkan teks. Analisis sentimen

digunakan untuk memperoleh sentimen dari setiap kalimat, yang dilakukan untuk mengetahui apa pendapat masyarakat dalam sebuah topik yang bersangkutan positif atau negatif (Astari et al., 2020). Analisis sentiment menjadi sangat penting karena banyak informasi dapat diperoleh dari berbagai platform media social, ulasan produk, komentar pengguna dan lainnya. Diperlukan pengolahan data besar dengan menggunakan Teknik otomatis seperti Naïve Bayes.

*Multinomial Naive Bayes* merupakan metode *Text Mining* dalam proses pengklasifikasian teks menggunakan probabilitas kelas pada dokumen (Hadaina & Budiyanto, 2022). *Multinomial Naive Bayes* mengidentifikasi kategori dokumen tidak hanya berdasarkan kata-kata yang muncul di dalamnya, namun juga berdasarkan berapa kali kemunculannya (Wardani et al., 2020).

Terdapat beberapa penelitian sebelumnya yang membahas analisis sentiment menggunakan metode Naive Bayes. Penelitian (Astari et al., 2020) menggunakan analisis sentiment untuk mengetahui dampak virus Corona melalui dokumen Twitter. Penelitian (Rilinka et al., 2021) menggunakan analisis sentiment untuk menilai komentar suatu topik pada media sosial Twitter(X). Data dibagi menjadi dua bagian yaitu positif atau negatif untuk melihat sentimen dari topik mengenai penghapusan Ujian Nasional. Penelitian (Mujahidin et al., 2022) menggunakan analisis sentiment untuk sentimen mengenai Sirkuit Internasional Mandalika menggunakan metode *Multinomial Naive Bayes*.

Berdasarkan uraian di atas Penelitian ini bertujuan untuk menganalisis sentimen dan mengidentifikasi respon publik atau masyarakat terhadap program makan siang gratis berdasarkan klasifikasi teks menggunakan metode Naive Bayes. Hasil penelitian diharapkan dapat menyediakan wawasan bagi pihak terkait mengenai seberapa baik atau tidaknya program ini diterima publik atau masyarakat.

## II. TINJAUAN PUSTAKA

### 2.1. Analisis Sentimen

Analisis sentiment merupakan proses menghasilkan informasi secara otomatis dengan memahami dan mengolah kata (Hasri & Alita, 2022). analisis sentiment memungkinkan pengungkapan opini publik terhadap suatu isu dan menjadi cara untuk mengumpulkan pendapat umum melalui jejaring sosial yang sering membahas pelayanan publik dan isu-isu terkini (Syarifuddin, 2020).

### 2.2. Text Mining

*Text mining* adalah penemuan pengetahuan dalam database tekstual (disingkat *Knowledge Discovery in Textual Database*, atau KDT), juga dikenal sebagai ekstraksi atau pengambilan data tekstual yang diminati. Pengetahuan data yang baru dibuat didefinisikan sebagai bagian dari proses mengekstraksi atau mengambil data tekstual yang sebelumnya tidak diketahui untuk dipahami dari kumpulan data tekstual atau korpus (pengumpulan data) yang besar. Teks tidak terstruktur yang secara lengkap dan tegas menangkap penggunaan bahasa, tertulis atau lisan (Yuyun et al., 2018).

### 2.3. TF-IDF

TF-IDF merupakan metode pembobotan kata yang menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF-IDF adalah proses untuk mengonversi teks menjadi representasi vektor, yang bertujuan untuk menilai pentingnya suatu kata dalam konteks dokumen. *Term Frequency* (TF) mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen, sedangkan *Inverse Document Frequency* (IDF) mengukur seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen. Pembobotan kata dihitung dengan mengalikan nilai TF dan IDF, dimana kata yang sering muncul akan

mendapatkan bobot rendah, sedangkan kata yang jarang muncul akan memiliki bobot tinggi (Mahendra et al., 2023).

#### 2.4. Multinomial Naive Bayes

*Multinomial Naive Bayes* adalah pendekatan yang bekerja dengan menghitung frekuensi setiap term dalam dokumen. Kategori dokumen dalam klasifikasi polinomial Naive Bayes tidak hanya didasarkan pada kata yang muncul di dalamnya, tetapi juga mempertimbangkan jumlah kemunculan kata terbanyak. Jadi dalam konteks *Multinomial Naive Bayes*, peran token sangatlah penting (Fibriyanti Arminda et al., 2023).

Algoritma ini bekerja berdasarkan konsep frekuensi kemunculan istilah, yaitu seberapa sering kata tertentu muncul dalam sebuah dokumen. Model ini menjelaskan dua fakta: apakah kata tersebut muncul dalam dokumen dan seberapa sering kata tersebut muncul dalam dokumen. Setiap kata dalam kelas bisa dihitung probabilitasnya untuk pelatihan, menggunakan persamaan 1 (Yuyun et al., 2018).

$$P(p|n) \propto P(p) \prod_{1 \leq k \leq n} P(t_k|p) \quad (1)$$

Dimana  $P(t_k|p)$ : probabilitas munculnya dokumen  $text(t_k)$ ,  $n$  adalah jumlah dokumen dan  $p$  adalah polaritas. Kemudian untuk menghitung polaritasnya atau dokumen yang mempunyai kemiripan menggunakan persamaan 2.

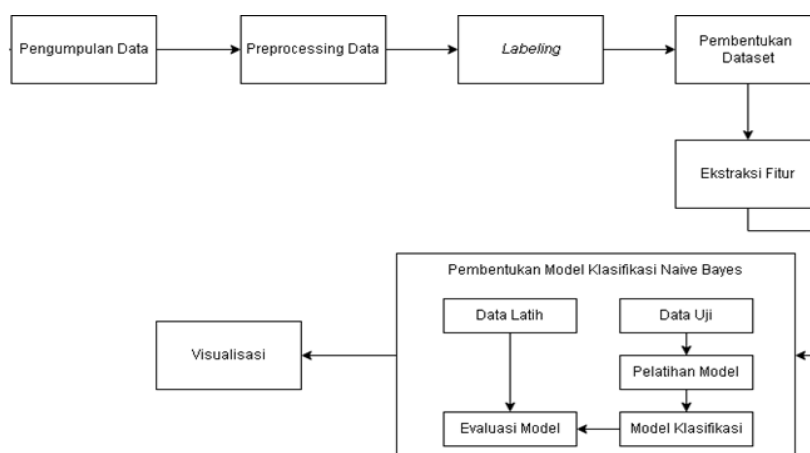
$$P(t_k|p) = \frac{\text{count}(t_k|p)+1}{\text{count}(tp)+|v|} \quad (2)$$

#### 2.5. Confusion matrix

*Confusion matrix* adalah alat pengukuran performa dalam masalah klasifikasi *machine learning* yang hasilnya bisa berupa dua kelas atau lebih. *Confusion matrix* menggambarkan tabel yang terdiri dari empat kombinasi berbeda antara nilai prediksi dan nilai aktual. Empat istilah yang digunakan untuk merepresentasikan hasil klasifikasi dalam *confusion matrix* adalah *True Positive*, *True Negative*, *False Positive*, dan *False Negative* (Hidayatullah et al., 2021).

### III. METODE PENELITIAN

Tahapan penelitian analisis sentimen opini pengguna Twitter(X) terhadap program makan siang gratis menggunakan metode Lexicon Based dan Multinomial Naive Bayes Classifier dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

## Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan dataset dengan menggunakan metode *Crawling Data* pada social media Twitter menggunakan tools Tweet-Harvest memanfaatkan Token API (Application Program Interface) dengan kata kunci Makan Siang Gratis mulai dari tanggal 17 Mei 2024 hingga 24 Mei 2024 berjumlah 2308 data dalam format .csv. Hasil *Crawling Data* dapat dilihat pada Gambar 2.

1	created_at	full_text	location	username
2	Thu May 23 23:30:19 +0000 2024	berharap dr program pemerintah yg baru makan siang gratis 7 millh lah yg jggetA biar slalu terhibur walau kenyataannya pahit oke gas dek yg smangat ya	new york	y4n_ydas
3	Thu May 23 23:29:03 +0000 2024	@DedynurPalakka @prabowo Koreksi makan siang di daki gratis		ordryngdman
4	Thu May 23 23:28:58 +0000 2024	@Boedianta4 Yadang jggetin aja kema makan siang gratis lih penting drpd pendidikan yakan sist @evaheni okegasokegas	Indonesia	pratiama92
5	Thu May 23 23:28:25 +0000 2024	@BosPurwa Mending studi banding ke mesjid jggetarjyan yg yakarta kalo utk makan siang gratis ngabizh duit aja kerja kelen.		EPPA3794
6	Thu May 23 23:28:25 +0000 2024	@szrirenova_ Kita cuma butuh sarapan soalnya makan siang gratis..... Janjinya	Makassar	stafuscolonial
7	Thu May 23 23:22:54 +0000 2024	@omni_siregar Lama-lama programnya ditambah. Ditambah sarapan dan makan siang gratis di rumah masing-masing	Malang	triefoweb88
8	Thu May 23 23:21:25 +0000 2024	@Lundipmefess Murah itu mah bagi @evaheni yakan sist kema pendidikan g penting uang negara mending buat makan siang gratis okegasokegas	Indonesia	pratiama92
9	Thu May 23 23:20:07 +0000 2024	@Iandikmalreza Ralat bukan makan siang gratis tp makanan bergizi untuk anak. Sdh revisi sebelum pelantikan.		FactSelfawan
10	Thu May 23 23:19:42 +0000 2024	@Gerindra kami blhas makan siang gratis di revisi asalkan biaya pendidikan di PTN menjadi lebih terjangkau		PalakkaSO
11	Thu May 23 23:18:54 +0000 2024	Tim Pakar Prabowo dan Gibran mengkalulasi biaya program prioritas makan siang gratis bagi anak sekolah balita dan wanita hamil untuk tahun pertama. Nilainya men. DKI Jakarta, Indonesia		zomett3
12	Thu May 23 23:18:38 +0000 2024	Negeri Kapitalis		
13	Thu May 23 23:15:14 +0000 2024	@DedynurPalakka @prabowo Ya elah dikoreksi kayak ujian tengah semester. Pa visi misi sebelumnya nggak dikoreksi dulu ya ? Yg tertanam di benak rakyat tetap MAKAN SIANG GRATIS		abehafulid2
14	Thu May 23 23:13:06 +0000 2024	@vivoNews Tadnya program makan siang gratis dikoreksi makan siang bergizi untuk anak-anak yang sudah dewasa dan berumur hanya g'gih jari... Tunggu saja nanti ak Cirebon, Jawa Barat		heryscomebody
15	Thu May 23 23:12:35 +0000 2024	Maaf ya kok makan siang gratisnya diganti jadi makan pagi gratis dan mungkin nanti diganti lagi jadi makan malam gratis serta akhirnya cuma cemilan gratis... -Salam KR Jakarta		astpnhair
16	Thu May 23 23:10:43 +0000 2024	@Gerindra Gas makan siang gratis! Solusi semua masalah di negeri ini! sv https://t.co/7TuaQPIAN	Jakarta, Indonesia	aresissha
17	Thu May 23 23:08:42 +0000 2024	@femykoto Sudahlah! Makan siang gratis lebih penting daripada masalah UKTI sv https://t.co/vmsITZfHm	Jakarta, Indonesia	aresissha
18	Thu May 23 23:07:59 +0000 2024	@Wimsaja @msaid_didu Lho ngga akan laper dong kan raiyat dikasih makan siang gratis	Aceh: -6.883714,107.608037	eritihesa
19	Thu May 23 23:06:49 +0000 2024	@Partasocmed Makan siang gratis lebih penting daripada study tour! sv https://t.co/kyoUqvDQUB	Jakarta, Indonesia	aresissha
20	Thu May 23 23:06:44 +0000 2024	@yanlansim semua akan 'di perbaiki' dan bisa jadi 'di eliminasi nantinya... makan siang gratis telah di perbaiki menjadi 'makan siang bergizi bagi anak' indahnya mempe Jakarta Capital Region, Indon! DemagaGarely		

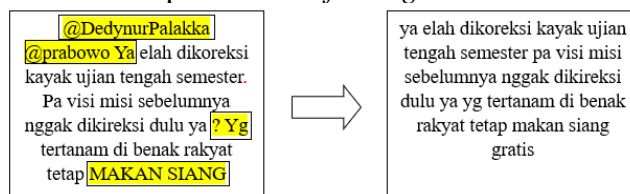
Gambar 2. Hasil *Crawling Data*

## Pre-Processing

Tahap Pre-processing data dilakukan untuk mengubah data *tweet* yang tidak terstruktur menjadi *tweet* yang lebih terstruktur sesuai dengan kebutuhan prosedur analisis sentiment. Tahapan pre-processing terdiri dari beberapa proses, yaitu *data cleaning*, *case folding*, *normalize*, *tokenize*, *stopword removal*, dan *stemming*.

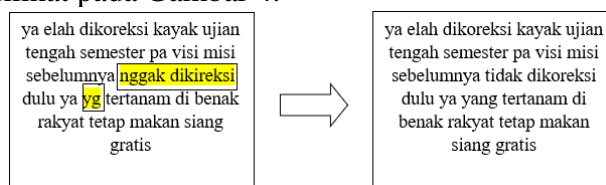
Tahapan pertama *data cleaning* dimana proses yang dilakukan dengan menghilangkan informasi yang tidak dibutuhkan seperti iklan, salah eja, tanda baca, karakter dan simbol yang tidak diperlukan, URL dengan fungsi `re.sub('simbol',text)` dan `remove`. Proses *data cleaning* dengan menggunakan filtering yaitu untuk menghilangkan beberapa elemen yang tidak diinginkan dan menghapus *duplicate tweet*.

Tahapan kedua, *case folding* adalah tahapan pengubahan huruf besar (*capital casing*) di dalam dataset menjadi huruf huruf kecil (*lower casing*). *Case folding* dijalankan dengan memeriksa keseluruhan ukuran karakter dan huruf, lalu jika ditemukan karakter yang menggunakan huruf kapital maka, maka huruf tersebut akan diubah menjadi *lower casing*. Gambar 3 menunjukkan ilustrasi proses *Case folding*.



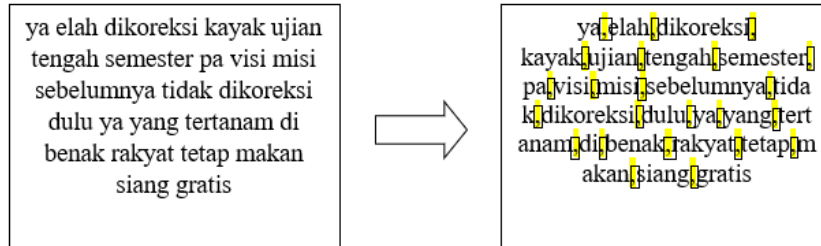
Gambar 3. Ilustrasi Proses *Case Folding*

Tahap ketiga, *normalize* dilakukan dengan tujuan untuk mengurangi data ke skala yang lebih konsisten dan terstandarisasi. Proses *normalize* mengubah kata-kata tidak baku menjadi kata-kata baku berdasarkan sebuah kamus. Ilustrasi kata-kata yang sudah dinormalisasikan bisa dilihat pada Gambar 4.



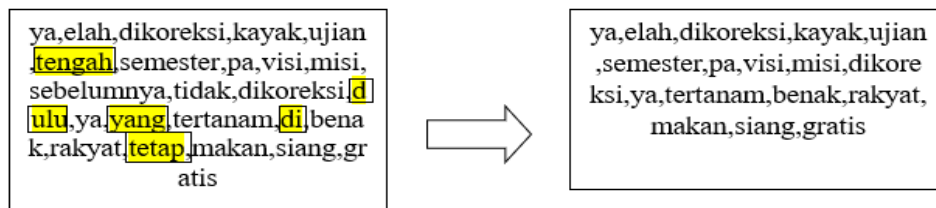
Gambar 4. Ilustrasi Proses Normalisasi

Tahapan keempat, *tokenize* dimana tahap ini memecahkan kalimat menjadi kata-perkata atau disebut dengan proses pemenggalan kata. Token ini bisa berupa kata, frasa, kalimat, atau entitas lain yang memiliki makna. Proses ini penting karena banyak metode analisis teks dan pemrosesan lebih lanjut bergantung pada representasi token ini. Ilustrasi ditunjukkan koma yang diberi kotak, Ilustrasi dari proses *tokenized* dapat dilihat pada Gambar 5.



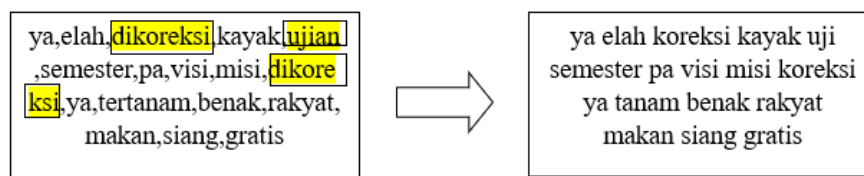
Gambar 5. Ilustrasi Proses Tokenized

Tahap kelima, *stopwords removal* merupakan proses dimana setiap kata-kata akan dianalisa dan diperiksa untuk kata-kata yang tidak relevan akan dihilangkan. Proses ini dilakukan untuk meningkatkan meningkatkan akurasi analisis. Ilustrasi Proses *Stopwords Removal* dapat dilihat pada Gambar 6.



Gambar 6. Ilustrasi Proses *Stopwords Removal*

Tahap keenam, *stemming* adalah tahap menyederhanakan kata menjadi kata yang sesuai dengan kamus atau kata dasar. *Stemming* dilakukan untuk mengenali setiap kata yang dimasukkan, proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya (*conflation*). Proses dimulai dengan memeriksa setiap kata untuk mengetahui apakah memiliki imbuhan atau tidak seperti dapat dilihat pada Gambar 7. Kata tanpa imbuhan dianggap sebagai kata dasar, dan proses diakhiri dengan hasil berupa kata dasar dari data tweet.



Gambar 7. Ilustrasi Proses *Stemming*

### Labeling

Tahapan selanjutnya setelah *pre-processing* adalah proses labeling dimana data diklasifikasi ke dalam sentimen positif dan negatif. Proses labeling menggunakan metode *lexicon based*, data di beri label berdasarkan kamus kata opini (*lexicon*) dari penelitian sebelumnya yaitu InSet Lexicon (Koto & Rahmanningtyas, 2017).

## Pembentukan Dataset

Tahap pembentukan dataset ini data yang telah diberi label positif dan negatif dilatih dengan *Multinomial Naive Bayes*, data dibuat menjadi data *training* dan data *test*. Skenario jumlah data *test* diambil 30%, 20%, dan 10% dari data asli.

## Ekstraksi Fitur

Setelah data di *Splitting* data *training* dan data *test* diubah menjadi numerik dengan pembobotan TF-IDF. Penggunaan TF-IDF dapat membantu mengidentifikasi kata-kata penting untuk sentiment tertentu, memungkinkan model pembelajaran mesin lebih fokus pada informasi yang relevan dan meningkatkan kinerja dalam klasifikasi sentiment(Wati et al., 2023).

## Pembentukan Model Klasifikasi dan Evaluasi

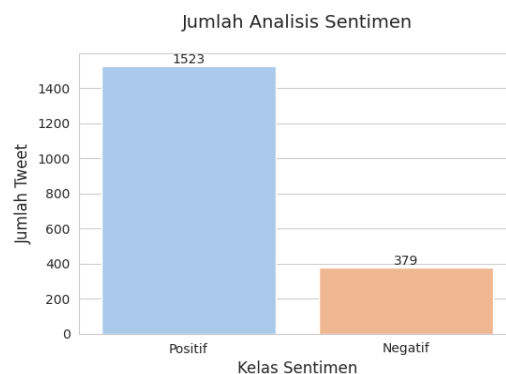
Klasifikasi yang digunakan dalam penelitian adalah Naïve Bayes. Pengklasifikasian dengan *Naive Bayes* melibatkan penerapan *Teorema Bayes* dengan asumsi independensi antar fitur untuk menentukan probabilitas suatu data termasuk dalam suatu kelas tertentu. *Naive Bayes* adalah metode probabilistik yang digunakan untuk memprediksi kelas dari data baru berdasarkan distribusi probabilitas yang telah dipelajari selama pelatihan.

Evaluasi Akurasi mengukur persentase prediksi yang benar dari total prediksi yang dibuat oleh model. Perhitungan tingkat akurasi dilakukan dengan menggunakan metode *Confusion Matrix* dimana seberapa baik klasifikasi yang telah dibuat, dapat mengenali *tuple* dari kelas yang berbeda.

## IV. HASIL DAN PEMBAHASAN

Analisis sentimen pada penelitian ini menggunakan kamus kata berisi kata-kata yang memiliki nilai sentimen tertentu dengan memberikan label dengan skor sentimen. Penggunaan metode lexicon akan menghitung skor sentimen keseluruhan suatu teks berdasarkan kata-kata yang ada dalam teks tersebut dan membandingkannya dengan nilai yang ada di kamus. Proses ini menggunakan kamus positif dan negatif InSet Lexicon.

Hasil proses *data cleaning* didapat 2275 dari 2308 kemudian dilakukan proses *labeling* menggunakan metode *lexicon* mendapatkan 1902 data dikarenakan data netral diubah menjadi *none* yang berarti data netral dihilangkan. Gambar 8 memperlihatkan hasil proses *labeling* dengan label sentimen positif 1523 dan 379 sentimen negatif.



Gambar 1. Diagram Hasil *Labeling*

Pembentukan dataset dilakukan dengan membagi data menjadi dua bagian data latih dan data uji. Skenario pembentukan data dibuat tiga kali proses *splitting* untuk mendapatkan hasil akurasi yang terbaik. Tabel 1 memperlihatkan skenario hasil dari

pembentukan dataset dimana 70% data latih dan 30% data uji menghasilkan 1331 data latih dan 571 data uji, untuk 80% data latih dan 20% data uji menghasilkan 1521 data latih dan 381 data uji, kemudian dengan 90% data latih dan 10% data uji menghasilkan 1711 data latih dan 191 data uji.

Ekstraksi Fitur yang dilakukan pada penelitian ini menggunakan TF-IDF dan *CountVectorizer*. Pada proses TF-IDF dilakukan untuk menentukan seberapa sering kata muncul di dalam dokumen, lalu *CountVectorizer* digunakan untuk menghitung frekuensi kemunculan kata pada data. Berdasarkan Tabel 1 hasil pengujian dengan data uji 30% mendapatkan 3361 fitur (kata-kata unik), data uji 20% mendapatkan 3637 fitur (kata-kata unik), dan 10% mendapatkan 3955 fitur (kata-kata unik). Hasil pembentukan dataset dan pembobotan TF-IDF dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pembentukan Dataset dan Pembobotan TF-IDF

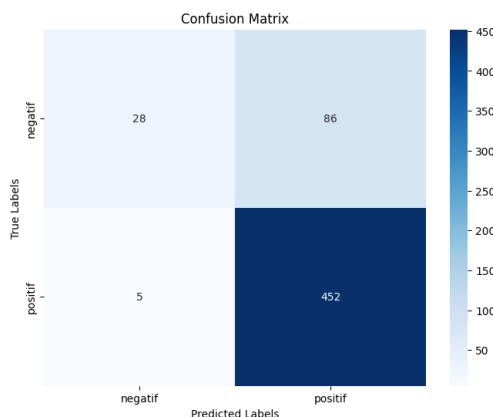
Data Latih(%)	Data Uji(%)	Hasil Data Latih	Hasil Data Uji	Hasil TF-IDF
70	30	1331	571	3361
80	20	1521	381	3637
90	10	1711	191	3955

Hasil keakuratan dari pelatihan *Multinomial Naive Bayes* didapat dari perhitungan *Confusion matrix*. Setiap data yang diuji melalui perhitungan *confusion matrix* untuk dihitung dan dari ke tiga data tersebut manakah yang menghasilkan akurasi yang paling tinggi. Tabel 2 memperlihatkan hasil evaluasi model menggunakan *Confusion matrix* pada 3 skenario. Penggunaan *lexicon* dapat mempengaruhi atau mendukung kinerja *Naive bayes* berdasarkan probabilitas kata-kata yang muncul pada teks, hal ini dapat dilihat pada Tabel 2 bahwa hasil nilai akurasi tertinggi pada data latih 70% dan data uji 30% sebesar 84,06%

Tabel 2. Hasil Evaluasi *Confusion matrix*

Data Latih (%)	Data Uji(%)	Accuracy(%)
70	30	84,06
80	20	82
90	10	81

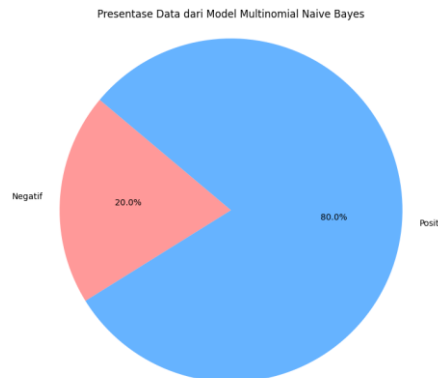
Evaluasi model menggunakan *Confusion matrix* pada pembentukan dataset untuk data latih 70% dan data uji 30%. Hasil evaluasi model dapat dilihat pada Gambar 9. Matrik menunjukkan bahwa model klasifikasi berhasil memprediksi 28 tweet negatif dengan benar (TN) dan 452 tweet positif dengan benar (TP). Namun, terdapat 86 kesalahan di mana tweet negatif diprediksi sebagai positif (FP) dan 5 kesalahan di mana tweet positif diprediksi sebagai negatif (FN).



Gambar 9. *Confusion Matrix* Data Uji 30%

Berdasarkan nilai yang didapat dari hasil evaluasi menggunakan *Confusion matrix* seperti yang terlihat pada gambar 9 didapatkan nilai *precision*, *recall* dan *F1-score*. Hasil *precision* sebesar 83,9% yang berarti model cukup baik dalam menghindari kesalahan prediksi positif (*false positif*). Hasil *recall* sebesar 98,9 % yang berarti model mampu mendeteksi hampir semua tweet positif dengan benar. Hasil *F1-score* sebesar 90,7% yang menunjukkan model sangat baik dalam menangani sentiment positif dengan keseimbangan yang baik antara *precision* dan *recall*.

Persentase data positif dan negatif dari akurasi *Multinomial Naive Bayes* berdasarkan perhitungan *Confusion Matrix* pada 571 data uji (30%), perhitungan tersebut mendapatkan hasil persentase untuk data negatif 20% dan data positif 80%, seperti dapat dilihat pada Gambar 10.



Gambar 10. Tampilan Diagram Pie Persentase dari Model *Multinomial Naive Bayes*

Secara keseluruhan, berdasarkan hasil analisis sentimen menggunakan metode Naive Bayes menunjukkan kinerja yang baik dalam menganalisis sentimen program makan siang gratis dengan akurasi sebesar 84,06%. Hal ini menunjukkan opini masyarakat cenderung positif.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil penelitian, analisis sentiment program makan siang gratis menggunakan metode *Multinomial Naive Bayes* didapat hasil akurasi 84,06%, *precision* 83,9%, *recall* 98,9%, dan *F1-score* 90,70% yang dihitung menggunakan *confusion matrix*. Hal ini menunjukkan bahwa mayoritas respon masyarakat cenderung memberikan positif dengan kata lain masyarakat mendukung dan optimis terkait program kerja makan siang gratis yang dapat langsung memberikan manfaat sosial kepada masyarakat dalam pemenuhan kebutuhan dasar.

Penggunaan teknik seperti Naive Bayes dalam pengklasifikasian sentimen dapat membantu pembuat kebijakan untuk mengetahui seberapa besar dukungan atau penolakan terhadap program tersebut berdasarkan opini yang ada di media sosial.

### 5.2 Saran

Pengembangan lebih lanjut terkait penelitian ini dapat dilakukan dengan meningkatkan kegunaan *lexicon based*, memasukan dan memperbanyak kosakata dan frasa yang relevan untuk topik yang ingin di analisis dan menggunakan metode ekstraksi fitur lain seperti, word embeddings (misalnya, Word2Vec, GloVe) atau transformer-based embeddings (misalnya, BERT). Penggunaan teknik *oversampling* atau *undersampling* untuk meningkatkan kinerja model dalam mendeteksi tweet negatif

## DAFTAR PUSTAKA

- Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- Fibriyanti Arminda, N., Sulistiyowati, N., & Nur Padilah, T. (2023). Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1817–1822. <https://doi.org/10.36040/jati.v7i3.7012>
- Hadaina, F., & Budiyanto, U. (2022). Implementasi Metode Multinomial Naive Bayes Untuk Sentiment Analysis Terhadap Data Ulasan Produk Colearn Pada Google Play Store. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia, September*, 660–666. <https://senafiti.budiluhur.ac.id/index.php>
- Hasri, C. F., & Alita, D. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 3(2), 145–160. <https://doi.org/10.33365/jatika.v3i2.2026>
- Hidayatullah, M., Alam, S., & Jaelani, I. (2021). Sentiment Analysis of Police Performance On Twitter Users Using Naïve Bayes Method. *RISTEC: Research in Information Systems and Technology*, 2(2), 86–97. <https://doi.org/10.31980/ristec.v2i2.1945>
- Kahi, F. R. B., Talakua, A., & Reynaldi, R. (2024). Analisis Sentimen Masyarakat Di Twitter Terhadap Pemerintahan Anies Baswedan Menggunakan Metode Naive Bayes Classifier. *Jurnal Minfo Polgan*, 13(1), 324–336. <https://doi.org/10.33395/jmp.v13i1.13636>
- Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017, 2018-Janua(December)*, 391–394. <https://doi.org/10.1109/IALP.2017.8300625>
- Mahendra, M. H., Murdiansyah, D. T., & Lhaksmana, K. M. (2023). Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer. *DIKE: Jurnal Ilmu Multidisiplin*, 1(2), 37–43. <https://doi.org/10.69688/dike.v1i2.35>
- Mujahidin, S., Hasyim, M. N., & Pratama, B. M. (2022). Implementasi Analisis Sentimen Opini Publik Mengenai Sirkuit Internasional Mandalika Pada Twitter Menggunakan Metode Multinomial Naive Bayes Classifier. *Bianglala Informatika*, 10(2), 129–136. <https://doi.org/10.31294/bi.v10i2.13544>
- Rilinka, R., Indriati, I., & Yudistira, N. (2021). Analisis Sentimen Penghapusan Ujian Nasional pada Twitter menggunakan Document Frequency Difference dan Multinomial Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu ...*, 5(3), 876–883. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/8659/3981>
- Sitanggang, A., Umaidah, Y., Umaidah, Y., Adam, R. I., & Adam, R. I. (2024). Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4902>
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>

- Wardani, N. S., Prahutama, A., & Kartikasari, P. (2020). *Analisis Sentimen Pemindehan Ibu Kota Negara dengan Klasifikasi Naïve Bayes untuk Model Bernoulli dan Multinomial*. 9, 237–246.
- Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), 84–93. <https://doi.org/10.34010/jamika.v13i1.9424>
- Yuyun, Hidayah, N., & Sahibu, S. (2018). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Resti*, 1(1), 19–25.