

dwi transiska - Analisis Sentimen Terhadap Penggunaan Chatgpt Menggunakan Algoritma Naïve Bayes

by Layanan Perpustakaan UHAMKA

Submission date: 23-Apr-2024 08:41AM (UTC+0700)

Submission ID: 2358754151

File name: cek_turnitin_-_Dwi_Transiska.docx (408.86K)

Word count: 3169

Character count: 20872



Analisis Sentimen Terhadap Penggunaan Chatgpt Menggunakan Algoritma Naïve Bayes

Dwi Transiska¹, Dimas Febriawan², Firman Noor Hasan^{3,*}

¹ Fakultas teknik, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

² Fakultas teknik, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

³ Fakultas teknik, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

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

Email Penulis Korespondensi: emailpenuliskorespondensi@email.com

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. Banyak kekhawatiran yang muncul mulai dari maraknya penyebaran hoaks dan kesalahpahaman di media sosial. Pengambilan data pada media sosial Twitter diproses menggunakan token API Twitter pada aplikasi Rapidminer. Data yang telah didapat kemudian dibersihkan melalui proses *preprocessing*. Setelah di bersihkan melalui *preprocessing* selanjutnya diberi pelabelan antara positif atau 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%.

19

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

Abstract— *Chatbots* can assist in facilitating collaboration between users and companies or organizations. Currently, artificial intelligence through chatbots has become a cornerstone in addressing various issues. One notable type of chatbot is ChatGPT, which utilizes Natural Language Processing (NLP) technology to understand and respond to user questions and requests. ChatGPT serves as a highly user-friendly tool for all demographics, featuring a simple yet engaging interface and prompt response to user commands, adding value to the user experience. Despite the numerous advantages offered, ChatGPT also raises concerns on the negative side. There are worries about the proliferation of misinformation and misunderstandings on social media. Data collection from the Twitter social media platform is processed using the Twitter API token in the Rapidminer application. The obtained data is then cleaned through a preprocessing process. After cleaning through preprocessing, it is labeled as either positive or negative sentiment, which will later be classified by the Naïve Bayes algorithm. This classification aims to distinguish between positive and negative sentiments. After classification, the data is then evaluated using a confusion matrix, resulting in an accuracy rate of 96.55%, a precision rate of 89.19%, and a recall rate 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 *hashtag*. 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*. karena isu ini membuat para pekerja di indonesia khawatir kehilangan pekerjaannya jika kecerdasan buatan diterapkan di perusahaan indonesia.

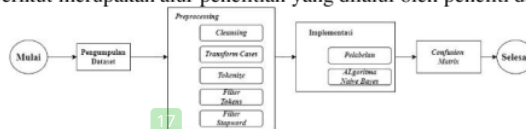
Analisis sentimen membantu mengidentifikasi apakah suatu komentar atau opini cenderung memiliki sentimen negatif atau positif. Hal ini berguna sebagai panduan untuk meningkatkan layanan atau kualitas produk. Pengumpulan data dilakukan menggunakan aplikasi Rapidminer dengan memasukan token API Twitter 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 penelitian sebelumnya 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].

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 merupakan alur penelitian yang dilalui oleh peneliti dalam penelitian kali ini.



Gambar 1. Alur Penelitian

Berikut penjelasan dari gambar 1:

- Pengumpulan data, peneliti menggunakan Rapidminer sebagai tools untuk menarik data dari Twitter.
- Preprocessing*, proses yang dilakukan dengan tujuan membersihkan elemen yang tidak dibutuhkan dalam penelitian.
- Pelabelan data, pelabelan data ini dilakukan secara manual pada *dataset* yang telah dibersihkan.
- Algoritma Naïve Bayes, Setelah pelabelan, data dibagi menjadi dua kelas yaitu data latih dan data uji. Lalu setelah itu diimplementasikan algoritma Naïve Bayes untuk pengklasifikasi nilai sentimen.
- Confusion matrix dilakukan sebagai bentuk evaluasi terhadap model yang digunakan dengan menggunakan algoritma Naïve Bayes.

2.2 Pengumpulan Data

Tahapan ini peneliti akan mengumpulkan *dataset* dari Twitter. Ilustrasi pengumpulannya sebagai berikut:





Gambar 2. Proses Pengumpulan Data.

Pertama-tama langkah yang harus di lakukan 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”, di ambil menggunakan tool Rapidminer. Setelah itu di kumpulkan dalam bentuk Spreadsheet Excel. Pengumpulan data ini di lakukan dengan dari tanggal 13 juni 2023 sampai 30 juni 2023 Data yang di kumpulkan sebanyak 300 data.

2.3 Preprocessing

Setelah melakukan pengumpulan data, proses selanjutnya yaitu *preprocessing* data. Proses ini memiliki beberapa tahapan, yaitu *cleansing*, *tranform cases*, *tokenize*, *Filter stopwords*, *Filter token* [8]. Berikut merupakan penjelasan dari setiap tahapan dalam proses *preprocessing*.

- a. *Cleansing* adalah proses penghilangan atau penghapusan elemen yang tidak penting dalam data. Contohnya seperti hastag(#), mention(@xxxx), emoticon dan lain sebagainya [9].
- b. *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[10].
- c. *Tokenizing* adalah proses pemisahan kalimat menjadi unit kecil yang lebih sederhana [11].
- d. *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 [12].
- e. *Filter stopword* yaitu adalah proses untuk menghilangkan kata yang tidak mengandung nilai sentiment contoh “dan”, “di”, “tetapi”, “atau” [13].

2.4 Algoritma Naive Bayes

Algoritma Naive Bayes adalah salah satu algoritma yang banyak digunakan dalam proses teks mining. Algoritma Naive Bayes mempunyai ciri dengan asumsi dasar yang sangat simpel, yaitu kemandirian antar setiap peristiwa atau kondisi [14]. Pada penelitian ini peneliti menggunakan algoritma Naive Bayes dikarenakan Naive 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 Naive Bayes juga mudah digunakan dalam penelitian dengan topik analisis sentiment [15].

2.5 Confusion Matrix

Confusion matrix adalah metode evaluasi model yang berguna untuk mengetahui nilai *accuracy*, *precision*, *recall* pada suatu model [16].

- a. *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

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing



Gambar 3. Preprocessing menggunakan Rapidminer

Peneliti menggunakan 300 data *tweet* netizen Twitter yang akan melalui tahap *preprocessing*. 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* [17].

**3.1.1 Cleansing**

Tahap pertama dari *preprocessing* adalah tahap *cleansing*, tahap ini dilakukan untuk membersihkan data dan menghapus elmen-elemen yang tidak dibutuhkan dalam proses penelitian seperti tanda baca, hastag(#), mention/username(@), URL [18]. Setelah proses pembersihan, data terbebas dari elemen-elemen yang tidak diperlukan. Hasil dari proses *cleansing* dapat dilihat pada tabel dibawah ini:

Tabel 1. Ilustrasi *Cleansing*

Sebelum	Sesudah	Sentimen
Kalau lagi ngerasa down dan butuh motivasi coba minta nasihat ChatGPT mana tau okey!!	Kalau lagi ngerasa down dan butuh motivasi coba minta nasihat ChatGPT mana tau okey	Positif
Emang bisa cek plagiasi dari #ChatGPT	Emang bisa cek plagiasi dari ChatGPT	Negatif
ChatGPT ngebantu banget buat ngerjain Tugas, Jadi ngebantu banget deh buat ngerjain tugas!!	ChatGPT ngebantu banget buat ngerjain Tugas Jadi ngebantu banget deh buat ngerjain tugas	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 [19]. Tahap ini penting dilakukan agar semua format data terjaga konsistensinya.

Tabel 2. Ilustrasi *Transform Case*

Sebelum	Sesudah	Sentimen
Kalau lagi ngerasa down dan butuh motivasi coba minta nasihat ChatGPT mana tau okey	kalau lagi ngerasa down dan butuh motivasi coba minta nasihat chatgpt mana tau okey	Positif
Emang bisa cek plagiasi dari ChatGPT	emang bisa cek plagiasi dari chatgpt	Negatif
ChatGPT ngebantu banget buat ngerjain Tugas Jadi ngebantu banget deh buat ngerjain tugas	chatgpt ngebantu banget buat ngerjain tugas jadi ngebantu banget deh buat ngerjain tugas	Positif



Tabel diatas 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 [20]. Berikut ini hasil dari *tokenizing*:

Tabel 3. Ilustrasi *Tokenizing*.

Sebelum	Sesudah	Sentimen
kalaupun lagi ngerasa down dan butuh motivasi coba minta nasihat chatgpt mana tau okey	['kalaupun', 'lagi', 'ngerasa', 'down', 'dan', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'mana', 'tau', 'okey']	Positif
emang bisa cek plagiasi dari chatgpt	['emang', 'bisa', 'cek', 'plagiasi', 'dari', 'chatgpt']	Negatif
chatgpt ngebantu banget buat ngerjain tugas jadi ngebantu banget deh buat ngerjain tugas	['chatgpt', 'ngebantu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebantu', '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 merupakan hasil dari tahap *Filter token*:

Tabel 4. Ilustrasi *Filter token*

Sebelum	Sesudah	Sentimen
['kalaupun', 'lagi', 'ngerasa', 'down', 'dan', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'mana', 'tau', 'okey']	['kalaupun', 'lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'mana', 'okey']	Positif



['emang', 'bisa', 'cek', 'plagiasi', 'dari', 'chatgpt']	['emang', 'bisa', 'plagiasi', 'dari', 'chatgpt']	Negatif
['chatgpt', 'ngebantu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebantu', 'banget', 'deh', 'buat', 'ngerjain', 'tugas']	['chatgpt', 'ngebantu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebantu', '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. Berikut ini hasil dari *Filter stopwords*:

Tabel 5. Ilustrasi *Filter stopwords*

Sebelum	Sesudah	Sentimen
['kalau', 'lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'mana', 'okey']	['lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'okey']	Positif
['emang', 'bisa', 'plagiasi', 'dari', 'chatgpt']	['bisa', 'plagiasi', 'chatgpt']	Negatif
['chatgpt', 'ngebantu', 'banget', 'buat', 'ngerjain', 'tugas', 'jadi', 'ngebantu', 'banget', 'buat', 'ngerjain', 'tugas']	['chatgpt', 'ngebantu', 'banget', 'ngerjain', 'tugas', 'ngebantu', 'banget', 'ngerjain', 'tugas']	Positif



Terlihat perbedaan pada tabel 4 kolom sebelah kiri sebelum dilakukan Filter stopword dan kolom sebelah kanan sesudah dilakukan Filter stopword.

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 [21]. Data bersih yang dimiliki berjumlah 116, lalu sebanyak 58 data yang di label kan manual dan 58 data yang lainnya menggunakan Rapidminer.

Tabel 6. Contoh Pelabelan Data Secara Manual

<i>Tweets</i>	Sentimen
['lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'okey']	Positif
['lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'okey']	Positif
['lagi', 'ngerasa', 'down', 'butuh', 'motivasi', 'coba', 'minta', 'nasihat', 'chatgpt', 'okey']	Positif

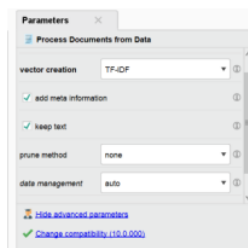
Tabel diatas memperlihatkan contoh data yang telah dilabeli. Setelah semua data telah terlabeli, maka didapatkan visualisasi perbandingan antara jumlah kelas sentimen positif dan kelas sentimen negatif. Berikut merupakan visualisasi perbandingan antara kedua kelas tersebut.



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 [22]. 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 stopword*. 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 Worldcloud

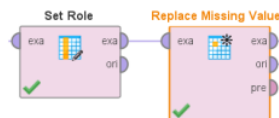


Gambar 6. Visualisasi kata dengan *worldcloud*

Gambar diatas 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.

1 3.5 Algoritma Naïve Bayes

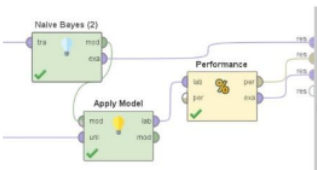
Sebelum memasuki tahap implementasi algoritma Naïve Bayes, *dataset* yang telah dilakukan pelabelan dimasukan ke dalam operator *set role* kemudian dimasukan ke dalam operator *replace missing values*. Fungsi dari kedua operator ini adalah untuk menentukan peran atribut atau kolom dalam analisis dan 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, *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*.

1
Berikut merupakan proses dari tahap implementasi algoritma Naïve Bayes.



Gambar 8. Implementasi algoritma Naïve Bayes



3.6 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

Accuracy yaitu ketepatan prediksi suatu model dalam mengklasifikasikan data, dan kurasi dihitung dengan membagi jumlah prediksi yang benar dengan total jumlah sampel [23]. *Precision* yaitu kalkulasi tingkat ketepatan hasil klasifikasi secara keseluruhan [24]. *Recall* adalah parameter yang mengindikasikan tingkat keberhasilan dalam mengidentifikasi kembali informasi secara akurat [25]. Dibawah ini adalah implementasi dari *confusion matrix*.

Tabel 7. 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 diatas :

- $Precision = \frac{TP}{TP+FP} = \frac{33}{33+4} = 0.89 = 89.19\%$
- $Recall = \frac{TP}{TP+FN} = \frac{33}{33+0} = 1 = 100\%$
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{33+79}{33+79+4+0} = 0.9655 = 96.55\%$

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%.

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

Gambar 10. Confusion matrix model penelitian



27

4. KESIMPULAN

Berdasarkan hasil analisis peneliti terhadap pengolahan data yang bersumber dari Twitter sebanyak 300 data dengan topik ChatGPT pada rentang waktu dari tanggal 13 juni 2023 sampai 30 juni 2023, menyatakan bahwa kehadiran ChatGPT menjadi hal yang mendapat respon positif di kalangan netizen twitter. Dalam segi implementasi algoritma Naïve Bayes, peneliti mendapatkan hasil precision sebesar 89,19%, *recall* sebesar 95,18% dan nilai *accuracy* sebesar 96,55%. Nilai *accuracy* tersebut membuktikan bahwa algoritma Naïve 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.

12

UCAPAN TERIMAKASIH

23

Terima kasih yang sebesar-besarnya saya ucapkan kepada Allah SWT atas segala rahmat dan bimbingan-Nya, serta kepada Bapak Dimas Febriawan dan Bapak Firman Noor Hasan, yang telah dengan sabar dan penuh dedikasi membimbing saya dalam penelitian ini. Ucapan terima kasih juga saya sampaikan kepada orang tua dan keluarga saya, yang selalu memberikan dukungan dan doa. Tak lupa, terima kasih kepada teman-teman yang telah membantu dan memberikan semangat selama proses penelitian. Kesuksesan ini adalah hasil dari dukungan dan doa kita semua.

REFERENCES

dwi transiska - Analisis Sentimen Terhadap Penggunaan Chatgpt Menggunakan Algoritma Naive Bayes

ORIGINALITY REPORT

9%

SIMILARITY INDEX

7%

INTERNET SOURCES

5%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.uhamka.ac.id Internet Source	1%
2	www.stmik-budidarma.ac.id Internet Source	1%
3	Dianati Duei Putri, Gigih Forda Nama, Wahyu Eko Sulistiono. "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier", Jurnal Informatika dan Teknik Elektro Terapan, 2022 Publication	<1%
4	jutif.if.unsoed.ac.id Internet Source	<1%
5	text-id.123dok.com Internet Source	<1%
6	Submitted to Academic Library Consortium Student Paper	<1%
7	Submitted to Universitas Brawijaya Student Paper	<1%

8

Destaria Wilandini, Purwantoro Purwantoro. "Penerapan Algoritma Naïve Bayes dalam mengklasifikasikan Media Sosial untuk mengamati Trend Kuliner", Jurnal Teknologi Terpadu, 2022

Publication

<1 %

9

Jonie Hermanto. "Klasifikasi Teks Humor Bahasa Indonesia Memanfaatkan SVM", Journal of Information System, Graphics, Hospitality and Technology, 2021

Publication

<1 %

10

castos.com

Internet Source

<1 %

11

repository.upnvj.ac.id

Internet Source

<1 %

12

www.researchgate.net

Internet Source

<1 %

13

conference.binadarma.ac.id

Internet Source

<1 %

14

Mochamad Alfian Rosid, Gunawan Gunawan, Edwin Pramana. "Centroid Based Classifier With TF – IDF – ICF for Classification of Student's Complaint at Appliation E-Complaint in Muhammadiyah University of Sidoarjo", JEEE-U (Journal of Electrical and Electronic Engineering-UMSIDA), 2015

Publication

<1 %

15	Naisah Marito Putry. "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS", EVOLUSI : Jurnal Sains dan Manajemen, 2022 Publication	<1 %
16	MASAKI MURATA, MAKOTO NAGAO. "An Estimate of Referent of Nouns in Japanese Sentences with Referential Property of Nouns", Journal of Natural Language Processing, 1996 Publication	<1 %
17	e-jurnal.pelitanusantara.ac.id Internet Source	<1 %
18	www.djournals.com Internet Source	<1 %
19	Oke Dwiraswati, Kemal Nazaruddin Siregar. "ANALISIS SENTIMEN PADA TWITTER TERHADAP PENGGUNAAN ANTIBIOTIK DI INDONESIA DENGAN NAIVE BAYES CLASSIFIER", Media Informasi, 2019 Publication	<1 %
20	klik.ulm.ac.id Internet Source	<1 %
21	ojs.amikom.ac.id Internet Source	<1 %

22	repository.iiq.ac.id Internet Source	<1 %
23	repository.mercubuana.ac.id Internet Source	<1 %
24	repository.usd.ac.id Internet Source	<1 %
25	teknikaftui.wordpress.com Internet Source	<1 %
26	Firman Noor Hasan. Jurnal Linguistik Komputasional (JLK), 2021 Publication	<1 %
27	doku.pub Internet Source	<1 %
28	Muhammad Saifurridho, Martanto Martanto, Umi Hayati. "Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi Shopee", Jurnal Informatika Terpadu, 2024 Publication	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off