



Analisis Sentimen Terhadap Penggunaan Chatgpt Berdasarka Twitter Menggunakan Algoritma Naïve Bayes

Dwi Transiska, Dimas Febriawan, Firman Noor Hasan*

Fakultas Teknik, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹dtrsiskaa@gmail.com, ²dimas.febriawan@uhamka.ac.id, ³*firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Abstrak—Chatbot dapat membantu dengan bekerja sama antara pengguna dan perusahaan atau organisasi. Pada masa sekarang, kecerdasan buatan melalui chatbot telah menjadi salah satu andalan dalam menangani berbagai masalah. Salah satu jenis chatbot yang terkenal adalah ChatGPT, yang memanfaatkan teknologi NLP (Natural Language Processing) untuk memahami dan menjawab pertanyaan dan permintaan pengguna. ChatGPT sebagai perangkat yang sangat mudah digunakan untuk semua kalangan, tampilan interface ChatGPT cukup sederhana namun tidak menimbulkan rasa bosan, dan kecepatan dalam merespon perintah yang diberikan, hal ini menjadi nilai tambah ChatGPT. Meskipun dengan segudang kemudahan yang ditawarkan, ChatGPT juga menimbulkan kekhawatiran pada sisi negatifnya. Sisi negatifnya yaitu banyak kekhawatiran yang muncul mulai dari maraknya penyebaran hoaks dan kesalahpahaman di media sosial. Kelebihan serta kekurangan yang telah di jelaskan di atas, peneliti mendapatkan dorongan untuk mengetahui kebenarannya dari respon masyarakat terkait ChatGPT lebih dalam sehingga di buatlah penelitian analisis sentimen ini dibuat. Apalagi penelitian terkait analisis sentimen ini bisa di bilang cukup menjadi jawaban atas simpang siur nya respon masyarakat di luaran terkait ChatGPT. Penelitian ini juga di mulai dari proses Pengambilan data pada media sosial Twitter menggunakan Rapidminer, pada proses ini peneliti menggunakan token API Twitter pada aplikasi Rapidminer sehingga dapat di peroleh. Data yang telah diperoleh kemudian dibersihkan melalui proses preprocessing dengan menggunakan fitur fitur yang tersedia pada Rapidminer, hasil dari proses inilah data menjadi bersih. Setelah di bersihkan melalui preprocessing selanjutnya diberi pelabelan yaitu berlabel positif atau berlabel negatif yang nantinya akan diklasifikasikan oleh algoritma Naïve Bayes. Klasifikasi ini bertujuan untuk membagi antara sentimen positif dan sentimen negatif. Setelah melakukan klasifikasi, data kemudian dievaluasi menggunakan confusion matrix dan didapatkan hasil nilai accuracy sebesar 96,55%, nilai precision sebesar 89,19%, dan nilai recall sebanyak 95,18%.

Kata Kunci: Algoritma; Analisis Sentimen; ChatGPT; Naïve Bayes; Twitter

Abstract—Chatbots can assist by fostering collaboration between users and companies or organizations. In today's world, artificial intelligence through chatbots has become one of the mainstays in handling various problems. One of the well-known types of chatbots is ChatGPT, which utilizes NLP (Natural Language Processing) technology to understand and respond to user queries and requests. ChatGPT, as a device that is very easy to use for all circles, has a fairly simple interface but does not cause boredom, and the speed in responding to commands given, this is an added value of ChatGPT. Despite the myriad of conveniences offered, ChatGPT also raises concerns on the negative side. The negative side is that there are many concerns that arise, starting from the rampant spread of hoaxes and misunderstandings on social media. The advantages and disadvantages that have been explained above, researchers are encouraged to find out the truth from the public's response regarding ChatGPT more deeply so that this sentiment analysis research is made. Moreover, research related to sentiment analysis can be said to be quite an answer to the confusion of public responses outside related to ChatGPT. This research also starts from the process of Data retrieval on Twitter social media using Rapidminer, in this process the researcher uses the Twitter API token on the Rapidminer application so that it can be obtained. The data that has been obtained is then cleaned through the preprocessing process using the features available in Rapidminer, the result of this process is that the data becomes clean. After being cleaned through preprocessing, it is then labeled as positive or negative which will later be classified by the Naïve Bayes algorithm. This classification aims to divide between positive sentiment and negative sentiment. After performing classification, the data is then evaluated using a confusion matrix and the results are obtained with an accuracy value of 96.55%, a precision value of 89.19%, and a recall value of 95.18%.

Keywords: Algorithm; ChatGPT; Naïve Bayes; Sentiment Analysis; Twitter

1. PENDAHULUAN

Perkembangan zaman yang pesat ini dapat memberikan dampak positif dan negatif pada media sosial seperti Twitter [1]. Twitter adalah sebuah platform media sosial yang dibuat untuk berbagi pesan singkat dengan sebutan "tweet". Twitter ini memberikan peluang bagi pengguna untuk menerima atau mengirim informasi secara instan, serta berkomunikasi dengan pengguna lain menggunakan link, gambar, video, dan hastag. Banyak pengguna Twitter memberikan pendapat setelah berinteraksi dengan ChatGPT, baik positif maupun negatif, menyampaikan pengalaman mereka.

Seiring berjalannya waktu penggunaan chatbot menjadi populer dikalangan netizen dunia. Chatbot memiliki kemampuan untuk membantu pekerjaan manusia dalam sebuah perusahaan maupun dalam aktivitas sehari-hari [2]. Pada masa kini, chatbot telah berkembang menjadi lebih kompleks dengan disertai kecerdasan buatan. Hal ini menyebabkan chatbot menjadi alternatif yang banyak digunakan netizen untuk membantu pekerjaan sehari-hari. ChatGPT hadir sebagai chatbot yang paling mutakhir pada awal tahun 2018, dan sangat populer digunakan pada tahun 2020 hingga kini, dengan memanfaatkan teknologi NLP (Natural Language Processing), ChatGPT mampu memahami dan memenuhi permintaan yang diminta oleh pengguna [3].

Sebagian masyarakat memandang ChatGPT sebagai perangkat yang sangat mudah digunakan untuk semua kalangan, tampilan interface ChatGPT cukup sederhana namun tidak menimbulkan rasa bosan, dan kecepatan



dalam merespon perintah yang diberikan, hal ini menjadi nilai tambah ChatGPT [4]. Meskipun dengan segudang kemudahan yang ditawarkan, ChatGPT juga menimbulkan kekhawatiran pada sisi negatifnya. Banyak kekhawatiran yang muncul mulai dari maraknya penyebaran hoaks dan kesalahpahaman di media sosial. Hingga kini belum ada eksplorasi khusus mengenai seberapa banyak informasi ChatGPT yang bisa dipercaya dan sudah tervalidasi. Hal ini tentu menimbulkan permasalahan yang cukup serius jika diterima sebagai fakta tanpa mencari tahu kebenarannya.

Hal ini sebenarnya dapat diminimalisir dengan lebih meningkatkan kemampuan berpikir kritis dengan cara membaca informasi dari multisumber serta membandingkannya dan memilih informasi terbaik. Tidak sedikit juga orang yang merasa bahwa ChatGPT sangat berguna, seperti menyelesaikan pekerjaan lebih cepat, lebih efisien, mudah digunakan, dan dapat diakses dimana saja, menjadikan ChatGPT lebih fleksibel. Namun, ada sebagian masyarakat yang merasa khawatir dengan kehadiran ChatGPT yang dapat menggantikan manusia dalam hal pekerjaan. Isu ini bermula karena salah satu perusahaan amerika serikat mengatakan bahwa kecerdasan buatan manusia mampu mengambil alih pekerjaan manusia itu sendiri. Hal ini bersumber dari hasil survei yang dirilis oleh platform pekerjaan yaitu [resumebuilder.com](https://www.resumebuilder.com). karena isu ini membuat para pekerja di indonesia khawatir kehilangan pekerjaannya jika kecerdasan buatan diterapkan di perusahaan indonesia.

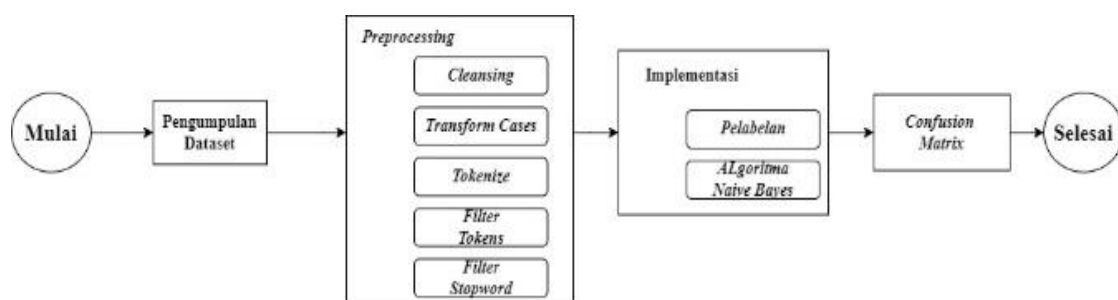
Analisis sentimen membantu mengidentifikasi apakah suatu komentar atau opini cenderung memiliki sentimen negatif atau positif. Hal ini berguna sebagai panduan untuk meningkatkan layanan atau kualitas produk. Pengumpulan data dilakukan menggunakan aplikasi Rapidminer dengan memasukan token API Twitter kemudian mencari tweet dengan memasukan keyword ChatGPT. Pemodelan dalam penelitian ini memanfaatkan algoritma Naïve Bayes sebagai algoritma untuk mengidentifikasi nilai sentimen yang terkandung dalam data [5]. Penelitian ini bertujuan untuk mengetahui opini publik terkait ChatGPT berdasarkan sentimen yang dikandung dalam opini tersebut. Nilai yang ada dalam setiap sentimennya akan diklasifikasikan oleh algoritma Naïve Bayes kedalam 2 kelas sentimen yaitu positif dan negatif berdasarkan data yang diambil dari Twitter [6].

Berdasarkan beberapa penelitian sebelumnya juga yang membahas tentang analisis sentimen pada pengguna Twitter terhadap ChatGPT oleh Yuma Akbar dan Tri Sugiharto. Diperoleh 1607 data setelah itu dibagi menjadi data uji dan data latih menggunakan algoritma C.45 dan Naïve Bayes dengan rasio 70/30. Hasil klasifikasi pada penelitian ini menghasilkan nilai akurasi sebesar 77.33% [7]. Selanjutnya penelitian yang membahas tentang respon masyarakat tentang berbelanja baju thrifting pada Twitter oleh Sania Wulandari dan Firman Noor Hasan. Dengan data awal sebanyak 900 dan diperoleh 368 sentimen positif dan 181 sentimen negatif serta menggunakan algoritma Naïve Bayes dengan rasio 60/40. Hasil klasifikasi pada penelitian ini menghasilkan nilai akurasi sebesar 95.92% [8]. Setelah itu Penelitian lain tentang ulasan yang diberikan pelanggan pada salah satu aplikasi coffee menggunakan algoritma naive bayes oleh Tia Anggita Sari, Estu Sinduningrum dan Firman Noor Hasan, dengan data sebanyak 1818 data ulasan dengan nilai akurasi sebesar 74,28% [9]. Penelitian yang terkait dengan aplikasi tiktok dengan algoritma Naïve Bayes dan SVM pada Google Play Store dengan perbandingan 80:20, dengan nilai akurasi sebesar 79% [10]. Dan yang terakhir yaitu penelitian terkait KPU 2024 dengan data dari Twitter sebanyak 300 data yang telah bersih, sebanyak 210 data yang termasuk sentiment positif dan 90 data yang negatif serta akurasi sebesar 67,13% [11].

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian ini menggunakan data yang diperoleh dari Twitter dengan menggunakan tools Rapidminer sebagai penarik data dan menggunakan algoritma Naïve Bayes yang berfungsi untuk mengklasifikasi nilai sentimen yang terkandung dalam data. Berikut gambar 1 merupakan alur penelitian yang dilalui oleh peneliti dalam penelitian kali ini.



Gambar 1. Alur Penelitian

Berikut penjelasan dari alur penelitian yang ada pada gambar 1:

a. Pengumpulan data

Peneliti menggunakan Rapidminer sebagai tools untuk penarikan data pada tahap ini. Pada tahapan ini juga data yang peneliti tarik merupakan data ulasan penggunaan ChatGpt yang berasal dari Twitter.



b. Preprocessing

Persiapan data atau Preprocessing merupakan tahapan untuk mempersiapkan data dengan fitur-fitur yang tersedia pada Rapidminer. Fitur-fitur ini yang akan dipakai serta bertujuan untuk membersihkan elemen yang tidak dibutuhkan dalam sentimen yang akan digunakan. Sehingga sentimen yang akan digunakan pada proses ini siap untuk proses selanjutnya.

c. Pelabelan data

Pelabelan data merupakan tahapan untuk menentukan kelas atau label dari sentimen yang akan dipakai. Dan pada penelitian ini digunakan pelabelan yang dilakukan secara manual menggunakan Microsoft Excel.

d. Algoritma Naïve Bayes

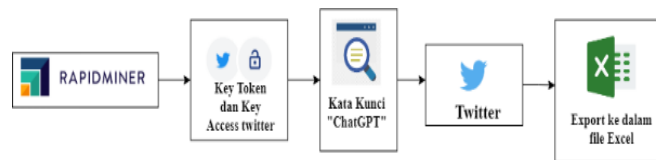
Setelah tahap pelabelan selesai dilakukan, data kemudian dibagi menjadi dua kelompok yaitu data latih dan data uji. Lalu setelah itu diimplementasikan algoritma Naïve Bayes untuk pengklasifikasi nilai sentimen.

e. Confusion matrix

Setelah implementasi algoritma Naïve Bayes dilakukan, kemudian dilakukan evaluasi terhadap model yang digunakan, tahapan ini menggunakan Confusion matrix sebagai fitur evaluasinya. Serta bertujuan untuk mengevaluasi kinerja dari model yang telah dipakai.

2.2 Pengumpulan Data

Gambar 2 pada tahapan ini peneliti akan mengumpulkan dataset dari Twitter. Ilustrasi pengumpulannya sebagai berikut:



Gambar 2. Proses Pengumpulan Data.

Pertama-tama langkah yang harus dilakukan adalah membuat akun Twitter Developer yang berfungsi untuk mengirim permintaan akses Key Token ke Twitter. Selanjutnya peneliti akan mendapatkan Key Token dan Key Access yang berfungsi sebagai Application Programming Interface (API). Kemudian peneliti akan mengambil data berupa tweet masyarakat dengan kata kunci pencarian "ChatGPT", diambil menggunakan tool Rapidminer. Setelah itu dikumpulkan dalam bentuk Spreadsheet Excel. Pengumpulan data ini dilakukan dengan tanggal 13 juni 2023 sampai 30 juni 2023 Data yang dikumpulkan sebanyak 300 data.

2.3 Preprocessing

Setelah melakukan pengumpulan data, proses selanjutnya yaitu preprocessing data. Proses ini memiliki beberapa tahapan, yaitu cleansing, transform cases, tokenize, Filter stopwords, Filter token [12]. Berikut merupakan penjelasan dari setiap tahapan dalam proses preprocessing.

- Cleansing adalah proses penghilangan atau penghapusan elemen yang tidak penting dalam data. Contohnya seperti hastag(#), mention(@xxxx), emoticon dan lain sebagainya [13].
- Transform Case adalah proses yang dilakukan dengan tujuan untuk mengubah format menjadi huruf kecil pada setiap huruf yang ada dalam dataset. Hal ini dilakukan untuk menjaga konsistensi format data yang akan digunakan dalam penelitian ini [14].
- Tokenizing adalah proses pemisahan kalimat menjadi unit kecil yang lebih sederhana [15].
- Filter token adalah proses penghapusan kata yang memiliki karakter terlalu sedikit ataupun terlalu banyak, pada proses ini peneliti menggunakan angka 4 sebagai batas minimum karakter dan 25 sebagai batas maksimumnya. Proses ini bertujuan untuk meningkatkan kualitas kalimat [16].
- Filter stopwords yaitu adalah proses untuk menghilangkan kata yang tidak mengandung nilai sentiment contoh "dan", "di", "tetapi", "atau" [17].

2.4 Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu algoritma yang banyak digunakan dalam proses teks mining. Algoritma Naïve Bayes mempunyai ciri dengan asumsi dasar yang sangat simpel, yaitu kemandirian antar setiap peristiwa atau kondisi [18]. Pada penelitian ini peneliti menggunakan algoritma Naïve Bayes dikarenakan Naïve Bayes memiliki kestabilan akurasi yang sangat tinggi, sehingga tingkat kemampuan untuk mengklasifikasikan suatu dataset melebihi rata-rata kemampuan yang dimiliki oleh algoritma lainnya. Selain itu algoritma Naïve Bayes juga mudah digunakan dalam penelitian dengan topik analisis sentiment [19].

2.5 Confusion Matrix

Confusion matrix adalah metode evaluasi model yang berguna untuk mengetahui nilai accuracy, precision, recall pada suatu model [20].

- Accuracy adalah tingkat keakuratan suatu prediksi dalam suatu model



- b. Precision adalah nilai total ketepatan hasil klasifikasi secara keseluruhan
- c. Recall adalah tingkat keberhasilan suatu model dalam mengidentifikasi informasi

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

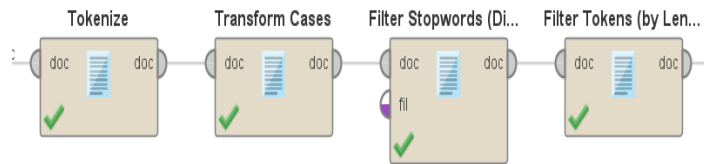
$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

Peneliti menggunakan 300 data tweet netizen Twitter yang akan melalui tahap preprocessing seperti terlihat pada gambar 3. Peneliti menggunakan tools Rapidminer sebagai alat untuk melakukan tahap preprocessing ini. Tahap Preprocessing ini akan menggunakan beberapa operator yaitu, cleansing, tranform cases, tokenize, Filter stopwords, Filter token [21].



Gambar 3. Preprocessing menggunakan Rapidminer

3.1.1 Cleansing

Tahap pertama dari preprocessing adalah tahap cleansing, tahap ini data akan dibersihkan dan menambahkan nilai yang hilang [22]. Setelah proses pembersihan, data terbebas dari elemen-elemen yang tidak diperlukan. Hasil dari proses cleansing dapat dilihat pada tabel 1 berikut:

Tabel 1. Ilustrasi Cleansing

Sebelum	Sesudah	Sentimen
Ngebantu banget gua ngerjain makalah dan cek plagiat di ChatGPT,,	Aku waktu ngerjain tugas makalah dan plagiat aku pakeknya ChatGPT dong dan ternyata ngebantu banget	Positif
Pernah coba ChatGPT buat bikin pantun Bahasa inggris dan kureng kata katanya ☹️ ChatGPT ngebantu banget buat ngerjain Tugas, Jadi ngebantu banget deh buat ngerjain tugas!!	Pernah coba ChatGPT buat bikin pantun Bahasa inggris dan hasilnya kata katanya masih kurang ChatGPT ngebantu banget buat ngerjain Tugas Jadi ngebantu banget deh buat ngerjain tugas	Negatif
		Positif

3.1.2 Transform Case

Tahap kedua pada proses preprocessing yaitu Transform Case. Tahap ini berfungsi untuk mengubah format seluruh karakter huruf pada dataset menjadi huruf kecil [23]. Tahap ini penting dilakukan agar semua format data terjaga konsistensinya.

Tabel 2. Ilustrasi Transform Case

Sebelum	Sesudah	Sentimen
Ngebantu banget gua ngerjain makalah dan cek plagiat di ChatGPT	aku waktu ngerjain tugas makalah dan plagiat aku pakeknya chatgpt dong dan ternyata ngebantu banget	Positif
Pernah coba ChatGPT buat bikin pantun Bahasa inggris dan kureng kata katanya ChatGPT ngebantu banget buat ngerjain Tugas Jadi ngebantu banget deh buat ngerjain tugas	pernah coba chatgpt buat bikin pantun bahasa inggris dan hasilnya kata katanya masih kurang chatgpt ngebantu banget buat ngerjain tugas jadi ngebantu banget deh buat ngerjain tugas	Negatif
		Positif

Tabel 2 memperlihatkan perbedaan saat terjadi perubahan sebelum dan sesudah proses Transform Cases. Terlihat beberapa huruf kapital pada kolom sebelum, kemudian setelah melalui tahap Transform Cases, huruf kapital di kolom kanan berubah menjadi huruf kecil.



3.1.3 Tokenizing

Tahap ketiga dari preprocessing yaitu tokenizing. Tahap ini berfungsi untuk memecah kalimat menjadi potongan – potongan kata atau unit Bahasa kecil, agar mudah di lakukan pengolahan kata dan analisis [24]. Berikut ini hasil dari tokenizing:

Tabel 3. Ilustrasi Tokenizing.

Sebelum	Sesudah	Sentimen
aku waktu ngerjain tugas makalah dan plagiat aku pakeknya chatgpt dong dan ternyata ngebanu banget	['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dong', 'dan', 'ternyata', 'ngebanu', 'banget']	Positif
pernah coba chatgpt buat bikin pantun bahasa inggris dan hasilnya kata katanya masih kurang	['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	Negatif
chatgpt ngebanu banget buat ngerjain tugas jadi ngebanu banget deh buat ngerjain tugas	['chatgpt', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebanu', 'banget', 'deh', 'buat', 'ngerjain', 'tugas']	Positif

Dapat dilihat perbedaan pada tabel 3, pada kolom sebelumnya masih berbentuk sebuah kalimat karena belum melalui tahap tokenizing. Sedangkan pada kolom sebelahnya yaitu kolom sesudah, data sudah menjadi bagian-bagian kecil atau telah terbagi menjadi unit-unit yang lebih sederhana untuk memudahkan dalam proses penelitian.

3.1.4 Filter token

Tahap keempat dari preprocessing yaitu Filter token. Tahap ini bertujuan untuk menghilangkan kata yang memiliki jumlah karakter terlalu sedikit ataupun terlalu banyak. Pada tahap ini peneliti menggunakan parameter 4-25 yang berarti kata yang memiliki jumlah karakter dibawah 4 dan lebih dari 25 akan dihilangkan dari data, sehingga nilai informasi yang didapatkan lebih jelas karena tidak adanya kata yang tidak memiliki nilai informatif. Berikut tabel 4 merupakan hasil dari tahap Filter token:

Tabel 4. Ilustrasi Filter token

Sebelum	Sesudah	Sentimen
['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dong', 'dan', 'ternyata', 'ngebanu', 'banget']	['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dong', 'dan', 'ternyata', 'ngebanu', 'banget']	Positif
['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	Negatif
['chatgpt', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebanu', 'banget', 'deh', 'buat', 'ngerjain', 'tugas']	['chatgpt', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas']	Positif

Beberapa kata seperti 'dan' dan 'cek' telah dihapus dari data karena panjang karakternya tidak memenuhi parameter yang digunakan pada proses Filter token ini.

3.1.5 Filter stopwords

Tahap kelima dari preprocessing yaitu Filter stopwords. Tahap ini menghilangkan kata-kata yang tidak memiliki arti penting, seperti kata penghubung dan kata-kata yang tidak memiliki nilai sentimen pada. Sehingga kualitas kejelasan kalimat menjadi lebih baik. Tabel 5 berikut hasil dari Filter stopwords:

Tabel 5. Ilustrasi Filter stopwords

Sebelum	Sesudah	Sentimen
['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dong', 'dan', 'ternyata', 'ngebanu', 'banget']	['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dan', 'ternyata', 'ngebanu', 'banget']	Positif
['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	Negatif
['chatgpt', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebanu', 'banget', 'buat', 'ngerjain', 'tugas']	['chatgpt', 'ngebanu', 'banget', 'ngerjain', 'tugas', 'ngebanu', 'banget', 'ngerjain', 'tugas']	Positif



Terlihat perbedaan pada tabel 5 kolom sebelah kiri sebelum dilakukan Filter stopwords dan kolom sebelah kanan sesudah dilakukan Filter stopwords.

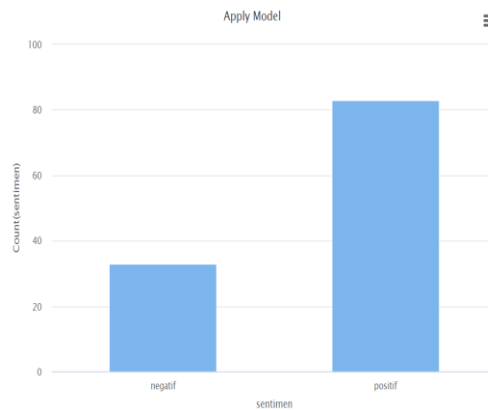
3.2 Pelabelan Data

Setelah melalui tahap preprocessing, peneliti lanjutkan penelitian ke tahap selanjutnya yaitu pelabelan data. Peneliti melakukan pelabelan data secara manual. Selanjutnya data yang telah dilabeli dibagi kedalam dua kategori kelas yaitu positif dan negatif [25]. Data bersih yang dimiliki berjumlah 116, lalu sebanyak 58 data yang di labelkan manual dan 58 data yang lainnya menggunakan Rapidminer.

Tabel 6. Contoh Pelabelan Data Secara Manual

Tweets	Sentimen
['aku', 'waktu', 'ngerjain', 'tugas', 'makalah', 'dan', 'plagiat', 'aku', 'pakeknya', 'chatgpt', 'dong', 'dan', 'ternyata', 'ngebantu', 'banget']	Positif
['pernah', 'coba', 'chatgpt', 'buat', 'bikin', 'pantun', 'bahasa', 'inggris', 'dan', 'hasilnya', 'kata', 'katanya', 'masih', 'kurang']	Negatif
['chatgpt', 'ngebantu', 'banget', 'ngerjain', 'tugas', 'ngebantu', 'banget', 'ngerjain', 'tugas']	Positif

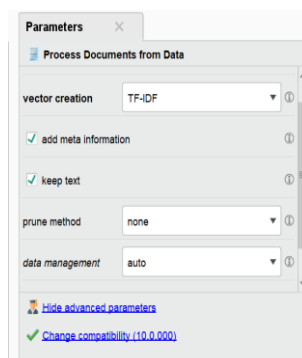
Tabel 6 memperlihatkan contoh data yang telah dilabeli. Setelah semua data telah terlabeli, maka didapatkan visualisasi perbandingan antara jumlah kelas sentimen positif dan kelas sentimen negatif. Gambar 4 berikut merupakan visualisasi perbandingan antara kedua kelas tersebut.



Gambar 4. Visualisasi perbandingan sentimen positif dan negatif.

3.3 TF-IDF

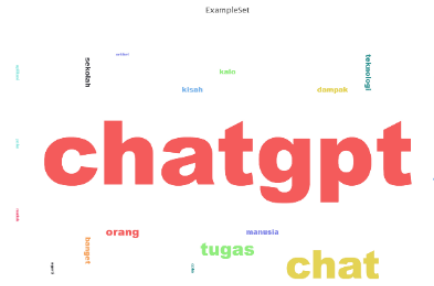
TF-IDF merupakan metode statistik yang digunakan untuk mengukur tingkat signifikansi suatu kata dalam sebuah dokumen [26]. Tahap ini dilakukan untuk memberikan nilai bobot kepada kata-kata yang ada pada tweet. Dalam menerapkan pembobotan TF-IDF, peneliti menggunakan operator proses dokumen yang mengintegrasikan operator dari tahap preprocessing, meliputi Transform Cases, tokenize, Filter token, dan Filter stopwords. Selanjutnya, peneliti menentukan parameter yang sesuai dan mengatur bagian pembuatan vektor (vector creation) ke TF-IDF, seperti yang terlihat pada gambar 5.



Gambar 5. TF-IDF

3.4 Wordcloud

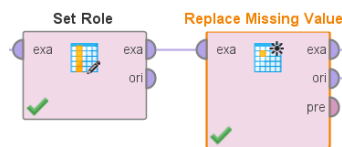
Gambar 6 menunjukkan hasil visualisasi data menggunakan wordcloud. Hasil ini menunjukkan bahwa semakin besar total kata pada dokumen maka akan semakin besar juga kata yang ditampilkan. Adapun pada penelitian ini dapat diketahui bahwa terdapat 20 kata.



Gambar 6. Visualisasi kata dengan worldcloud

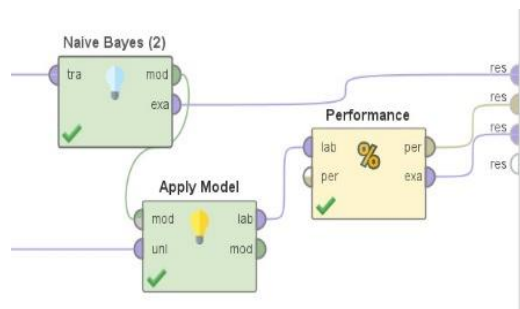
3.5 Algoritma Naïve Bayes

Sebelum memasuki tahap implementasi algoritma Naïve Bayes, dataset yang telah dilakukan pelabelan dimasukan ke dalam operator set role kemudian dimasukan ke dalam operator replace missing values. Fungsi dari kedua operator ini adalah untuk menentukan peran atribut atau kolom dalam analisis dan mencegah kolom yang kosong atau missing values. Kedua operator ini digunakan dengan tujuan untuk mengolah data agar sentimen pada data tersebut seimbang.



Gambar 7. Operator set role dan replace missing values

Setelah melewati kedua operator diatas seperti gambar 7, dataset baru dapat memasuki tahap implementasi Naïve Bayes . Setelah pengisian label kosong menggunakan operator replace missing values selanjutnya disambungkan dengan operator apply model dan operator algoritma Naïve Bayes. Selanjutnya untuk mengetahui nilai confusion matrix pada pengimplementasian algoritma Naïve Bayes maka peneliti menggunakan operator performance. Berikut merupakan proses dari tahap implementasi algoritma Naïve Bayes.



Gambar 8. Implementasi algoritma Naïve Bayes

Gambar 8 merupakan tahapan menggunakan perbandingan 80: 20 untuk data latih dan data uji nya. Data latih di gunakan untuk melatih model dan data uji di gunakan untuk menguji kinerja model. Klasifikasi Naïve Bayes dilakukan untuk analisis sentimen. Performance pada tahapan ini bekerja untuk mengevaluasi kinerja dari model sampai perbaikan kinerja model. Hasil prediksi dari sentimen yang ada di dapatkan seperti pada tabel 7 berikut:

Tabel 7. Hasil Predik dari sentimen

Jenis Data	Positif	Negatif
Data latih dan	158	60
Data uji		

3.6 Confusion Matrix

Gambar 9 dapat terlihat Accuracy prediksi suatu model dalam mengklasifikasikan data, dan kurasi dihitung dengan membagi jumlah prediksi yang benar dengan total jumlah sampel [27]. Precision yaitu kalkulasi tingkat ketepatan hasil klasifikasi secara keseluruhan [28]. Recall adalah parameter yang mengindikasikan tingkat keberhasilan dalam mengidentifikasi kembali informasi secara akurat [29]. Tabel 8 ini adalah implementasi dari confusion matrix.



	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN)	Sensitivity $\frac{TP}{(TP + FN)}$
Negative	False Positive (FP)	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
	Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Gambar 9. Confusion matrix

Tabel 8. Confusion matrix

	True Positive	True Negative	Class Precision
Pred. Positive	33	4	89.19%
Pred. Negative	0	79	100%
Class Recal	100%	95.18%	

Berikut merupakan perhitungan dari tabel 8:

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{33}{33+4} = 0.89 = 89.19\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{33}{33+0} = 1 = 100\%$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{33+79}{33+79+4+0} = 0.9655 = 96.55\%$$

Perhitungan diatas menjelaskan bahwasanya dari 115 data yang digunakan pada penelitian ini menghasilkan 79 True Positive (TP), 4 False Positive (FP), 0 False Negative (FN), dan 33 True Negative (TN). Hasil akhir dari confusion matrix kali ini mendapatkan hasil precision sebesar 89,19%, recall sebesar 95,18%. Penelitian ini menghasilkan nilai accuracy sebesar 96,55%. Gambar 10 terlihat hasil dari confusion matrix model penelitian.

		Predicted Values	
		Pred. Positif	Pred. Negatif
Predicted Values	Pred. Positif	33	4
	Pred. Negatif	0	79

Gambar 10. Confusion matrix model penelitian

4. KESIMPULAN

Penelitian terkait analisis sentimen terhadap penggunaan ChatGPT berdasarkan Twitter menggunakan algoritma Naive Bayes ini dengan data awalnya berjumlah 300 data dengan topik ChatGPT pada rentan waktu pengambilan dari tanggal 13 juni 2023 sampai 30 juni 2023, dan diolah dengan beberapa fitur pada proses Preprocessing yang meliputi Transform Cases, tokenize, Filter token, dan Filter stopword yang ada pada Rapidminer. Penelitian ini juga menggunakan rasio perbandingan untuk data latih dan data ujinya sebesar 80:20. hasil dari prediksi ini juga didapatkan 158 data positif dan 60 data negatif. Itu artinya kehadiran ChatGPT menjadi hal yang mendapat respon positif di kalangan netizen twitter. Dalam segi implementasi algoritma Naive Bayes, peneliti mendapatkan hasil precision sebesar 89,19%, recall sebesar 95,18% dan nilai accuracy sebesar 96,55%. Nilai accuracy tersebut membuktikan bahwa algoritma Naive Bayes merupakan algoritma yang memiliki nilai akurasi yang paling baik bahkan mendekati sempurna. Hasil akurasi pada penerapan metode ini berdasarkan dari tweet user Twitter yang menggunakan ChatGPT sebagai kebutuhan sehari hari, jika di lihat berdasarkan pelabelan dan akurasi, sentimen positif lebih banyak dari pada sentimen negatif. Hal ini dikarenakan mayoritas pengguna ChatGPT merasa terbantu



dengan hadirnya chatbot yang mampu meringankan pekerjaan manusia sehari-hari. Oleh karena itu saran untuk masyarakat gunakanlah ChatGPT dengan bijak dan sesuai kebutuhan. Pengembangan penelitian ini dapat menggunakan metode klasifikasi lain untuk membandingkan hasil akurasi dengan berbagai metode klasifikasi lainnya dan data yang di ambil pun mungkin dapat lebih banyak.

REFERENCES

- [1] Uky Yudatama et al., *Memahami Teknologi Informasi*. 2023.
- [2] Y. Eka and P. Heri, "Implementasi Chatbot 'ALITTA' Asisten Virual Dari BALITTAS Sebagai Pusat Informasi Di BALITTAS," *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, vol. 12, no. 1, pp. 24–35, 2019, doi: <https://doi.org/10.35457/antivirus.v13i1.714>.
- [3] S. Alfandi and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *ZONASI: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [4] D. Atmajaya, A. Febrianti, and H. Darwis, "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indonesian Journal of Computer Science*, vol. 12, no. 4, pp. 2173–2181, 2023, doi: 10.33022/ijcs.v12i4.3341.
- [5] al as Hafizh Dhery and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.
- [6] Styawati, N. Andi, A. Zaenal, and S. Heni, "Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 904–910, Oct. 2021, doi: 10.29207/resti.v5i5.3380.
- [7] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naive Bayes," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 115–122, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v4i3.1368>
- [8] S. Wulandari and F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 768–776, 2024, doi: 10.30865/mib.v8i2.7520.
- [9] T. A. Sari, E. Sinduningrum, and F. Noor Hasan, "Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naive Bayes," *Klik : Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 773–779, 2023, doi: 10.30865/klik.v3i6.884.
- [10] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naive bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekn.v10i2.419.
- [11] D. P. Ray, F. N. Hasan, and A. R. Dzikrillah, "Analisis Sentimen Terhadap KPU 2024 Berdasarkan Tweet Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *Klik: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 2235–2243, 2024, doi: 10.30865/klik.v4i4.1587.
- [12] F. H. Noor and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier," *JLK*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [13] A. C. Khotimah and E. Utami, "Comparison Naive Bayes Classifier, K-Nearest Neighbor And Support Vector Machine In The Classification Of Individual On Twitter Account," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 3, pp. 673–680, 2022, doi: 10.20884/1.jutif.2022.3.3.254.
- [14] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 807–814, 2021, doi: <http://dx.doi.org/10.30645/j-sakti.v5i2.378>.
- [15] R. Ade, Z. N. Muhammad, and B. Auliya, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," *JLK*, vol. 2, no. 1, 2019, doi: <https://doi.org/10.26418/jlk.v2i1.17>.
- [16] A. Hendra, "Analisis Sentimen Review Halodoc Menggunakan Naive Bayes Classifier," *JISKA*, vol. 6, no. 2, pp. 78–89, 2021, doi: <https://doi.org/10.14421/jiska.2021.6.2.78-89>.
- [17] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, S. Khomsah, and A. S. Aribowo, "Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 648–654, 2020, doi: <https://doi.org/10.29207/resti.v4i4.2035>.
- [18] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *Journal of Information System Research*, vol. 4, no. 4, pp. 1136–1144, 2023, doi: 10.47065/josh.v4i4.3803.
- [19] L. Siliyani, I. Agis Junizar, U. Nuraeni, E. Tohidi, and I. Ali, "Penerapan Algoritma Naive Bayes Untuk Mengetahui Kepuasan Mahasiswa Terhadap Layanan Administrasi Keuangan," *Kopertip*, vol. 04, no. 03, pp. 72–78, 2020, doi: <https://doi.org/10.32485/kopertip.v4i3.122>.
- [20] A. Ridhovan et al., "Penerapan Metode Residual Network (RESNET) Dalam Klasifikasi Penyakit Pada Daun Gandum," *JIPI*, vol. 07, no. 01, pp. 58–65, 2022, doi: <https://doi.org/10.29100/jipi.v7i1.2410>.
- [21] Yusuf Wahyu Setiya Putra et al., *Pengantar Aplikasi Mobile*. 2023.
- [22] E. Andrian and A. Rahman Isnain, "Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 788–796, 2024, doi: 10.30865/mib.v8i2.7530.
- [23] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 4, p. 2022, 2022, doi: <https://doi.org/10.35870/jti>.
- [24] S. D. F. Kartika and A. T. Purnomo, "Klasifikasi Berita Menggunakan Metode Multinomial Naive Bayes," *SCAN*, vol. 16, no. 3, pp. 1–8, 2021, doi: <https://doi.org/10.33005/scan.v16i3.2870>.



- [25] A. F. Setya and F. N. Hasan, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store," *Jurnal ICT : Information Communication & Technology*, vol. 23, no. 1, pp. 75–80, 2023, [Online]. Available: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>
- [26] A. S. Jeremy, M. F. Tresna, and N. Aryo, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepkabolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, vol. 1, no. 1, pp. 43–49, 2021, doi: <https://doi.org/10.52985/insyst.v1i1.36>.
- [27] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *Jurnal Ilmiah Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: <https://doi.org/10.33557/jurnalmatrik.v22i1.843>.
- [28] A. Rahim and L. E. Taufiq, "Convolutional Neural Network untuk Kalasifikasi Penggunaan Masker," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 2, pp. 109–115, 2020, doi: <http://dx.doi.org/10.35585%2Finspir.v10i2.2569>.
- [29] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.