ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP PROGRAMMAKAN SIANG GRATIS MENGGUNAKAN METODE *NAÏVE BAYES*

Zona Arya Kusuma; Widi Widayat, S.Kom., M.Eng.

Program Studi Teknik Informatika, Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta

Abstrak

Twitter adalah platform media sosial yang memungkinkan interaksi tanpa batasan. Banyak topik yang dibahas di Twitter, termasuk program makan siang gratis yang menjadi bagian dari agenda pasangan calon presiden dan wakil presiden. Penelitian ini bertujuan untuk menganalisis sentimen terkait program makan siang gratis menggunakan klasifikasi Naïve Bayes. Sebelum dilakukan klasifikasi, dilakukan beberapa tahap prepocessing. Hasil penelitian menggunakan 1008 dataset yang dieksekusi menggunakan bahasa pemrograman python. Evaluasi dilakukan menggunakan Confusion Matrix menghasilkan akurasi 89%, presisi 90%, recall 90%, dan F1-score 89% dengan pembagian 80% untuk pelatihan dan 20% pengujian. Accuracy 89% menujukkan bahwa keseluruhan model bekerja dengan baik dalam mengklasifikasikan tweet. Recall 90% menunjukkan kemampuan model dalam menangkap hampir semua instance dari kelas positif, membuktikan minim nya kesalahan dalam bentuk False Negative. Precisioun 90% menunjukkan saat model memprediksi tweet, 90% dari prediksi bersifat benar. F1-Score 89% membuktikan adanya keseimbangan antara presisi dan recall, menunjukkan kinerja model yang konsisten dalam klasifikasi sentimen.

Kata Kunci: Analisis Sentimen, Twitter, Naïve Bayes.

Abstract

Twitter is a social media platform that allows interaction without restrictions. Many topics are discussed on Twitter, including the free lunch program which is part of the agenda of the presidential and vice presidential candidate pairs. This research aims to analyze sentiment related to free lunch programs using Naïve Bayes classification. Before classification, several preprocessing stages were carried out. The results of the study used 1008 datasets that were executed using the python programming language. Evaluation is done using Confusion Matrix resulting in 89% accuracy, 90% precision, 90% recall, and 89% F1-score with 80% division for training and 20% testing. Accuracy of 89% indicates that the overall model works well in classifying tweets. Recall 90% shows the model's ability to capture almost all instances of the positive class, proving minimal errors in the form of False Negative. Precisioun 90% shows that when the model predicts tweets, 90% of the predictions are correct. F1-Score of 89% proves a balance between precision and recall, showing consistent performance of the model in sentiment classification.

Keywords: Sentiment Analysis, Twitter, Naïve Bayes.

1. PENDAHULUAN

Dewasa ini, peralihan pemanfaatan teknologi telah banyak mengubah cara masyarakat dalam berinteraksi, Salah satu contohnya adalah interaksi dalam hal berbagi informasi, seseorang bisa membagikan informasi secara perlahan namun bersifat masif (Taufik et al., 2022). Kita semua bisa menyadari, dengan semakin cepatnya laju perkembangan teknologi ternyata berdampak pada semakin

cepatnya penyebaran informasi kepada publik melalui jejaring media sosial. Salah satu platform media sosial yang banyak digunakan untuk kebutuhan informasi adalah Twitter. Berdasarkan data dari (Kominfo Republik Indonesia, 2012) data pengguna twitter saat ini mencapai 19,5 Juta orang, jumlah pengguna yang banyak ini berbanding lurus dengan semakin banyaknya informasi yang bererdar pada platform twitter, banyak topik-topik sosial dan isu politik yang menjadi perbincangan yang menarik di platformtwitter.

Dari banyaknya topik-topik isu yang beredar di jejaring sosial twitter, salah satu isu yang tengah ramai diperbincangan adalah mengenai pelaksanaan program makan siang gratis, yang merupakan salah satu agenda pasangan calon presiden dan wakil presiden nomer urut 02, yakni Prabowo Subianto dan Gibran Rakabuming. Dikarenakan jumlah anggaran diambil dari dana BOS dan APBN membuat masyarakat Indonesia mengalami kekhawatiran dapat mempengaruhi biaya pendidikan dan gaji guru (Media Antaranews , 2024) oleh karena itu program tersebut mendapat banyak respon dari masyarakat Indonesia.

Dikarenakan banyaknya respon dari masyarakat, menimbulkan berbagai perspektif terhadap perihal yang menyebabkan munculnya berbagai macam opini pro dan kontra terhadap program tersebut, maka tujuan dari penelitian ini adalah melakukan analisis sentimen terhadap program makan siang gratis untuk mengetahui pandangan dari masyarakat Indonesia terhadap program makan siang gratis tersebut. Analisis sentimen adalah sebuah metode untuk mengetahui apa yang dipikirkan oleh masyarakat umum tentang suatu objek tertentu yang harus dipahami (Sulaeman et al., 2024).

Penelitian yang akan dilakukan menggunakan dataset berupa tweet berbahasa Indonesia. Dataset akan dikumpulkan dengan menggunakan kata kunci “program makan siang gratis”. Tahapan penelitian yang akan dilakukan diantaranya adalah pengumpulan data (*crawling*), *preprocessing data*, pelabelan data, klasifikasi data, dan evaluasi.

Proses classifier adalah metode yang dapat mengelompokkan data ke dalam beberapa kategori (Ade Dwi Dayani et al., 2024). Dalam penelitian ini, classifier yang dipilih adalah naïve bayes.

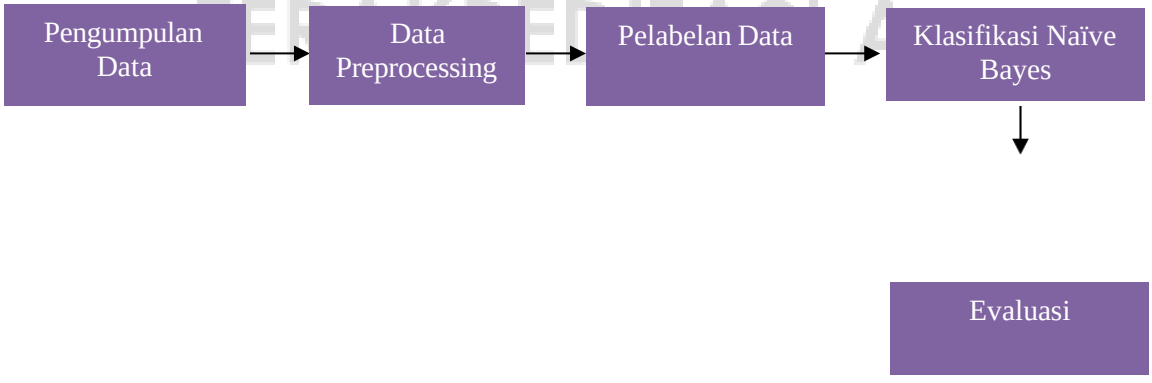
Naïve Bayes Classifier adalah jenis algoritma pembelajaran mesin yang umum diterapkan dalam tugas klasifikasi teks, seperti analisis sentimen. Analisis sentimen digunakan untuk menentukan sentimen atau gagasan yang tercermin dalam teks, seperti pada postingan media sosial dan label produk (Murni et al., 2023

Pada penelitian terdahulu telah banyak dilakukan penelitian analisis sentiment dengan menggunakan data Twitter, salah satunya adalah penelitian analisis sentimen terhadap topik kenaikan harga bahan bakar minyak (BBM). Penelitian tersebut menggunakan sejumlah 4419 dataset, menghasilkan sentimen positif sebanyak 53.3% (2357 data), negatif 31,2% (1378 data), dan netral 15,5% (684 data) menggunakan klasifikasi *Naïve Bayes* (Khatib Sulaiman et al., 2023). Penelitian serupa telah dilakukan oleh (Duei Putri et al., 2022) mengenai *Analisis Sentimen Kinerja Dewan*

Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. Penelitian ini menghasilkan sentiment positif sebanyak 75% (95 data), negatif 82% (758 data), dan netral 79% (693 data) dengan *accuracy score* 0.8 berdasarkan data pengujian yang mencakup 20% dari total data. Penelitian serupa telah dilakukan (Florensius Sianipar et al., 2023) mengenai *Analisis Sentimen Pembangunan Kereta Cepat Jakarta- Bandung di Media Sosial Twitter Menggunakan Metode Naïve Bayes.* Penelitian tersebut menghasilkan sentiment positif sebanyak 668 data, negatif 673 data dan netral 665 data dengan *accuracy score* 71%, *precision 73%* dan *recall 89%*. Masih seputar analisis sentiment terhadap pengguna twitter, (Tiara Susilawati et al., 2024) melakukan penelitian mengenai *Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes.* Penelitian tersebut menghasilkan masyarakat cenderung mendukung boikot produk Israel dengan *accuracy 95%, precision 96%, recall 95%, dan F1 score 95%*. Penelitian yang dilakukan menunjukan tingkat keberhasilan tinggi dalam mengidentifikasi sentiment yang dianalisis. Penelitian serupa telah dilakukan oleh (Hariyanti et al., 2024) mengenai *Analisis Sentimen Terhadap Putusan Mahkaman Konstitusi Tentang Batasan Umur Capres dan Cawapres Menggunakan Metode Naïve Bayes.* Penelitian tersebut menghasilkan 6,6% sentiment positif, 2,5% netral, dan 90,9% sentiment negatif. Berdasarkandata tersebut, menandakan ketidakpuasan yang dominan di kalangan publik.

Maka dari itu, penelitian akan dilakukan untuk menganalisis terhadap sentimen masyarakat terhadap program makan siang gratis menggunakan metode klasifikasi Naïve Bayes. Metode Naïve Bayes dipilih karena menghasilkan akurasi yang cukup baik dan fleksibilitasnya dalam menerima berbagai tipe input serta kemampuannya dalam memprosesdata secara cepat.

2. METODE

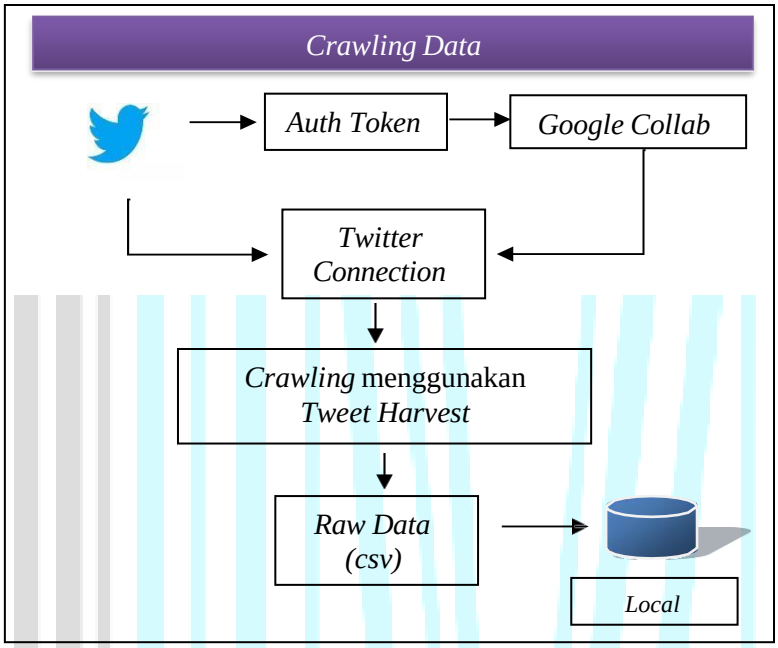


Gambar 1. Tahapan Metodologi Peneltian

Gambar 1 (Chouhan et al., 2023) merepresentasikan tahapan dalam melakukan penelitian yang digunakan untuk menyelidiki data yang diungkapkan oleh publik di platform media sosial Twitter mengenai program makan siang gratis dalam tweet berbahasa Indonesia. Berikut merupakan uraian dari tahapan penelitian :

2.1 Pengumpulan data

Proses pengumpulan data dengan teknik *crawling* data menggunakan aplikasi *google colab* dan tools *tweet harvest* serta melakukan eksekusi kode *python* terhadap tweet berbahasa Indonesia yang dengan keyword "program makan gratis Indonesia" dilakukan dari tanggal 1 November 2023 sampai 15 April 2024. Tahapan crawling data diperlihatkan pada gambar 2.



Gambar 2. Tahapan *Crawling* Data

2.2 Preprocessing Data

Data yang telah didapatkan dari proses *crawling* kemudian dilakukan tahap *preprocessing*. *Preprocessing* dilakukan untuk memisahkan data twitter dari informasi yang tidak diperlukan sehingga data lebih terstruktur (Patel & Passi, 2020).

Berikut tahapan dalam melakukan *preprocessing* data:

1) *Cleaning* untuk mengurangi bagian teks di seluruh prosesnya. Proses tersebut dengan menghapus kolom yang tidak diperlukan.

   Contoh: Membersihkan tanda baca, menghapus nama user, dan menghapus url (Chang et al., 2023)
2) *Case Folding* secara khusus, adalah mengubah ukuran huruf kapital menjadi bentuk huruf kecil (Yusran et al., 2023).
3) *Normalization* dilakukan untuk mengatasi munculnya istilah-istilah yang tidak standar dengan melakukan proses konversi terhadap kata tidak baku menjadi kata baku(Imaduddin et al., 2023).
4) *Stopword Removal* merupakan proses untuk menghapus kata penghubung.
5) *Tokenization* adalah proses membagi teks menjadi unit-unit kecil yang disebut token, yang darinya data teks akan diekstraksi untuk membentuk kata-kata yang lebih bermakna (Nazim

Uddin et al., 2023).

6) *Stemming* adalah proses mengekstraksi makna dari sebuah kata sehingga hanya kata- kata yang paling penting saja yang tersisa (Qorib et al., 2023).

2.3 Pelabelan Data

Tujuan pelabelan data adalah untuk mengklasifikasikan semua data yang tersedia sebagai penentu. Proses pelabelan dilakukan dengan library *TextBlob* dengan memberikan keterangan positif, negatif, maupun netral.

2.4 Klasifikasi *Naïve Bayes*

Proses klasifikasi akan memisahkan data menjadi dua kategori: data pelatihan dan data pengujian. Data pelatihan digunakan untuk mengembangkan model yang dapat mengidentifikasi apakah sebuah tweet bersifat positif atau negatif. Klasifikasi Naïve Bayes termasuk dalam jenis supervised learning. Supervised learning adalah metode dimana model dilatih menggunakan data yang sudah diberi label (Erfina & Alamsyah, 2023). Klasifikasi Naive Bayes yang digunakan merupakan Multinomial Naïve Bayes. Algoritma Multinomial Naive Bayes adalah salah satu metode pembelajaran berbasis probabilitas yang menggunakan teori Bayes yang diterapkan dalam Pemrosesan Bahasa Alami (NLP). Model ini mengilustrasikan dua fakta: apakah kata yang disebutkan di atas muncul di dokumen atau tidak, dan seberapa sering kata tersebut muncul di dokumen (Yuyun et al.,2021).

Multinomial Naïve Bayes dapat dirumuskan sebagai berikut

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

Keterangan:

$P(p|n)$: Peluang suatu dokumen (tk) muncul

n : Jumlah total dokumen

P : sentimen

Selanjutnya, untuk menghitung sentimen, digunakan rumus sebagai berikut;

$$P(t_k|p) = \frac{count(t_k|p)+1}{count(t_p)+|V|} \quad (2)$$

Keterangan:

$(t_k|p)$: jumlah kemunculan tk dalam dokumen yang memiliki sentimen p

(tp) : jumlah token dalam dokumen dengan sentimen p

2.5 Evaluasi

Setelah data diklasifikasikan, hasilnya akan dievaluasi dengan menggunakan metode *confusion matrix*. Langkah ini bertujuan guna menilai kinerja model yang telah dibangun (Yunita & Kamayani, 2023). Hasil evaluasi meliputi *recall*, *accuracy*, dan *precision*. *Confusion matrix* diperlihatkan pada tabel 1.

Tabel 1. *Confusion matrix*

Actual class	Predicted Class	
	Positive	Negative
Positive	True Positive (TP)	**False Negative (FN)**
Negative	**False Positive (FP)**	**True Negative (TN)**

Berdasarkan Tabel 1 diperoleh empat nilai keluaran :

2.5.1 Accuracy

Mengukur jumlah prediksi yang benar (positif maupun negatif) keseluruhanprediksi yang dibuat.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2.5.2 Recall

Mengukur jumlah dari sampel yang benar-benar terdeteksi (positif maupunnegatif)

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

2.5.3 Precision

Mengukur jumlah dari prediksi yang benar-benar terdeteksi (positif maupunnegatif)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

2.5.4 F1-Score

Mengukur kinerja model klasifikasi

$$\text{F1-Score} = 2 * \frac{\text{Precisioun} * \text{Recall}}{\text{Precisioun} + \text{Recall}} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

Proses pengumpulan data dengan teknik *crawling* data menggunakan aplikasi *google colab* dan tools *tweet harvest* serta melakukan eksekusi kode *phython* terhadap tweet berbahasa Indonesia yang dengan keyword “program makan gratis Indonesia”. Proses pengambilan data dilakukan dari tanggal 1 November 2023 sampai 15 April 2024 dan mendapatkan sebanyak 1423 dataset. Hasil *crawling* diperlihatkan pada tabel 2.

Tabel 2. Data *Crawling* tweet

Kata kunci	Jumlah Data	Jumlah Data tanpa duplikasi
Program makan gratis Indonesia	1423	**1008**
Total	**1423**	**1008**

3.2 Preprocessing

Preprocessing dilakukan dengan menggunakan jumlah dataset tanpa duplikasi sebanyak 1008 dataset. Pada preprocessing ini, dilakukan dengan menggunakan bahasa *python* dan di implementasikan pada aplikasi *google colab*. Sebelum dilakukan klasifikasi, dilakukan beberapa tahap pemrosesan teks, termasuk pembersihan data, perubahan kecil huruf, normalisasi, penghapusan kata-kata umum, tokenisasi, stemming, dan pelabelan data. Tujuan dilakukan tahap preprocessing adalah memudahkan dalam pengolahan data, dan membuat hasil data memiliki akurasi yang baik.

3.2.1 Cleaning

Dalam tahapan ini, merupakan proses pembersihan tanda baca, penghapusan nama user, dan penghapusan URL. Hasil *cleaning* diperlihatkan pada tabel 3.

Tabel 3. Hasil *Cleaning*

Sebelum	Sesudah
@tonnysofijan @bul_ala @AH_SiregarXIX Gini nih kak kalau ga mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yg gak masuk akal untuk Indonesia saat ini. Darimana uangnya kalau BKN naikin BBM dan naikin pajak blm buat IKN	**Gini nih kak kalau ga mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yg gak masuk akal untuk Indonesia saat ini. Darimana uangnya kalau BKN naikin BBM dan naikin pajak blm buat IKN**

3.2.2 Case Folding

Dalam tahapan ini, proses yang dilakukan adalah mengonversi huruf besar menjadihuruf kecil. Hasil dari perubahan ukuran huruf diperlihatkan pada tabel 4.

Tabel 4. Hasil Case Folding

Sebelum	Sesudah
Gini nih kak kalau ga mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yg gak masuk akal untuk Indonesia saat ini. Darimana uangnya kalau BKN naikin BBM dan naikin pajak blm buat IKN	gini nih kak kalau ga mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yg gak masuk akal untuk indonesia saat ini. darimana uangnya kalau bkn naikin bbm dan naikin pajak blm buat ikn

3.2.3 Normalization

Dalam tahapan ini, proses yang akan dilakukan adalah perubahan istilah tidak baku menjadi istilah baku, sebagai contoh perubahan kata “ga” menjadi “tidak”, “bkn” menjadi “bukan” dan “blm” menjadi “belum”. Hasil Normalization diperlihatkan padatabel 5

Tabel 5. Hasil *Normalization*

Sebelum	Sesudah
gini nih kak kalau ga mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yg gak masuk akal untuk indonesia saat ini. darimana uangnya kalau bkn naikin bbm dan naikin pajak blm buat ikn	gini nih kak kalau tidak mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yang tidakk masuk akal untuk indonesia saat ini. darimana uangnya kalau bukan naikin bbm dan naikin pajak belum buat ikn

3.2.4 Stopwords Removal

Dalam tahapan ini, proses yang akan dilakukan adalah penghapusan kata hubung,contoh, seperti kata “yang”. Hasil *Stopwords Removal* diperlihatkan pada tabel 6.

Tabel 6. Hasil *Stopwords Removal*

Sebelum	Sesudah
gini nih kak kalau tidak mau dikaitkan ya jadi program yang kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program yang tidakk masuk akal untuk indonesia saat ini. darimana uangnya kalau bukan naikin bbm dan naikin pajak belum buat ikn	gini nih kak kalau tidak mau dikaitkan ya jadi program kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program tidakk masuk akal untuk indonesia saat ini. darimana uangnya kalau bukan naikin bbm dan naikin pajak belum buat ikn

3.2.5 Tokenization

Dalam tahapan ini, proses yang akan dilakukan adalah membagi kalimat menjadi kata atau unit kecil seperti token. Setelah dilakukannya *Stopwords Removal* maka setelahnya dilakukan *Tokenization*. Hasil *Tokenization* diperlihatkan pada tabel 7.

Tabel 7. Hasil *Tokenization*

Sebelum	Sesudah
gini nih kak kalau tidak mau dikaitkan ya jadi program kalian banggain itu seperti program makan siang gratis bagi kami pendukung 01 itu tuh program tidak masuk akal untuk indonesia saat ini. darimana uangnya kalau bukan naikin bbm dan naikin pajak belum buat ikn	[gini, nih, kak, kalau, tidak, mau, dikaitkan, ya, jadi, program, kalian, banggain, itu, seperti, program, makan, siang, gratis, bagi, kami, pendukung, 01, itu, tuh, program, tidak, masuk , akal, untuk, indonesia, saat , ini, darimana, uangnya, kalau, bukan, naikin, bbm, dan, naikin, pajak, belum, buat, ikn]

3.2.6 Stemming

Pada tahapan ini, proses yang akan dilakukan adalah mengubah kata menjadi bentuk kata dasar, sebagai contoh penghapusan kata imbuhan. Hasil *Stemming* diperlihatkanpada tabel 8.

Tabel 8. Hasil *Stemming*

Sebelum	Sesudah
[gini, nih, kak, kalau, tidak, mau, dikaitkan, ya, jadi, program, kalian, banggain, itu, seperti, program, makan, siang, gratis, bagi, kami, pendukung, 01, itu, tuh, program, tidak, masuk , akal, untuk, indonesia, saat , ini, darimana, uangnya, kalau, bukan, naikin, bbm, dan, naikin, pajak, belum, buat, ikn]	gini kak kalau tidak mau kait ya jadi program kalian bangga itu seperti program makan siang gratis bagi kami pendukung 01 itu program tidak masuk akal untuk Indonesia saat ini darimana uangnya kalau bukan naik bbm dan naik pajak belum buat ikn

3.3 Labeling

Setelah proses preprocessing selesai, langkah selanjutnya adalah pelabelan dengan menetapkan keterangan positif, negatif, atau netral. Sebelum diberikan keterangan sentiment, data hasil stemming diubah kedalam bahasa inggris dikarenakan library *TextBlob* dirancang menggunakan bahasa inggris dalam pemberian label. Tahapan ini menggunakan sejumlah 1008 dataset. Terdapat 3 jenis indikator label yang digunakan, indikator positif jika polaritas antara 0.0 hingga 1.0, indikator netral jika polaritas sama dengan 0.0, dan negatif jika polaritas antara 0.0 hingga -1.0. Penetapan nilai polaritas telah didefinisikan library *TextBlob* tanpa melakukan perhitungan sendiri (Parlika et al., 2020). Hasil *Labeling* dapat diperlihatkan pada tabel 9 dan tabel 10.

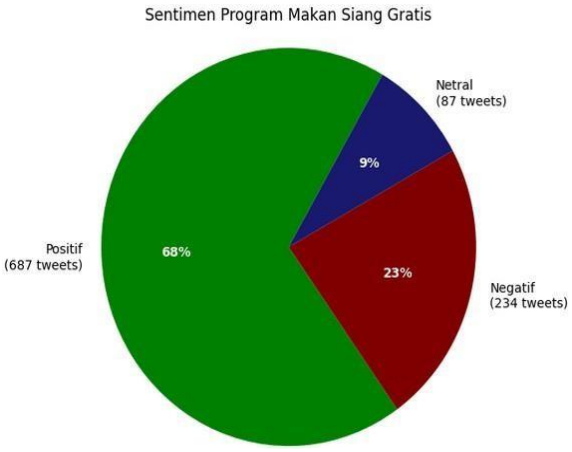
Tabel 9. Hasil *Labeling*

Labeling	Jumlah data
Positif	687
Negatif	234
Netral	87
Total data	1008

Tabel 10. Keterangan *Labeling*

Tweet	Labeling
Alhamdulillah, jika Indonesia bisa maju di bawah kepemimpinan Prabowo, saya akan melihat program-programnya, makan siang gratis, tinggal melanjutkan program 3 periode Jokowi.	Positif
program makan siang caleg gratis, program bodoh untuk rakyat Indonesia	Negatif
tidak, tidak miskin secara struktural, jadi jika orang Indonesia tidak bekerja, bisnis memiliki program makan siang gratis untuk anak-anak sekolah, beban orang tua akan lebih ringan dan mereka dan mereka bisa fokus pada ekonomi keluarga.	Netral

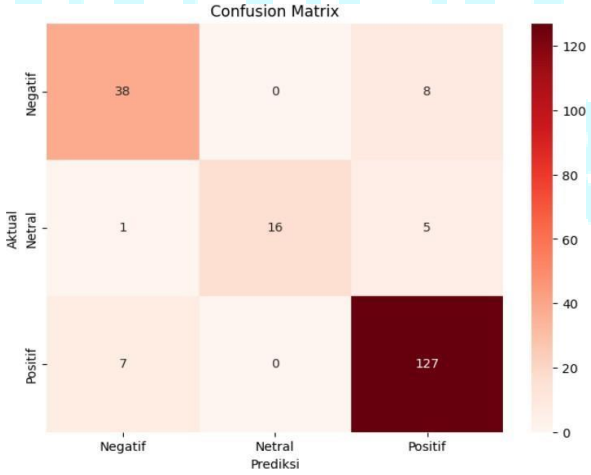
Pelabelan menggunakan library TextBlob menunjukkan bahwa 68% (687) data memiliki sentimen positif, 23% (234) data memiliki sentimen negatif, dan 9% (87) data memiliki sentimen netral, seperti yang diperlihatkan pada Gambar 2.



Gambar 3. Hasil *Labeling*

3.4 Klasifikasi *Naïve Bayes*

Proses klasifikasi 1008 data menggunakan algoritma Naïve Bayes dilakukan dengan menggunakan metode split data, dimana data latih diterapkan dengan 80%, dan 20% sebagai data uji. Penggunaan perbandingan 80:20 membantu memaksimalkan penggunaan data untuk pelatihan dan menyediakan subset data yang cukup besar untuk evaluasi kinerja model. Hasil klasifikasi *Naïve Bayes* menghasilkan *accuracy* sebesar 89%, *precision* 90%, dan *recall* 90%, berdasarkan evaluasi dengan metode *confusion matrix*. Metode *confusion matrix* diperlihatkan pada gambar 3 dan klasikasi *Naive Bayes*untuk setiap kelas diperlihatkan pada tabel 10.



Gambar 4. *Confusion Matrix*

Tabel 11. Hasil klasifikasi *Naïve Bayes*

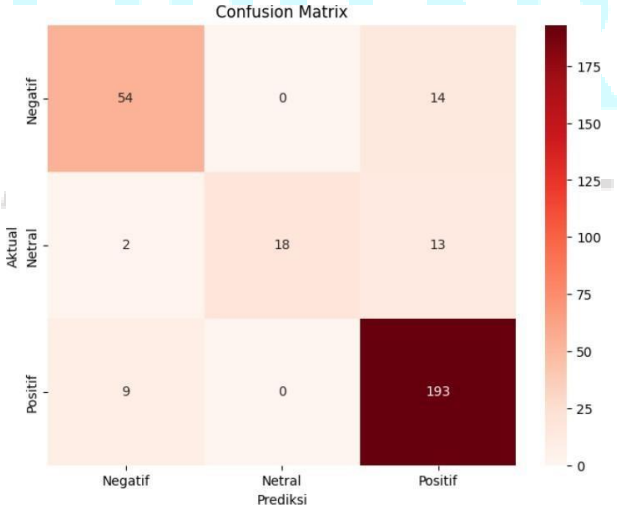
Label	Precision	Recall	F1-Score	Accuracy
Positif	0.91	0.95	0.93	**0.89**
Negatif	0.83	0.83	0.83	
Netral	**1.00**	**0.73**	**0.84**	

Hasil klasifikasi dapat dilihat pada tabel 10, sentimen negatif memiliki tingkat presisi yang baik (83%) yang menunjukkan sebagian tweet diklasifikasikan konsisten sebagai negatif, tingkat *recall* sebesar (83%) menunjukkan bahwa semua sentimen yang terdeteksi berhasil diklasifikasikan sebagai negatif, meskipun ada beberapa sentimen negatif yang terlewatkan dan *F1-Score* sebesar (83%) menunjukan keseimbangan yang baik antara presisi dan recall, Sentimen netral memiliki *precision* sebesar (100%) menunjukan bahwa model sangat tepat dalam memprediksi sentimen netral, benar-benar memiliki nilai netral. Meskipun *recall* dan *F1-Score* cenderung lebih rendah, menunjukkan bahwa ada beberapa tweet netral yang tidak terdeteksi.

Sentimen positif memiliki *precision* (91%) menandakan bahwa tweet yang diklasifikasikan sebagai positif benar memiliki sentimen positif, *recall* sebesar (95%) menunjukkan bahwa model sangat baik dalam mendeteksi sentimen positif dan *F1-Score* sebesar (93%), menunjukkan keseimbangan yang baik antara *precision* dan *recall.*

3.5 Penggunaan Split data rasio 70:30

Proses klasifikasi 1008 data menggunakan algoritma Naïve Bayes dilakukan dengan menggunakan metode split data, dimana data latih diterapkan dengan 70%, dan 30% sebagai data uji. Pengunaan perbandingan 70:30 membantu model dapat menggeneralisasi dengan baik dan performa dapat diandalkan pada data baru yang tidak terlihat. Hasil dari klasifikasi *Naïve Bayes* menunjukkan *accuracy* sebesar 87%, *precision* 88%, dan *recall* 87%. Dengan evaluasi metode menggunakan *confusion matrix.* Metode *confusion matrix* dapat dilihat pada gambar 4 dan klasikasi *Naive Bayes* untuk setiap kelas diperlihatkan pada tabel 11.



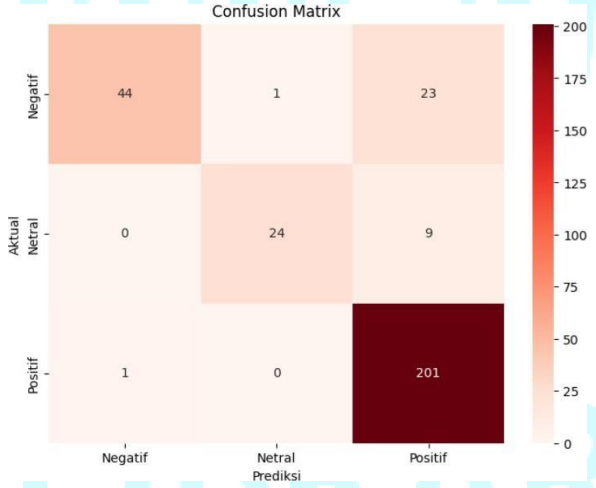
Gambar 5. *Confusion Matrix*

Tabel 12. Hasil klasifikasi *Naïve Bayes*

Label	Precision	Recall	F1-Score	Accuracy
Positif	0.88	0.96	0.92	**0.87**
Negatif	0.83	0.79	0.81	
Netral	**1.00**	**0.55**	**0.71**	

3.6 Komparasi Support Vector Machine

Proses klasifikasi 1008 data menggunakan model SVM dilakukan dengan menggunakan metode split data, dimana data latih diterapkan dengan 70%, dan 30% sebagai data uji. Hasil dari model SVM menunjukkan tingkat akurasi sebesar 88%. Dengan evaluasi metode menggunakan *confusion matrix.* Metode *confusion matrix* diperlihatkan pada gambar 5 dan klasikasi *Support Vector Machine* untuk setiap kelas diperlihatkan pada tabel 12.



Gambar 6. *Confusion Matrix*

Tabel 13. Hasil Support Vector Machine

Label	Precision	Recall	F1-Score	Accuracy
Positif	0.86	1.00	0.92	**0.88**
Negatif	0.98	0.65	0.78	
Netral	**0.96**	**0.73**	**0.83**	

4. PENUTUP

Berdasarkan hasil penelitian menggunakan metode *Naïve Bayes Classifier* menggunakan 1008 tweet yang diposting antara tanggal 1 November 2023 hingga 15 April 2024, dengan melakukan langkah-langkah *preprocessing* serta pemberian label sentimen menggunakan library *Textblob*. Langkah selanjutnya, melibatkan pembuatan model *Multinomial Naïve Bayes* dengan membagi data menjadi

80% untuk pelatihan dan 20% untuk pengujian. Evaluasi dilakukan menggunakan metode *confusion matrix*, yang menghasilkan metrik klasifikasi *Naive Bayes* dengan *accuracy* sebesar 89%, *recall* 90%, *precision* 90%, dan *F1-Score* sebesar 89%. *Accuracy* 89% menujukkan bahwa keseluruhan model bekerja dengan baik dalam mengklasifikasikan tweet. *Recall* 90% menunjukkan kemampuan model dalam menangkap hampir semua instance dari kelas positif, membuktikan minim nya kesalahan dalam bentuk *False Negative. Precisioun* 90% menunjukkan saat model memprediksi tweet, 90% dari prediksi bersifat benar. *F1-Score* 89% membuktikan adanya keseimbangan antara presisi dan recall, menunjukkan kinerja model yang konsisten dalam klasifikasi sentimen.

DAFTAR PUSTAKA

Ade Dwi Dayani, Yuhandri, & Widi Nurcahyo, G. (2024). Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine.

Chang, V., Liu, L., Xu, Q., Li, T., & Hsu, C. H. (2023*). An improved model for sentiment analysis on luxury hotel review. Expert Systems, 40(2).*

Chouhan, K., Yadav, M., Rout, R. K., Sahoo, K. S., Jhanjhi, N. Z., Masud, M., & Aljahdali, S. (2023). *Sentiment Analysis with Tweets Behaviour in Twitter Streaming API. Computer Systems Science and Engineering, 45(2).*

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier.

Erfina, A., & Alamsyah, M. R. N. R. (2023*). Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. Data and Metadata.*

Florensius Sianipar, J., Ramadhan, Y. R., & Jaelani, I. (2023). Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes.

Hariyanti, Y., Kacung, S., & Santoso, B. (2024). Analisis Sentimen Terhadap Putusan Mahkamah Konstitusi Tentang Batasan Umur Capres dan Cawapres Menggunakan Metode Naive Bayes.

Imaduddin, H., Yusfida A'la, F., & Nugroho, Y. S. (2023). *Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach.*

Khatib Sulaiman, J., Danirmala, T., Sulistyo Nugroho, Y. (2023). Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter.

Murni, M., Riadi, I., & Fadlil, A. (2023). Analisis Sentimen *HateSpeech* pada Pengguna Layanan Twitter dengan Metode *Naïve Bayes Classifier* (NBC).

Nazim Uddin, M., Ferdous Bin Hafiz, M., Hossain, S., & Mohammad Mominul Islam, S. (2023). *Drug Sentiment Analysis using Machine Learning Classifiers.*

Parlika, R., Ilham Pradika, S., Hakim, A. M., & Kholilul, R. N. M. (2020). Analisis Sentimen Twitter Terhadap *Bitcoin* dan *Cryptocurrency* Berbasis *Python TextBlob*.

Patel, R., & Passi, K. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. Internet of Things, 1(2).

Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cotae, P. (2023). *Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications, 212*.

Sulaeman, M., Bachrun', M., Romadhoni, U., Matheos Sarimole, F., & Septian, W. (2024). Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine.

Taufik, A., Kom, S., Bernadus Gunawan Sudarsono, M., & Kom, M. (2022). Pengantar Teknologi Informasi.

Tiara Susilawati. (2024). Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes.

Yunita, R., Kamayani, M. Perbandingan Algoritma SVM dan Naïve Bayes Pada Analisis Sentimen Kebijakan Penghapusan Kewajiban Skripsi

Yusran, M., Siswanto, S., & Islamiyati, A. (2023). *SISTEMASI: Jurnal Sistem Informasi Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking*.

Yuyun, Nurul Hidayah, & Supriadi Sahibu. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 820–826.