

Analisa Sentimen terhadap Twitter Pemilu 2024 menggunakan Perbandingan Algoritma Naïve Baiyes

Rian Damara¹, Reyhan Tri Rahmaddyan², Ajif Yunizar Pratama Yusuf³ dan Tb Ai Munandar⁴

Universitas Bhayangkara Jakarta Raya^{1,2}

Email : 202110715266@mhs.ubharajaya.ac.id, 202110715140@mhs.ubharajaya.ac.id, ajif.yunizar@dsn.ubharajaya.ac.id, tbaimunandar@gmail.com

ABSTRAKSI

Dalam era digital, analisis sentimen menjadi penting untuk memahami pandangan publik terhadap berbagai isu, termasuk pemilihan umum (Pemilu). Dalam konteks Pemilu 2024, penelitian ini bertujuan untuk menganalisis sentimen yang diekspresikan di platform Twitter terkait dengan Pemilu 2024. Dua algoritma klasifikasi utama, yaitu Naïve Bayes, digunakan untuk membandingkan performa mereka dalam mengklasifikasikan sentimen positif, negatif, dan netral. Data Twitter yang dikumpulkan menggunakan teknik crawling selama periode kampanye Pemilu 2024 digunakan sebagai dataset. Kemudian, data tersebut diproses untuk menghilangkan noise dan melakukan preprocessing teks, termasuk tokenisasi, stemming, dan penghapusan stop words. Setelah itu, Naïve Bayes diterapkan untuk mengklasifikasikan sentimen dari tweet-tweet yang dikumpulkan. Naïve Bayes, dengan pendekatannya yang berbasis probabilistik dan asumsi independensi fitur, menawarkan solusi yang cepat dan sederhana untuk masalah klasifikasi. Hasil analisis menunjukkan bahwa algoritma mampu mengklasifikasikan sentimen dengan baik. Dalam pengujian dengan data uji yang terpisah, Naïve Bayes mencapai akurasi sekitar 82%. Namun demikian, algoritma ini memiliki kelebihan dan kelemahan yang perlu dipertimbangkan dalam konteks analisis sentimen Twitter terkait Pemilu 2024. Misalnya, Naïve Bayes yang lebih efisien dalam hal waktu dan sumber daya. Penelitian ini menyimpulkan bahwa meskipun Naive Baiyes memberikan hasil yang lebih akurat, pemilihan algoritma terbaik bergantung pada kebutuhan spesifik dari analisis, seperti kecepatan pemrosesan dan ketersediaan sumber daya. Penelitian lebih lanjut disarankan untuk mengeksplorasi metode hybrid dan teknik deep learning untuk meningkatkan akurasi dan efisiensi analisis sentimen pada platform media sosial. Data yang diolah berjumlah 1500 tweets. Hasil dari Penelitian ini menunjukkan bahwa klasifikasi data twitter dengan algoritma Naive Baiyes menghasilkan akurasi sebesar 80%.

Kata Kunci: analisis sentimen, twitter, pemilu 2024, naïve bayes

ABSTRACT

In the digital era, sentiment analysis has become essential for understanding public opinion on various issues, including general elections. In the context of the 2024 General Election (Pemilu), this study aims to analyze sentiments expressed on the Twitter platform regarding the event. A primary classification algorithm, Naïve Bayes, was used to classify sentiments into positive, negative, and neutral categories and compare its performance. Twitter data was collected using a crawling technique during the 2024 election campaign period and used as the dataset. The data was then processed to remove noise and underwent text preprocessing, including tokenization, stemming, and stop word removal. Subsequently, the Naïve Bayes algorithm was applied to classify the sentiment of the collected tweets. Naïve Bayes, with its probabilistic approach and feature independence assumption, offers a fast and straightforward solution for classification tasks. The analysis results show that the algorithm was able to classify sentiments effectively. In tests using a separate test set, Naïve Bayes achieved an accuracy of approximately 82%. However, this algorithm has strengths and weaknesses that must be considered in the context of sentiment analysis on Twitter related to the 2024 election. For example, Naïve Bayes is more efficient in terms of time and resources. The study concludes that although Naive Baiyes produced accurate results, selecting the best algorithm depends on specific analysis needs, such as processing speed and resource availability. Further research is recommended to explore hybrid methods and deep learning techniques to enhance the accuracy and efficiency of sentiment analysis on social media platforms. The processed data consisted of 1,500 tweets. This study shows that the classification of Twitter data using the Naïve Bayes algorithm achieved an accuracy of 80%.

Keywords: sentiment analysis, twitter, pemilu 2024, naïve bayes

Penulis Korespondensi

Rian Damara

Tanggal Submit : 03/07/2024

Tanggal Diterima : 17/07/2025

Tanggal Terbit : 25/07/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Indonesia merupakan salah satu negara yang menganut sistem demokrasi [1]. Template Pemilihan umum (Pemilu) adalah salah satu peristiwa penting dalam kehidupan suatu negara, di mana masyarakat memiliki kesempatan untuk memilih pemimpin mereka melalui proses demokratis. Dalam era digital saat ini, media sosial telah menjadi platform utama di mana opini, pandangan, dan sentimen terkait dengan pemilu dapat diekspresikan secara luas dan cepat oleh masyarakat. Pemilihan umum merupakan salah satu pilar penting dalam demokrasi, yang memungkinkan warga negara untuk memilih pemimpin dan wakil mereka. Di era digital ini, media sosial telah menjadi platform yang signifikan untuk menyuarakan opini publik. Salah satu platform media sosial yang paling banyak digunakan untuk diskusi politik adalah Twitter. Twitter, dengan formatnya yang ringkas dan real-time, memungkinkan pengguna untuk berbagi pandangan, diskusi, dan debat tentang isu-isu politik termasuk pemilu.

Salah satu cara untuk memahami dinamika opini publik selama pemilihan umum adalah dengan melakukan analisis sentimen terhadap cuitan-cuitan yang diposting di media sosial, terutama yang menggunakan tagar (#) yang terkait dengan pemilu tersebut. Analisis sentimen adalah proses menilai pendapat, emosi, dan sikap yang diungkapkan dalam teks. Dalam konteks Twitter, analisis sentimen dapat membantu memahami bagaimana perasaan dan pandangan pengguna terhadap berbagai elemen pemilu, termasuk kandidat, partai politik, dan kebijakan. Untuk melakukan analisis ini, berbagai algoritma pembelajaran mesin dapat digunakan. Salah satu algoritma yang sering digunakan dalam analisis sentimen adalah Naïve Bayes.

Dalam konteks Indonesia, dengan mendekati Pemilu 2024, tagar #Pemilu2024 akan menjadi penting untuk dipelajari karena akan mencerminkan beragam pandangan dan sikap masyarakat terhadap proses pemilihan dan calon pemimpin. Pemilihan algoritma yang tepat sangatlah krusial karena akan berpengaruh pada akurasi dan kehandalan hasil analisis tersebut.

Naïve Bayes adalah algoritma berbasis probabilitas yang sering digunakan dalam klasifikasi teks. Naïve Bayes menggunakan teorema Bayes dengan asumsi bahwa fitur-fitur dalam data bersifat independen. Meskipun asumsi ini jarang sepenuhnya benar dalam kenyataan, Naïve Bayes sering kali memberikan hasil yang baik dalam analisis sentimen dan memiliki keunggulan dalam hal kecepatan dan efisiensi.

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap tweet-tweet yang berkaitan dengan Pemilu 2024 di Indonesia, dengan menggunakan algoritma Naïve Bayes. Data yang diolah berjumlah 1500 tweets. Hasil dari Penelitian ini menunjukkan bahwa klasifikasi data twitter dengan algoritma Naive Baiyes menghasilkan akurasi sebesar 80%[2].

II. PENELITIAN YANG TERKAIT

Penelitian terkait analisis sentimen telah banyak dilakukan oleh para peneliti dan menjadi topik yang sangat populer dalam dunia penelitian. Analisis sentimen digunakan untuk memahami pandangan dan opini publik terhadap berbagai isu, termasuk produk, layanan, dan peristiwa tertentu seperti pemilihan umum. Dengan kemajuan teknologi dan peningkatan penggunaan media sosial, analisis sentimen menjadi alat yang sangat penting dalam mengolah dan memahami data teks yang dihasilkan oleh pengguna secara real-time.

Pada penelitian kali ini, peneliti menggunakan metode Naïve Bayes sebagai teori utama untuk analisis sentimen. Naïve Bayes adalah salah satu algoritma yang sering digunakan dalam

klasifikasi teks karena beberapa alasan. Pertama, kesederhanaannya membuat algoritma ini mudah dipahami dan diterapkan. Kedua, Naïve Bayes mampu bekerja dengan baik pada dataset yang besar, yang seringkali menjadi tantangan dalam analisis data teks yang berasal dari media sosial seperti Twitter.

Prinsip utama dari Naïve Bayes adalah asumsi independensi antar fitur. Dalam konteks analisis teks, ini berarti setiap kata dalam sebuah dokumen dianggap independen dari kata lainnya. Meskipun asumsi ini jarang berlaku dalam situasi nyata (karena kata-kata dalam teks biasanya memiliki ketergantungan kontekstual satu sama lain), algoritma ini tetap memberikan hasil yang baik dalam banyak kasus praktis.

Naïve Bayes bekerja dengan cara menghitung probabilitas dari setiap kategori sentimen (positif, negatif, netral) berdasarkan kata-kata yang muncul dalam teks. Probabilitas ini kemudian digunakan untuk menentukan kategori sentimen dari teks baru yang akan diklasifikasikan. Dalam prosesnya, teks yang akan dianalisis harus melalui beberapa tahap preprocessing, seperti penghapusan noise, tokenisasi, stemming, dan penghapusan stop words, untuk memastikan data yang digunakan bersih dan relevan.

Dalam penelitian ini, data Twitter yang dikumpulkan selama periode kampanye Pemilu 2024 digunakan sebagai dataset. Data ini dikumpulkan menggunakan teknik crawling dan kemudian diproses untuk menghilangkan noise serta melakukan preprocessing teks. Setelah itu, algoritma Naïve Bayes diterapkan untuk mengklasifikasikan sentimen dari tweet-tweet yang telah dikumpulkan. Hasil analisis menunjukkan bahwa algoritma ini mampu mengklasifikasikan sentimen dengan baik, dengan akurasi mencapai sekitar 80% dalam pengujian dengan data uji yang terpisah.

Namun, meskipun Naïve Bayes memiliki banyak kelebihan, algoritma ini juga memiliki beberapa kelemahan. Salah satunya adalah asumsi independensi fitur yang tidak selalu sesuai dengan kenyataan dalam teks alami. Selain itu, Naïve Bayes juga bisa menjadi kurang efektif jika terdapat kata-kata yang sangat jarang muncul tetapi memiliki makna penting. Oleh karena itu, penting untuk mempertimbangkan kebutuhan spesifik dari analisis yang dilakukan, seperti kecepatan pemrosesan dan ketersediaan sumber daya, saat memilih algoritma yang paling sesuai.

Penelitian ini menyimpulkan bahwa meskipun Naïve Bayes memberikan hasil yang cukup akurat, eksplorasi lebih lanjut terhadap metode hybrid dan teknik deep learning disarankan untuk meningkatkan akurasi dan efisiensi analisis sentimen di platform media sosial. Dengan terus berkembangnya teknologi dan metode analisis, diharapkan hasil yang lebih baik dan lebih akurat dapat dicapai di masa mendatang.

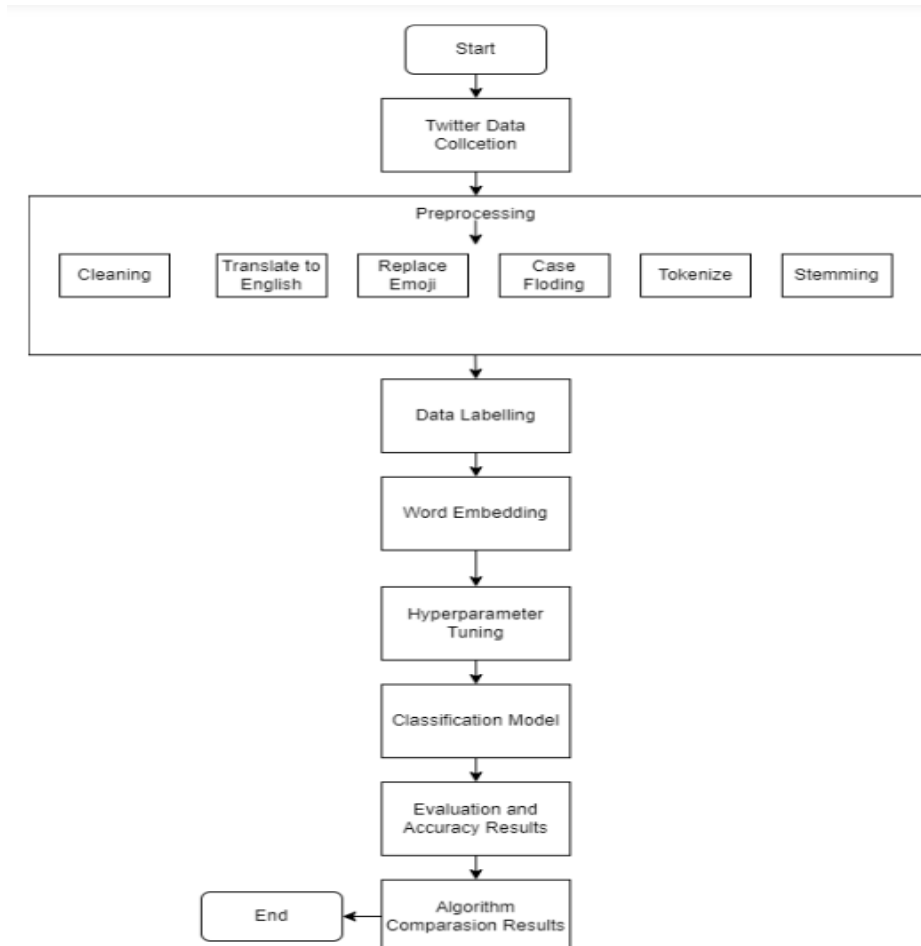
Penelitian oleh Kristiyanti D, Umam AWahyudi M, Amin R, Marlinda L[3] menunjukkan bahwa Naïve Bayes dapat digunakan secara efektif untuk analisa sentimen. Mereka menemukan bahwa meskipun asumsi independensi sering tidak realistis dalam praktek, algoritma ini tetap memberikan hasil yang baik.

III. METODE PENELITIAN

A. Tahap Penelitian

Penelitian ini diawali dengan tahap pengumpulan data menggunakan library Python 'sncrape' untuk mengakses API Twitter. Data yang terkumpul merupakan opini publik terkait calon presiden Indonesia untuk pemilihan tahun 2024. Kata kunci yang digunakan dalam proses pengambilan data ini mencakup "Prabowo", "Ganjar", dan "Anies". Dengan memanfaatkan Google Colab sebagai platform, kami berhasil mengumpulkan sejumlah 1500 data tweet. Penggunaan library Python 'sncrape'

memungkinkan akses API Twitter untuk mengumpulkan tweet berdasarkan filter yang ditentukan. Data yang terkumpul mencakup opini publik terkait calon presiden Indonesia untuk Pemilu 2024. Dengan memanfaatkan Google Colab sebagai platform, sejumlah 1500 data tweet berhasil dikumpulkan menggunakan kata kunci "Prabowo", "Ganjar", dan "Anies". Tahapan penelitian selanjutnya dijelaskan pada Gambar 1.



Gambar 1. Flowchart Metodologi Penelitian

B. Pre-Processing

preprocessing data merupakan langkah penting dalam analisis data teks. Preprocessing digunakan untuk menghindari dataset yang kurang sempurna[4]. Preprocessing yang telah dilakukan pada data tweet kali ini yaitu penghapusan duplikasi dan nilai kosong, seperti menghapus tweet yang duplikat untuk memastikan setiap tweet unik dan Menghapus baris dengan nilai kosong untuk menghindari kesalahan dalam analisis. Lalu pembersihan text "cleantext" untuk menyaring data yang tidak relevan atau tidak akurat. Setelah itu, Tahapan tersebut disebut dengan case folding dimana case folding ini digunakan untuk mengkonversi alfabet ke huruf kecil atau lower case [5]. Selanjutnya, proses normalisasi yaitu mengubah kata-kata tidak baku menjadi bentuk baku menggunakan kamus substitusi (misalnya, "yg" menjadi "yang", "engga" menjadi "tidak"). Kemudian ada Penghapusan Stop Word yaitu menghapus kata-kata umum yang tidak memberikan informasi penting (misalnya, "dan", "yang", "di"). Selanjutnya, Stemming merupakan proses yang digunakan untuk menemukan kata dasar dari sebuah kata

dengan menghilangkan imbuhan yang terdiri dari awalan, sisipan, akhiran dan kombinasi dari awalan dan akhiran [6], menggunakan library sastrawi. Kemudian Tokenisasi dan Penerjemahan "tokenization" akan memecah kalimat menjadi kata - kata yang membentuknya, memisahkan kalimat dengan spasi atau tanda baca. Misalnya memisahkan teks menjadi kata kata individual dan menerjemahkan teks dari bahasa Indonesia ke bahasa Inggris menggunakan google translator untuk analisis lebih lanjut.

C. Naïve Bayes

Algoritma Naive Bayes adalah salah satu metode data mining yang termasuk kedalam sepuluh klasifikasi data mining yang paling populer diantara algoritma-algoritma lainnya. Persamaan Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset

yang diberikan. Naive Bayes menggunakan metode statistik sederhana yang mengacu pada teorema Bayes dengan mengasumsikan independensi antara fitur-fitur yang ada terhadap kelas tertentu, tanpa adanya keterkaitan antar fitur tersebut [7]. Teknik ini banyak digunakan untuk menyelesaikan banyak masalah, seperti prediksi ketidakhadiran[8], prediksi ketepatan waktu[9], mendeteksi virus komputer[10], dan penentuan kelayakan penerima kartu Jakarta pintar plus[11]. Dengan metode Naive Bayes terlebih dahulu mencari Nilai Probabilitas dan likelihood maksimum dari setiap atribut untuk masing-masing kelas. Persamaan dari teorema bayes:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik P(H|X)
- : Probabilitas Hipotesis H berdasarkan kondisi X
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

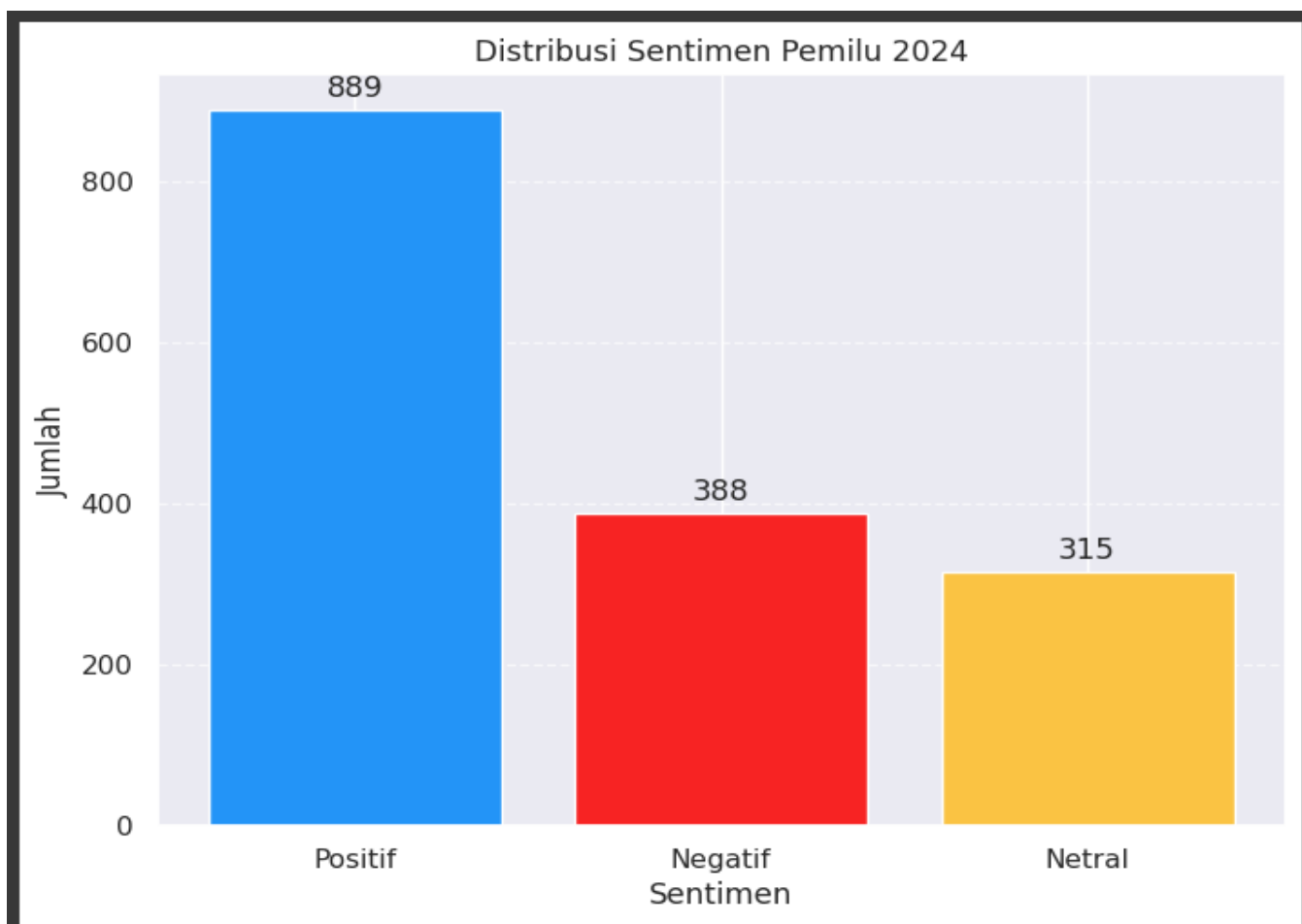
Penentuan class dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di class yang satu dengan nilai probabilitas suatu sampel berada di class yang lain. Untuk menentukan class yang cocok dari suatu sampel dilakukan dengan cara membandingkan nilai posterior untuk masing-masing class, dan mengambil class dengan nilai posterior yang tertinggi. Nilai evidence selalu tetap untuk setiap kelas pada satu

sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai prosterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dengan menggunakan $c|x_1, \dots, x_n$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) \quad (2) \\ &= P(C)P(X_1|c)(X_2, \dots, X_n|C, X_1) \\ &= P(C)P(X_1|c)P(X_2|C, X_1)(X_3, \dots, X_n|C, X_1, X_2) \\ &= P(C)P(X_1|c) P(X_2|C, X_1)P(X_2|C, X_1, X_2) \dots \\ &P(X_n|C, X_1, X_2, \dots, X_{n-1}) \end{aligned}$$

IV. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan menggunakan data dari twitter dengan tagar (#) pemilu2024 dan mendapatkan 1500 data. Dari data yang terkumpul dibagi menjadi tiga kelas yaitu positif, negatif dan netral. Data yang digunakan menggunakan ulasan berbahasa Indonesia mengenai pendapat, kritik maupun saran dari pengguna twitter pada pemilu 2024. Setelah itu untuk melakukan membersihkan data yang tidak penting, simbol, angka, tanda baca digunakan proses preprocessing seperti menghapus duplikat dan nilai kosong, lalu pembersihan text atau cleningtext, casefolding, penghapusan stop word, stemming dan tokenization. Kemudian hasil yang didapatkan yaitu 889 kelas positif, 388 kelas negatif dan 315 diidentifikasi sebagai kelas netral.



Gambar 2. Grafik Batang

- [6] A. Guterres, Gunawan, and J. Santoso, “Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based,” *Teknika*, vol. 8, no. 2, pp. 142–147, Oct. 2019, doi: 10.34148/teknika.v8i2.224.
- [7] R. Y. H. and S. N. M. R. Sari, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website,” *IJSE-Indonesian Journal on Software Engineering*, vol. 5, no. 2, pp. 51–60, 2016.
- [8] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, “Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 578, Dec. 2021, doi: 10.32493/informatika.v5i4.7575.
- [9] R. G. and S. P. J. Nata Permana, “Comparison Of C4.5 Algorithm and Naïve Bayes for Prediction Of Student Study Timeliness (Case Study: Departement of Statistics Mulawarman University),” *“Comparison Of C4.5 Algorithm and Naïve B EKSPONENSIAL*, vol. 13, no. 2, pp. 161–170, *ays for Prediction Of Student Study Timeliness (Case Study: Departement of Statistics Mulawarman University)*, 2022.
- [10] J. Elektronik et al, “Implementasi Algoritma Naive Bayes Classifier (NBC) Dan Information Gain Untuk Mendeteksi DDoS,” *vol. 11, no. 2, pp. 273–282*, 2022.
- [11] E. Widya Ningsih, “Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus,” 2019.