

Implementasi Algoritma *Naive Bayes* untuk Analisis Sentimen Terhadap Program Makan Siang Gratis

Tundo¹, Dea Noer Rachmawati^{2*}

¹ Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

² Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Corresponding Email: asna8mujahid@gmail.com¹, dearachma21@gmail.com^{2*}

Histori Artikel:

Dikirim 22 Juli 2024; Diterima dalam bentuk revisi 2 Agustus 2024; Diterima 15 Agustus 2024; Diterbitkan 20 September 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Program kerja yang dijanjikan oleh pasangan Calon Presiden dan Calon Wakil Presiden 2024 yaitu, Prabowo Subianto dan Gibran Rakabuming Raka, salah satunya adalah program makan siang gratis, program ini menjadi upaya dalam meningkatkan kesejahteraan masyarakat, namun hal ini menjadi perhatian publik pada sosial media, salah satunya adalah platform X. Respon publik terhadap program makan siang gratis menjadi sorotan utama. Dalam penelitian ini, peneliti menganalisis sentimen terkait program makan siang gratis dengan metode Naive Bayes menggunakan bahasa pemrograman python pada Google Colab untuk menganalisis sentimen terhadap pengguna media sosial X. hasil yang didapatkan dari 920 data tweet, terdapat 167 tweet bernilai positif, 744 tweet bernilai negatif. Hasil evaluasi dengan confusion matrix didapatkan akurasi sebesar 86.95%, dengan precision 93%, recall 61%, dan F1-Score 65%. Dari hasil penelitian ini dapat disimpulkan bahwa sebagian besar publik yang berkomentar pada sosial media X memberikan respon negatif terhadap program makan siang gratis. Hasil ini dapat menjadi bahan evaluasi pemerintah dalam menentukan kebijakan yang efektif dan efisien bagi masyarakat.

Kata Kunci: Program Makan Siang Gratis; Sentimen Analisis; Naive Bayes; X; Rapidminer.

Abstract

The work program promised by the 2024 Presidential Candidate and Vice Presidential Candidate pair, namely, Prabowo Subianto and Gibran Rakabuming Raka, one of which is a free lunch program, this program is an effort to improve community welfare, but this has attracted public attention on social media, one the other is platform X. The public response to the free lunch program is the main focus. In this research, researchers analyzed sentiment related to the free lunch program using the Naive Bayes method using the Python programming language on Google Colab to analyze sentiment towards social media users. X. the results obtained from 920 tweets data, there are 167 positive value tweets, 744 tweets are negative. The evaluation results with the confusion matrix showed an accuracy of 86.95%, with precision 93%, recall 61%, and F1-Score 65%. From the results of this research it can be concluded that the majority of the public who commented on social media X gave a negative response to the free lunch program. These results can be used as material for government evaluation in determining effective and efficient policies for society.

Keyword: Free Lunch Program; Sentiment Analysis; Naive Bayes; X; Rapidminer.

1. Pendahuluan

Visi dan misi adalah elemen penting dalam pencalonan presiden dan wakil presiden. Pasangan calon presiden dan wakil presiden di Indonesia biasanya mencantumkan visi dan misi mereka dalam dokumen rencana pembangunan jangka panjang nasional (RPJPN). Penyusunan visi dan misi ini diatur dalam Pasal 10 ayat (1) Peraturan Panitia Pemilihan Umum (PKPU) Nomor 19 Tahun 2023, yang menyatakan bahwa visi, misi, serta rencana kerja calon presiden dan wakil presiden harus disusun sesuai dengan ketentuan yang berlaku (Aspa & Muchram, 2024). Salah satu program yang diusung oleh pasangan calon Prabowo Subianto dan Gibran Rakabuming Raka dalam Pemilihan Presiden 2024 adalah Program Makan Siang Gratis. Program ini mendapat perhatian luas karena menasarkan kelompok rentan seperti siswa sekolah, santri, ibu hamil, dan balita. Tujuan dari program ini adalah untuk mengatasi kemiskinan, *stunting*, serta masalah gizi buruk pada anak-anak di Indonesia. Namun, keberhasilan program pemerintah tidak hanya bergantung pada efisiensi distribusi bantuan atau pengelolaan anggaran. Persepsi masyarakat terhadap program tersebut sangat berperan dalam keberlanjutan dan kesuksesannya di masa depan (Puad *et al.*, 2024). Pandangan dan penilaian masyarakat dapat memengaruhi implementasi kebijakan, di mana kebijakan yang mendapat dukungan publik cenderung lebih mudah diterima dan dilaksanakan. Sebaliknya, kebijakan yang menuai reaksi negatif dari masyarakat akan menghadapi berbagai tantangan dalam penerapannya.

Di era digital ini, opini masyarakat sering kali diekspresikan melalui media sosial, salah satunya adalah *Platform X*. Platform ini menjadi wadah utama bagi masyarakat untuk menyampaikan pandangan, komentar, dan kritik terhadap isu-isu yang sedang berkembang. Kebijakan Program Makan Siang Gratis, misalnya, telah mendapat sorotan besar di *Platform X*. Banyaknya opini yang beredar di media sosial mengenai program ini menunjukkan pentingnya analisis terhadap respons publik. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Program Makan Siang Gratis menggunakan metode analisis sentimen berbasis algoritma *Naive Bayes*. Analisis sentimen merupakan teknik yang digunakan untuk mengukur dan mengkategorikan opini masyarakat terhadap isu tertentu, yang umumnya dibagi menjadi dua kategori, yaitu sentimen positif dan negatif (Verena *et al.*, 2021). Metode ini memungkinkan peneliti untuk mengidentifikasi dan mengklasifikasikan opini publik secara otomatis berdasarkan teks yang dihasilkan oleh pengguna media sosial.

Algoritma *Naive Bayes* dipilih sebagai metode utama dalam analisis ini karena kemampuannya untuk menangani data teks dalam jumlah besar dengan cepat dan akurat. *Naive Bayes* adalah algoritma berbasis probabilitas yang menggunakan teorema *Bayes* untuk memprediksi kelas dari suatu data berdasarkan pola yang ada pada data sebelumnya (Azhar *et al.*, 2022). Dalam penelitian ini, algoritma *Naive Bayes* akan digunakan untuk memodelkan klasifikasi sentimen terkait dengan Program Makan Siang Gratis, dengan data yang diambil dari *tweet* di *Platform X*. Algoritma ini telah banyak digunakan dalam penelitian-penelitian serupa untuk menganalisis sentimen di berbagai bidang, termasuk kebijakan publik, kampanye politik, dan respons masyarakat terhadap layanan pemerintah. Data yang dianalisis dalam penelitian ini mencakup *tweet* yang mengandung kata kunci seperti "Program Makan Siang Gratis" dan "Makan Siang Gratis" yang diambil dari *Platform X*. Data tersebut kemudian dikategorikan menjadi dua sentimen utama: positif dan negatif. Sentimen positif mencerminkan dukungan masyarakat terhadap program tersebut, sedangkan sentimen negatif menunjukkan ketidaksetujuan atau kritik terhadap pelaksanaannya (Syarifuddin, 2020). Proses ini memberikan gambaran yang jelas tentang bagaimana masyarakat merespons kebijakan yang diusulkan oleh pemerintah.

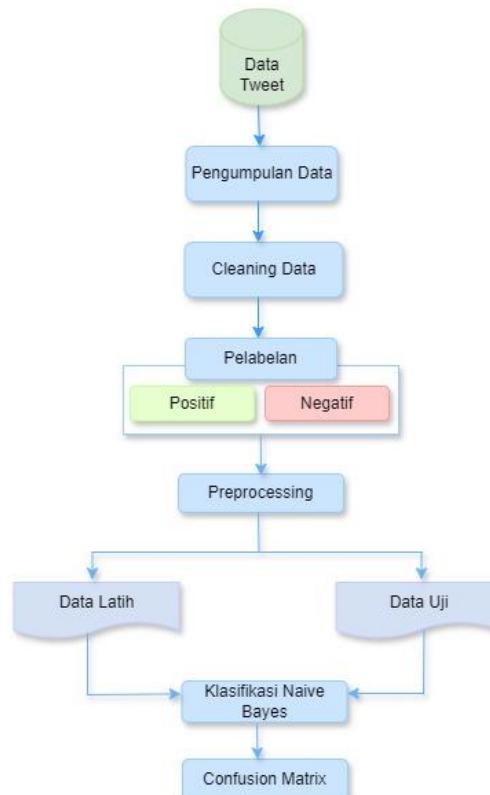
Analisis sentimen berbasis *Naive Bayes* tidak hanya membantu mengukur dukungan atau penolakan terhadap kebijakan, tetapi juga dapat memberikan saran yang berguna bagi pemerintah dalam merancang kebijakan yang lebih efektif dan efisien (Wijaya, 2021). Hasil dari analisis ini dapat menjadi masukan berharga bagi pemerintah dalam mengevaluasi Program Makan Siang Gratis, terutama terkait aspek-aspek yang mungkin belum sesuai dengan harapan masyarakat. Jika program ini lebih banyak mendapatkan sentimen negatif, pemerintah dapat mempertimbangkan langkah-langkah korektif untuk memperbaiki pelaksanaan atau komunikasi kebijakan. Sebaliknya, jika

majoritas sentimen positif, program ini dapat dilanjutkan dengan beberapa penyesuaian untuk mencapai hasil yang lebih optimal. Tujuan utama penelitian ini adalah untuk mengetahui bagaimana masyarakat memandang dan menanggapi Program Makan Siang Gratis yang diusulkan oleh pasangan calon presiden dan wakil presiden Prabowo Subianto dan Gibran Rakabuming Raka. Dengan menganalisis sentimen publik melalui data yang diambil dari media sosial, pemerintah dapat memperoleh informasi yang bermanfaat terkait penerimaan masyarakat terhadap program ini. Selain itu, hasil penelitian ini diharapkan dapat memberikan rekomendasi kepada pemerintah dalam menyusun kebijakan yang lebih responsif terhadap kebutuhan dan keinginan masyarakat. Hal ini sangat penting untuk memastikan bahwa kebijakan yang diterapkan benar-benar efektif dan mampu memberikan manfaat optimal bagi seluruh lapisan masyarakat.

Secara keseluruhan, penelitian ini diharapkan dapat memberikan kontribusi bagi pemahaman yang lebih baik tentang bagaimana opini masyarakat terhadap kebijakan pemerintah dapat diukur secara sistematis melalui analisis data media sosial. Selain itu, penelitian ini juga bertujuan untuk memperkuat literatur mengenai penggunaan algoritma *Naive Bayes* dalam analisis sentimen kebijakan publik, serta bagaimana hasil dari analisis ini dapat digunakan sebagai bahan evaluasi untuk perbaikan kebijakan di masa mendatang.

2. Metode Penelitian

Pada penelitian ini beberapa langkah penelitian yang dillakukan untuk menganalisis sentimen masyarakat terhadap program makan siang gratis menggunakan metode *Naive Bayes*. Tahapan penelitian tersebut seperti gambar 1 sebagai berikut:



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan tujuan untuk mengumpulkan informasi atau fakta yang relevan guna menjawab pertanyaan penelitian. Data yang digunakan berasal dari *Platform X*, yang diperoleh melalui proses *crawling* dengan menggunakan bahasa pemrograman Python. Proses pengumpulan data dimulai dengan mencari tweet yang mengandung kata kunci seperti "Program Makan Siang Gratis" dan "Makan Siang Gratis". Data yang tidak relevan atau tidak dibutuhkan, seperti tautan, retweet, atau mention, dihilangkan agar hanya tersisa data yang relevan untuk proses penelitian lebih lanjut. Tujuan utama dari pengumpulan data adalah untuk memperoleh data yang akurat dan relevan, sehingga hasil analisis yang dihasilkan dapat mencerminkan pandangan masyarakat secara lebih objektif (Azhar *et al.*, 2022). Setelah data dikumpulkan, langkah berikutnya adalah *cleaning data*. Pembersihan data ini merupakan proses penting dalam analisis teks, di mana karakter atau teks yang tidak relevan dihapus untuk meningkatkan kualitas data yang akan dianalisis. Langkah ini mencakup penghapusan simbol, tanda baca, angka, dan karakter lainnya yang tidak diperlukan dalam analisis sentimen. Data yang tidak bersih atau mengandung banyak elemen yang tidak relevan dapat mengganggu akurasi model dalam memproses teks, sehingga proses pembersihan ini menjadi esensial untuk menghasilkan hasil yang lebih baik (Syarifuddin, 2020). Setelah pembersihan, data kemudian diberi label sesuai dengan kelas sentimen. Proses pelabelan ini adalah pengkategorian data ke dalam dua kelas utama, yaitu sentimen positif dan negatif. Sentimen positif meliputi tweet yang mendukung Program Makan Siang Gratis, sedangkan sentimen negatif terdiri dari tweet yang berisi kritik atau ketidakpuasan terhadap program tersebut. Pelabelan data ini dilakukan secara manual untuk memastikan setiap tweet dikategorikan dengan benar (Wijaya, 2021).

Tahap selanjutnya adalah *preprocessing* data, di mana data mentah yang telah diberi label diproses lebih lanjut agar siap digunakan dalam model pembelajaran mesin. Proses ini melibatkan beberapa langkah seperti transformasi teks menjadi huruf kecil, normalisasi kata, penghapusan kata-kata yang tidak penting (*stopwords*), tokenisasi, dan *stemming*. Setiap langkah bertujuan untuk mengubah data yang tidak terstruktur menjadi lebih terstruktur dan mudah dipahami oleh model pembelajaran mesin. Proses ini juga bertujuan untuk meningkatkan akurasi dan performa model dalam mengklasifikasi data (Nurmala *et al.*, 2020). Setelah data diproses, algoritma *Naive Bayes* digunakan untuk memprediksi probabilitas dari masing-masing kelas sentimen. Algoritma ini bekerja berdasarkan prinsip probabilitas, di mana data dianggap saling independen satu sama lain. Data yang sudah melalui proses *preprocessing* dibagi menjadi dua bagian, yaitu dataset latih dan dataset uji. Dataset latih digunakan untuk melatih model, sementara dataset uji digunakan untuk mengukur performa model yang telah dilatih. Dalam penelitian ini, rumus dasar *Naive Bayes* digunakan untuk menghitung probabilitas sentimen positif atau negatif dari setiap tweet yang dianalisis (Atimi & Pratama, 2022). Model *Naive Bayes* dievaluasi menggunakan matriks kebingungan (*confusion matrix*) untuk mengukur akurasi, presisi, recall, dan *F1-Score*. Evaluasi ini penting untuk memahami seberapa baik model dalam mengklasifikasikan tweet ke dalam sentimen positif atau negatif (Priyanto *et al.*, 2024).

$$P(C|x) = \frac{P(C)xP(x|C)}{P(x)}$$

Keterangan :

$P(C|x)$ = adalah probabilitas posterior kelas C dengan diberikan data x.

$P(C)$ = adalah probabilitas prior dari kelas C.

$P(x|C)$ = adalah probabilitas likelihood data x pada kelas C.

$P(x)$ = adalah probabilitas dari data x secara keseluruhan

2.2 Evaluasi *Confusion Matrix*

Evaluasi model dilakukan dengan *Confusion Matrix* untuk mengevaluasi kinerja model *Naive Bayes*. *Confusion matrix* menggambarkan performa model klasifikasi dengan membandingkan prediksi yang dibuat pada suatu model dengan nilai sebenarnya dari data yang diamati. Representasi pada *confusion*

matrix dilakukan untuk menghitung hasil prediksi model terhadap data uji. *Confusion matrix* dapat digambarkan sebagai berikut:

Tabel 1. *Confusion Matrix*

	True Positif	True Negatif
Pred Positif	Tru Positif (TP)	False Negatif (FN)
Pred Negatif	False Positif (FP)	True Negatif (TN)

Tahap ini menghitung jumlah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), dan *False Negatif* (FN) dari model *Naive Bayes*. Setelah itu, dilakukan pencarian nilai akurasi, presisi, dan *recall*. Berikut rumus pada proses *confusion matrix*:

1) Akurasi

Akurasi digunakan untuk mengukur seberapa baik model yang digunakan dalam melakukanklasifikasi sentimen. Dimana semakin besar nilai akurasi dari metode tersebut maka semakin bagus kinerja metode tersebut. Adapun rumus untuk menghitung akurasi sebagai berikut:

$$\text{Accuracy} : \frac{TP+TN}{TP+TN+FP+FN}$$

2) Presision

Presision digunakan untuk mengukur seberapa baik model dalam memprediksi kelas positif dari total prediksi positif yang dilakukan. Adapun rumus untuk menghitung presision sebagai berikut:

$$\text{Precision} : \frac{TP}{TP+FP}$$

3) Recall

Recall digunakan untuk mengevaluasi matrix yang menggambarkan seberapa baik suatu model dalam mengidentifikasi kelas positif dengan benar. Dalam klasifikasi, *recall* menunjukkan kemampuan model yang digunakan dalam mengenali sentimen positif yang sebenarnya. Adapun rumus untuk menghitung recall sebagai berikut:

$$\text{Recall} : \frac{TP}{TP+FN}$$

4) F1-Score

F1-Score adalah matrik evalausi yang menggambarkan keseimbangan antara presisi (*Precision*) dan sensitivitas (*Recall*). Dalam klasifikasi *F1-Score* memberikan gambaran mengenai seberapa baik kita dalam mengklasifikasi baik komentar positif maupun negatif secara akurat. Nilai *F1-Score* akan memberikan informasi mengenai seberapa baik model yang di gunakan dalam menggabungkan kemampuan presisi dan sensitivitas, sehingga peneliti bisa memahami seberapa efektif model yang digunakan dalam mengklasifikasi opini pada twitter.

$$\text{F1-Score} : 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi seharusnya. Ini memungkinkan pengembang dalam mengidentifikasi model. *Confusion matrix* tidak hanya memberikan informasi tentang performa model, tetapi juga memberikan wawasan untuk meningkatkan kualitas dan akurasi model secara keseluruhan.

3. Hasil dan Pembahasan

3.1 Hasil

Pada penelitian ini, peneliti menyajikan hasil penelitian dan pembahasan yang diperoleh dari berbagai tahap proses penelitian yang telah dilakukan. Proses penelitian melibatkan beberapa langkah yang saling terkait dan memiliki kontribusi pada pemahaman yang lebih baik mengenai sentimen publik terhadap program makan siang gratis. Berikut adalah tahap-tahap yang diimplementasikan dalam proses penelitian:

3.1.1 Pengumpulan Data

Pada penelitian ini proses pengumpulan data dilakukan menggunakan bahasa Pemrograman Python yang dijalankan pada google colab. dengan kata kunci "Program Makan Siang Gratis". Yang diambil pada media sosial khususnya *platform X* untuk memperoleh data yang relevan dengan proses *Crawling* data. Berikut adalah kode program proses *crawling* data



```
# Crawl Data

filename = 'program_makan_gratis_ok.csv'
search_keyword = 'program makan siang gratis lang:id'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 2. Kode program proses pengumpulan data

	conversation_id_str	created_at	favorite_count	full_text	id_str
0	1795323892925616623	Wed May 29 07:05:55 +0000 2024	0	@Dennysiregar7 Cari dana utk program makan sia...	1795713665100083257
1	1795708596338241709	Wed May 29 06:47:46 +0000 2024	0	Presiden Terpilih Prabowo Subianto Ganti Progr...	1795708596338241709
2	1795702671254643070	Wed May 29 06:40:20 +0000 2024	1	@satriohendri mogaa2 program makan siang gratis...	1795706726823649467
3	1795684100524937604	Wed May 29 06:28:07 +0000 2024	0	@kikyasaputrii Program makan siang gratis adl p...	1795703648888197595
4	1795700048598315044	Wed May 29 06:13:52 +0000 2024	0	Prabowo juga menambahkan alasan program makan ...	1795700064473739692

Gambar 3. Hasil dari proses pengumpulan data tweet

Hasil data yang didapat dari proses crawling data sebanyak 1013 data. Data ini di simpan dalam bentuk file dengan format CSV. Data yang didapat masih berupa data mentah maka, perlu dilakukan pembersihan data untuk menghindari hal-hal yang dapat mengganggu pemrosesan data.

3.1.2 Cleaning Data

Cleaning data adalah proses untuk mengidentifikasi, mengoreksi, dan menghapus kesalahan atau ketidaksempurnaan yang dapat mempengaruhi hasil analisis. Pada penelitian ini proses pembersihan data dilakukan dengan bahasa pemrograman python pada Google Colab. beberapa langkah yang dilakukan diantaranya sebagai berikut:

- 1) Tahap pertama adalah penghapusan mention "@" pada tweet.
- 2) Tahap kedua adalah penghapusan hastga "#" pada tweet.
- 3) Selanjutnya adalah tahap penghapusan retweet "RT" pada tweet.
- 4) Kemudian dilakukan penghapusan tautan URL atau link pada data tweet.
- 5) Penghapusan numerik yang ada pada tweet.
- 6) Penghapusan semua karakter selain huruf seperti simbol dan tanda baca dan lainnya.

Proses pembersihan data ini bertujuan untuk memastikan bahwa data yang digunakan untuk analisis sentimen terbebas dari data yang tidak relevan. berikut kode program dan hasil dari proses cleaning data:

```

def clean_twitter_text(text):
    #menghapus karakter khusus
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r'RT[\s]+', '', text)
    text = re.sub(r'https://\S+', '', text)
    text = re.sub(r'[0-9]+', '', text)
    text = re.sub(r'[^A-Za-z ]', '', text)

    return text
df['full_text'] = df['full_text'].apply(clean_twitter_text)

```

	full_text
0	Cari dana utk program makan siang gratis wong ...
1	apa yg salah dengan program makan siang gratis...
2	moga program makan siang gratis yg sdh digant...
3	Program makan siang gratis adl program yg dib...
4	Program makan siang gratis itu sebenarnya memb...

Gambar 4. Proses *cleaning* data

Setelah dilakukan proses *cleaning* data, dilakukan proses penghapusan data yang bersifat duplikat atau data tweet yang lebih dari satu. Pada penelitian ini proses penghapusan data tweet yang bersifat duplikat dilakukan dengan menggunakan kode program sebagai berikut:

```

df = df.drop_duplicates(subset=("full_text"))

[12] df.duplicated().sum()
      0

[13] df.shape
      (920, 1)

```

Gambar 5. Proses penghapusan data duplikat

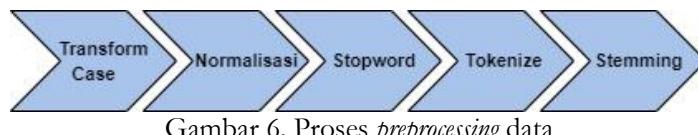
Setelah dilakukan proses pembersihan data dan penghapusan data duplikat, tersisa 920 data tweet siap digunakan pada tahap proses selanjutnya.

3.1.3 Pelabelan Data

Pada penelitian ini pelabelan dilakukan secara manual pada Microsoft Excel. Pelabelan dilakukan pada data dengan membagi sentimen menjadi dua kelas yaitu sentimen positif dan sentimen negatif. Sentimen yang berlabelkan positif adalah tweet yang berisi dukungan pada program makan siang gratis, sedangkan sentimen negatif adalah sentimen yang berisi ketidaksetujuan atau keluhan terhadap program makan siang gratis.

3.1.4 Preprocessing

Pada penelitian ini dilakukan dengan tahap preprocessing data. Preprocessing data adalah proses persiapan data yang akan diolah. Tahap ini untuk memastikan bahwa data yang digunakan pada saat analisis adalah data yang konsisten dan siap untuk masuk ke tahap pemodelan. Proses implementasi pada preprocessing melibatkan beberapa langkah penting sebagai berikut:



Gambar 6. Proses *preprocessing* data

1) Transform Case

Proses *Transform Case* dilakukan untuk mengubah teks teks menjadi huruf kecil sebagai penyamaan huruf yang digunakan, tahap ini bertujuan untuk menghindari perbedaan antara huruf kapital dan huruf kecil sehingga kata-kata yang identik diperlakukan sebagai entitas yang sama. Seperti kata “Makan Siang” menjadi “makan siang”. Berikut kode program dan hasil pada tahap proses transform case:

```
# mengubah teks menjadi lower
data['full_text'] = data['full_text'].str.lower()
data.head()
```

	full_text	Sentimen
0	cari dana utk program makan siang gratis lah i...	negatif
1	apa yang salah dengan program makan siang grat...	positif
2	semoga program makan siang gratis yg sudah di...	negatif
3	program makan siang gratis adl program yg dib...	negatif
4	program makan siang gratis itu sebenarnya memb...	positif

Gambar 1. Proses *trasnform case*

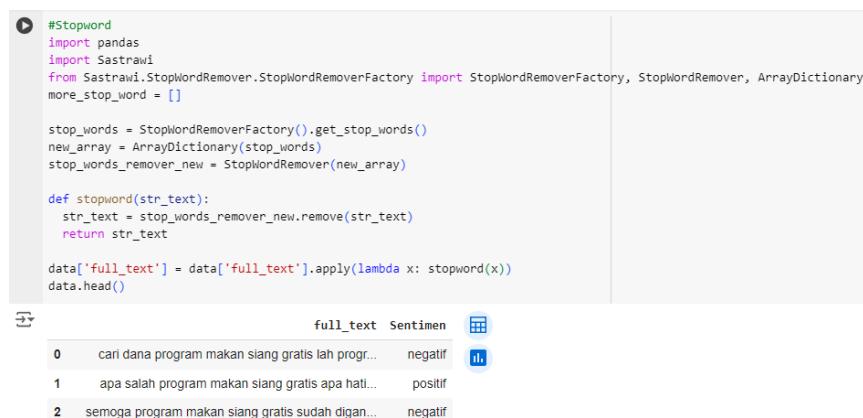
2) Normalisasi

Pada penelitian ini proses normalisasi bertujuan untuk memastikan bahwa teks yang dianalisis konsisten. Dalam hal ini, normalisasi dapat melibatkan beberapa langkah spesifik, termasuk mengubah kalimat singkat menjadi kalimat yang lebih standar atau lengkap. Seperti kata “krna” menjadi “karena”. Berikut kode program dan hasil pada tahap normalisasi:

Gambar 2. Proses Normalisasi

3) Stopword

Stopwords adalah kata-kata umum yang sering muncul dalam bahasa tetapi tidak berkontribusi pada makna dalam analisis teks. Pada penelitian ini proses *stopword* bertujuan untuk menghilangkan kata-kata yang tidak memberikan makna dalam konteks analisis. Penghapusan kata-kata ini membantu memfokuskan perhatian pada kata-kata yang lebih relevan dan berisi informasi yang lebih berarti. seperti “di”, “dan”, “itu”. Berikut kode program dan hasil pada tahap *stopword*:



```
#Stopword
import pandas
import Sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary
more_stop_word = []

stop_words = StopWordRemoverFactory().get_stop_words()
new_array = ArrayDictionary(stop_words)
stop_words_remover_new = StopWordRemover(new_array)

def stopword(str_text):
    str_text = stop_words_remover_new.remove(str_text)
    return str_text

data['full_text'] = data['full_text'].apply(lambda x: stopword(x))
data.head()
```

	full_text	Sentimen
0	cari dana program makan siang gratis lah progr...	negatif
1	apa salah program makan siang gratis apa hati...	positif
2	semoga program makan siang gratis sudah digan...	negatif

Gambar 3. Proses Stopword

4) Tokenize

Proses *tokenize* dilakukan untuk memecahkan teks menjadi kata atau unit-unit yang lebih kecil proses ini dapat dengan mudah memberi pemahaman struktur pada teks dan menganalisis teks secara efektif. Seperti kata “program kerja makan siang gratis” menjadi “program”, “kerja”, “makan”, “siang”, “gratis”. Berikut kode program dan hasil pada tahap *tokenize*:



```
#Tokenisasi
tokenized = data['full_text'].apply(lambda x:x.split())
tokenized
```

	full_text
0	[cari, dana, program, makan, siang, gratis, la...
1	[apa, salah, program, makan, siang, gratis, ap...
2	[semoga, program, makan, siang, gratis, sudah...
3	[program, makan, siang, gratis, program, dibua...
4	[program, makan, siang, gratis, sebenarnya, me...
...	
915	[moga, ngomong, gini, kemarin, milihnya, capre...
916	[aku, setuju, program, makan, siang, gratis]
917	[ditawari, program, satu, keluarga, satu, sarj...
918	[makan, siang, gratis, merupakan, program, efe...
919	[plot, twist, udah, terpilih, baru, ganti, nam...

Gambar 4. Proses Tokenisasi

5) Stemming

Stemming bertujuan untuk mengubah kata-kata menjadi bentuk dasar atau akar kata mereka. Proses ini melibatkan penghapusan akhiran atau prefiks dari kata untuk menyederhanakan variasi bentuk kata yang berbeda menjadi bentuk dasarnya. Seperti “melihat” menjadi “lihat”. Pada tahap ini proses stemming dengan melakukan import sastrawi untuk membangun program analisis teks, kemudian data di import dalam bentuk csv. Berikut kode program dan hasil pada proses stemming:



```
#Stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(full_text):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in full_text:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

tokenized.to_csv("/content/drive/MyDrive/Skripsi/program_makan_siang_gratis_stemming1.csv", index=False)
data_clean = pd.read_csv("/content/drive/MyDrive/Skripsi/program_makan_siang_gratis_stemming1.csv", encoding='latin1')
```

cuma kaum iq bilang program wowo sama samsul masuk akal emang nya gampang distribusin makan siang gratis seluruh masyarakat ta
mending gausah ada makan siang gratis terlalu paksa mending fokus lanjutkan program jokowi aja udah maksimal gimik kaya gin u
gojetin aja selamat makan tu program bodoh makan siang gratis

Gambar 5. Proses Stemming

3.1.5 Klasifikasi Naive Bayes

Proses klasifikasi naive bayes adalah salah satu metode yang digunakan dalam machine learning untuk memprediksi kelas dari dataset berdasarkan fitur yang ada[12]. Kelas sentimen pada dataset saat ini masih berlabelkan positif dan negatif. Maka, diperlukan proses untuk mengkonversi teks kelas sentimen menjadi biner dengan mengubah kelas sentimen positif menjadi “1”, dan negatif menjadi “0”. kode program yg digunakan untuk proses konversi sentimen sebagai berikut:

```
data_clean = data_clean.replace({'positif':1, 'negatif':0})
data_clean.head()
```

	full_text	Sentimen
0	cari dana program makan siang gratis lah progr...	0
1	apa salah program makan siang gratis apa hati ...	1
2	moga program makan siang gratis sudah ganti ja...	0
3	program makan siang gratis program buat perint...	0
4	program makan siang gratis sebenarnya bantu ba...	1

Gambar 6. Proses mengkonversi teks kelas sentimen menjadi biner

Untuk melihat seberapa banyak sentimen yang ada pada kelas positif dan negatif terkait program makan siang gratis dapat dilihat sebagai berikut:

```
data_clean['Sentimen'].value_counts()
```

Sentimen	count
0	744
1	176

Gambar 7. Proses dan hasil sentimen pada kelas positif dan negatif

Hasil pada sentimen yang ada dapat dilihat bahwa dari 920 data tweet 744 tweet memiliki kelas sentimen negatif dan 176 tweet memiliki kelas sentimen positif. Kemudian, dilakukan visualisasi untuk melihat kata yang sering muncul pada setiap sentimen. Berikut kode program dan hasilvisualisasi pada sentimen positif:

```
[ ] all_text_s0 = ' '.join(word for word in data_negatif["full_text"])
wordcloud = WordCloud(colormap='Reds', width=1000, height=1000, mode='RGBA', background_color='white').generate(all_text_s0)
plt.figure(figsize=(9, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Visualisasi Kata Negatif")
plt.margins(x=0, y=0)
plt.show()
```

Gambar 8. Proses visualisasi kelas sentimen positif



Gambar 9. Hasil visualisasi kelas sentimen positif

Berikut adalah contoh kode program untuk analisis sentimen menggunakan metode *Naive Bayes* dalam Python di Google Colab, beserta hasil visualisasi untuk sentimen negatif:

```
[ ] all_text_s0 = ' '.join(word for word in data_negatif["full_text"])
wordcloud = WordCloud(colormap='Reds', width=1000, height=1000, mode='RGBA', background_color='white').generate(all_text_s0)
plt.figure(figsize=(9, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Visualisasi Kata Negatif")
plt.margins(x=0, y=0)
plt.show()
```

Gambar 10. Proses visualisasi kelas sentimen negatif



Gambar 11. Hasil visualisasi kelas sentimen negatif

Berikut kode program dan hasil visualisasi sentimen berdasarkan diagram batang. Terlihat bahwa dataset yang memiliki sentimen negatif lebih banyak dibanding yang memiliki kelas sentimen positif:

The figure is a bar chart titled "Visualisasi Sentimen". The x-axis is labeled "Sentimen" and has two categories: 0 and 1. The y-axis is labeled "Jumlah" and ranges from 0 to 700. A legend box is present on the right side, titled "Sentimen", with two entries: 0 represented by a red square and 1 represented by a blue square. The bar for category 0 (red) reaches approximately 720 on the y-axis. The bar for category 1 (blue) reaches approximately 180 on the y-axis.

Sentimen	Jumlah
0	720
1	180

Gambar 12. Proses visualisasi sentimen positif dan negatif dengan diagram batang

Kemudian dilakukan proses pembagian data dengan banding 80:20 data latih dan data uji untuk melakukan klasifikasi pada model *naiive bayes*. Berikut kode program pada proses pembagian data:

```
#pembagian data latih dan data uji  
X = data_clean['full_text']  
y = data_clean['Sentimen']  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)
```

Gambar 13. Proses pembagian data uji dan data latih

Setelah dilakukan pembagian data latih dan data uji, dilakukan proses untuk mengonversi teks menjadi representasi numerik yang dapat diproses oleh model machine learning teks dengan kode program sebagai berikut:

```
#countvectorizer untuk mengubah fitur numerik
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_text_vec = vectorizer.transform(X_test)
```

Gambar 14. Proses mengonversi teks menjadi numerik

Tahap selanjutnya pada proses naive bayes dalam penelitian ini adalah melakukan klasifikasi mode *naive bayes* dengan multinomial naive bayes dengan kode program dan hasil sebagai berikut:

```
[67] naive_bayes = MultinomialNB()
naive_bayes.fit(X_train_resampled, y_train_resampled)

MultinomialNB()
```

Gambar 15. Proses klasifikasi *Naive Bayes*

3.1.6 Evaluasi

Selanjutnya dilakukan evaluasi model yang digunakan pada penelitian ini untuk melihat seberapa baik model *naive bayes* dalam melakukan analisis teks pada sentimen dengan *confusion matrix*. Berikut kode program dan hasil pada proses evaluasi dengan menggunakan *confusion matrix*:

```
#Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred, target_names=['negatif', 'positif'])

print("Akurasi Model Naive Bayes : ", accuracy)
print("\nKlasifikasi :\n", classification_rep)

Akurasi Model Naive Bayes :  0.8695652173913043

Klasifikasi :
      precision    recall  f1-score   support
negatif       0.86     1.00     0.93     153
positif       1.00     0.23     0.37      31

accuracy          0.87      --      0.87     184
macro avg       0.93     0.61     0.65     184
weighted avg    0.89     0.87     0.83     184
```

Gambar 16. Proses evaluasi dan hasil dengan *confusion matrix*

Berdasarkan hasil pengujian metode naive bayes pada dataset terkait program makan siang gratis pada *platform X*, menunjukkan nilai akurasi sebesar 86.95%, *precision* 93% *recall* 61% dan *F1-Score* 65% dari 167 sentimen positif dan 744 sentimen negatif.

3.2 Pembahasan

Penelitian ini menggunakan algoritma *Naive Bayes* untuk menganalisis sentimen publik terhadap Program Makan Siang Gratis yang diusung oleh pasangan calon presiden dan wakil presiden, Prabowo Subianto dan Gibran Rakabuming Raka. Berdasarkan hasil penelitian, dari total 920 tweet yang dianalisis, 744 tweet mengandung sentimen negatif, sementara hanya 176 tweet yang mengandung sentimen positif. Jumlah sentimen negatif yang jauh lebih tinggi ini menunjukkan bahwa program ini belum sepenuhnya mendapatkan dukungan dari masyarakat.

Sebagaimana diuraikan oleh Puad *et al.* (2024), sentimen negatif di media sosial sering kali terkait dengan ekspektasi yang tidak terpenuhi oleh pemerintah. Ketika sebuah program tidak sesuai dengan harapan publik atau jika terdapat masalah dalam pelaksanaannya, kritik publik akan lebih mudah terlihat di media sosial. Hal ini dapat menjadi penanda adanya kekhawatiran masyarakat terhadap efektivitas Program Makan Siang Gratis, yang meskipun memiliki tujuan yang baik, mungkin belum terlaksana sesuai harapan. Selain itu, sentimen negatif juga dapat menunjukkan kurangnya kepercayaan masyarakat terhadap kemampuan pemerintah dalam mengimplementasikan program tersebut dengan baik, sebagaimana dicatat dalam penelitian Azhar *et al.* (2022), di mana persepsi masyarakat sangat dipengaruhi oleh bagaimana kebijakan dipromosikan dan dijalankan.

Dalam penelitian ini, algoritma *Naive Bayes* memberikan akurasi sebesar 86.95%, yang menunjukkan bahwa model tersebut mampu mengklasifikasikan sentimen dengan cukup baik. Namun, nilai *recall* yang relatif rendah, yaitu 61%, menunjukkan bahwa model memiliki keterbatasan dalam mengidentifikasi seluruh sentimen negatif yang ada. Hal ini sejalan dengan temuan Syarifuddin (2020), yang mencatat bahwa ketidakseimbangan dalam data dapat memengaruhi performa model, terutama ketika data yang dianalisis memiliki lebih banyak sentimen negatif daripada positif. Ketidakseimbangan ini mungkin memengaruhi kemampuan algoritma untuk mengenali seluruh karakteristik sentimen negatif secara menyeluruh, sehingga menyebabkan rendahnya nilai *recall*. Penelitian oleh Verena *et al.* (2021) menunjukkan bahwa algoritma *Naive Bayes* sangat efektif untuk memproses data teks dalam jumlah besar. Namun, algoritma ini bekerja dengan asumsi independensi antar fitur, yang tidak selalu sesuai dengan realitas data teks yang cenderung memiliki keterkaitan antar kata. Dalam penelitian ini, asumsi independensi fitur dalam *Naive Bayes* mungkin menjadi salah satu faktor yang memengaruhi performa model. Untuk mengatasi keterbatasan ini, beberapa metode lain seperti *Relevance Frequency Feature Selection* dapat digunakan untuk meningkatkan performa model dalam mengidentifikasi pola sentimen yang lebih kompleks (Verena *et al.*, 2021).

Penelitian ini juga mengindikasikan bahwa komunikasi yang lebih baik dari pihak pemerintah dapat membantu mengurangi sentimen negatif yang muncul. Seperti yang disampaikan oleh Muchram (2024), persepsi masyarakat terhadap program pemerintah sangat dipengaruhi oleh bagaimana program tersebut dikomunikasikan kepada publik. Jika pemerintah dapat meningkatkan transparansi dan memperbaiki komunikasi publik mengenai tujuan dan manfaat dari Program Makan Siang Gratis, maka kemungkinan besar penerimaan publik terhadap program ini akan meningkat. Hal ini didukung oleh penelitian Wijaya (2021), yang menyebutkan bahwa program-program yang didukung oleh komunikasi yang baik cenderung mendapatkan respon yang lebih positif dari masyarakat. Pemerintah perlu melakukan evaluasi terhadap pelaksanaan Program Makan Siang Gratis dan memperhatikan umpan balik yang diterima dari masyarakat. Seperti yang disarankan oleh Priyanto *et al.* (2024), evaluasi program berbasis sentimen publik dapat menjadi alat yang efektif untuk menyesuaikan kebijakan agar lebih sesuai dengan kebutuhan dan ekspektasi masyarakat. Dengan demikian, program ini dapat lebih efektif dalam mencapai tujuannya serta meningkatkan kesejahteraan kelompok rentan yang menjadi sasaran utama.

4. Kesimpulan

Berdasarkan penelitian yang dilakukan mengenai sentimen publik pada sosial media X terhadap program makan siang gratis dengan metode Naive Bayes menggunakan bahasa pemrograman python pada Google Colab, didapatkan hasil sebanyak 920 data tweet dengan 176 tweet bernilai sentimen positif, 744 tweet bernilai sentimen negatif. menunjukan nilai akurasi sebesar 86.95%, *precision* 93% *recall* 61% dan *F1-Score* 65%. Disimpulkan bahwa sebagian besar publik yang berkomentar pada sosial media x memberikan reaksi negatif terhadap program makan siang gratis. Hasil ini dapat menjadi bahan evaluasi pemerintah dalam menentukan kebijakan yang efektif dan efisien dengan kebutuhan masyarakat.

5. Ucapan Terima Kasih

Terima kasih kepada pihak-pihak yang memberikan dukungan dan motivasi, sehingga penulis dapat menyelesaikan penelitian ini. penulis ingin mengucapkan terima kasih kepada :

- 1) Bapak Tundo M.Kom, selaku pembimbing yang telah memberikan ide dan arahan kepada penulis dalam menyelesaikan penulisan ini.
- 2) Orang Tua yang memberi dukungan dan doa yang tiada henti kepada penulis.
- 3) Teman-teman Mahasiswa/i seperjuangan yang mendukung dan memberi motivasi kepada penulis selama penelitian hingga penulisan karya tulis ilmiah ini.
- 4) Semua pihak yang tidak dapat disebutkan, yang memberi dukungan secara tulus.

6. Daftar Pustaka

- Ardiyanto, A., & Sela, E. I. (2024). Analisis Sentimen Opini Mahasiswa Terhadap Aplikasi Portal Mahasiswa UTY Menggunakan Metode Naïve Bayes Classifier. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(1), 525-540. DOI: <https://doi.org/10.35870/jimik.v5i1.508>.
- Atimi, R. L., & Pratama, E. E. (2022). Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia. *Jurnal Sains Dan Informatika*, 8(1), 88-96.
- Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(1), 267-281. DOI: <http://dx.doi.org/10.30645/j-sakti.v6i1.443>.
- Christiana, C., Kirana, R. N., & Ardyan, S. (2024). Analisis Pendukung Paslon 02 di Pemilu 2024 Pada Platform Media Sosial Twitter “X”. *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, 3(5), 28-35.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115-123. DOI: <https://doi.org/10.33365/jti.v14i2.679>.
- Muchram, M. (2024). Analisis Hubungan Latar Belakang Pendidikan, Visi Misi Dan Berita Media Online Capres 2024 Terhadap Pasar Modal. *Jurnal Media Akademik (JMA)*, 2(2). DOI: <https://doi.org/10.62281/v2i2.154>.
- Nurmalasari, M., Temesvari, N. A., & Maula, S. N. M. (2020). Analisis Sentimen terhadap Opini Masyarakat dalam Penggunaan Mobile-JKN untuk Pelayanan BPJS Kesehatan Tahun 2019. *Indonesian of Health Information Management Journal (INOHIM)*, 8(1), 35-44. DOI: <https://doi.org/10.47007/inohim.v8i1.208>.
- Priyanto, I., Dewanti, E. M., Tundo, T., Nurdin, M., & Kasiono, R. (2024). PENERAPAN ALGORITMA METODE NAÏVE BAYES UNTUK PENENTUAN PENERIMAAN BANTUAN PROGRAM INDONESIA PINTAR (PIP). *Jurnal Manajemen Informatika Jayakarta*, 4(2), 162-172. DOI: <https://doi.org/10.52362/jmijayakarta.v4i2.1355>.
- Puad, S., Garno, G., & Irawan, A. S. Y. (2023). Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1560-1566. DOI: <https://doi.org/10.36040/jati.v7i3.6920>.

Syarifuddinn, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), 23-28. DOI: <https://doi.org/10.33480/inti.v15i1.1347>.

Tanggraeni, A. I., & Sitokdana, M. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785-795. DOI: <https://doi.org/10.35957/jatisi.v9i2.1835>.

Toy, K. V. S., Sari, Y. A., & Cholissodin, I. (2021). Analisis Sentimen Twitter menggunakan Metode Naïve Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11), 5068-5074.

Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter. *Jambura Journal of Electrical and Electronics Engineering*, 3(2), 78-83. DOI: <https://doi.org/10.37905/jjeee.v3i2.10885>.

Yasir, M., & Suraji, R. (2023). Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube. *Jurnal Cahaya Mandalika ISSN 2721-4796 (online)*, 3(2), 180-192. DOI: <https://doi.org/10.36312/jcm.v3i2.1520>.