

# Analisis Sentimen *Chat GPT* sebagai Masa Depan Pekerja pada Media Sosial *Youtube* menggunakan *Algoritma Klasifikasi Naive Bayes*

M. Yusril Pratama, Ulfa Arma Putri, Putri Afifah D.A, Dian Puspita, Fandi Kurniawan

Jurusan Sistem Dan Teknologi Informasi, Fakultas Teknik Dan Ilmu Komputer  
Universitas Muhammadiyah Kotabumi  
Kotabumi, Indonesia

muham.2059201012@umko.ac.id, ulfaa.2059201037@umko.ac.id, putri.2059201019@umko.ac.id,  
dianp.2059201036@umko.ac.id, fandi.kurniawan@umko.ac.id

**Abstract-** This research focuses on implementing sentiment analysis in conversations on YouTube with the help of Chat GPT and Naive Bayes classification techniques. The experiment uses conversation data between YouTube users and GPT chatbots to understand how the Naive Bayes algorithm can recognize and classify sentiment in those interactions. The results show high accuracy in sentiment classification, although challenges such as language variations and conversational context can influence the analysis results. This study highlights the great potential of Naive Bayes algorithms in helping understand human interactions with technology on social media and emphasizes the need to adapt technology for the future of work, especially in the context of the complexity of interactions on platforms such as YouTube.

**Keywords:** Chat GPT, Sentiment Analisis, Naive Bayes

**Abstrak-** Penelitian ini fokus pada implementasi analisis sentimen dalam percakapan di YouTube dengan bantuan Chat GPT dan teknik klasifikasi Naive Bayes. Eksperimen menggunakan data percakapan antara pengguna YouTube dan chatbot GPT untuk memahami bagaimana algoritma Naive Bayes dapat mengenali dan mengelompokkan sentimen dalam interaksi tersebut. Hasilnya menunjukkan akurasi yang tinggi dalam klasifikasi sentimen, meskipun tantangan seperti variasi bahasa dan konteks percakapan dapat mempengaruhi hasil analisis. Studi ini menyoroti potensi besar algoritma Naive Bayes dalam membantu memahami interaksi manusia dengan teknologi di media sosial dan menekankan perlunya penyesuaian teknologi untuk masa depan pekerjaan, terutama dalam konteks kompleksitas interaksi di platform seperti YouTube.

**Kata Kunci:** Chat GPT, Analisis Sentimen, Naive Bayes

## 1. Pendahuluan

Di era perkembangan digital yang pesat saat ini, dunia maya telah mengubah cara kita berkomunikasi, berbagi informasi, dan bahkan mencari pekerjaan. *YouTube*, salah satu platform media sosial terbesar di dunia, telah mengalami perkembangan signifikan. Selain menjadi tempat utama bagi berbagi video, *YouTube* juga telah menjadi ruang potensial bagi transformasi masa depan dalam dunia pekerjaan. Di tengah kemajuan teknologi chatbot yang didukung oleh kecerdasan buatan, seperti *GPT (Generative Pre-trained Transformer)*. *ChatGPT* merupakan hasil pengembangan yang dilakukan oleh *OpenAI*, sebuah perusahaan penelitian di bidang kecerdasan buatan yang bertempat di San Francisco, California, Amerika Serikat. Proyek

*ChatGPT* dimulai pada tahun 2020 dan secara resmi diperkenalkan kepada publik pada tahun 2021. Model *ChatGPT* ini sendiri merupakan evolusi dari model bahasa generatif terdahulu sebelumnya, yakni *GPT-3*, yang telah diadaptasi dan ditingkatkan untuk menciptakan pengalaman berinteraksi dengan teks yang lebih canggih dan bermanfaat [1], peran mereka semakin relevan dalam ekosistem media sosial. *Chatbot GPT* tidak lagi hanya sekadar program komputer yang memberikan respons teks. Mereka mewakili sebuah visi masa depan pekerjaan yang menghadirkan layanan pelanggan yang lebih cepat, pembelajaran berbasis interaksi, dan berbagai kemungkinan lainnya. *ChatGPT* sendiri merupakan sebuah Artificial Intelligence besar yang

Vol.14 no.2 | Desember 2023

EXPLORE : ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: <http://dx.doi.org/10.36448/jsit.v14i2.3391>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

telah dilatih untuk memahaminya. kemampuan untuk meningkatkan sistem komputer yang sanggup melaksanakan tugas-tugas yang memerlukan kecerdasan manusia. *AI* dapat diprogram untuk memproses informasi, belajar dari pengalaman, dan membuat keputusan berdasarkan data yang diberikan. Bidang ini mencakup berbagai teknik seperti machine learning, neural networks, dan natural language processing. berbagai aplikasi luas. Ada kesalahpahaman yang umum bahwa *AI* hanya terkait dengan robotika, padahal sebenarnya *AI* mencakup berbagai subdomain dan tidak terbatas pada robotika. Salah satu aspek *AI* adalah analisis data besar (big data) yang memungkinkan pengolahan dan pemanfaatan data dalam berbagai konteks [2].

Namun, untuk memahami bagaimana *chatbot GPT* akan membentuk masa depan pekerjaan di media sosial, kita perlu melihat lebih dalam dari sekadar respon yang dihasilkan. Kita perlu melihat bagaimana orang merespons *chatbot* ini secara kognitif, supaya mendapatkan aspirasi masyarakat terkait teknologi ini. Analisis sentimen adalah alat yang memungkinkan kita untuk melakukan ini. Dalam konteks penelitian ini, kami akan mengambil pendekatan analisis sentimen untuk memahami bagaimana interaksi dengan *chatbot GPT* dipahami dan diterima oleh pengguna *YouTube*. Salah satu metode klasifikasi sentimen yang paling umum dan efektif adalah algoritma klasifikasi Naive Bayes. Dengan menggunakan metode ini, kami akan menganalisis sentimen dari percakapan yang terjadi di *YouTube*, membuka jendela ke dalam persepsi publik tentang *chatbot GPT* sebagai masa depan pekerjaan di *platform* ini.

Melalui analisis ini, diharapkan penelitian kami akan memberikan wawasan yang dalam tentang bagaimana teknologi *chatbot GPT* tidak hanya menjadi bagian dari masa depan pekerjaan di media sosial, tetapi juga bagaimana ia akan membentuk dan dipengaruhi oleh harapan dan pengalaman pengguna. Dengan demikian, penelitian ini tidak hanya melihat ke depan menuju masa depan pekerjaan, tetapi juga menggali pemahaman mendalam tentang koneksi manusia dengan kecerdasan buatan di era digital yang terus berkembang.

## 2. Metodologi

### A. Metode

Penelitian ini bertujuan untuk memberikan wawasan yang dalam tentang bagaimana teknologi *chatbot GPT* tidak hanya menjadi bagian dari masa depan pekerjaan di media sosial, tetapi juga bagaimana ia akan membentuk dan dipengaruhi oleh harapan dan pengalaman pengguna. Dengan demikian, penelitian ini tidak hanya melihat ke depan menuju masa depan pekerjaan, tetapi juga menggali pemahaman mendalam tentang koneksi manusia dengan kecerdasan buatan di era digital yang terus berkembang.

### B. Analisis Sentimen

Analisis sentimen merupakan proses otomatis dalam

pengolahan informasi teks dengan tujuan mengevaluasi dan mengidentifikasi sentimen atau pendapat yang terdapat dalam teks seperti postingan, ulasan produk, tweet, atau artikel media sosial lainnya. Dalam analisis sentimen, teks diklasifikasikan menjadi tiga jenis sentimen, yaitu positif, negatif, atau netral. Pendekatan ini dapat membantu industri untuk memahami pendapat publik terhadap produk atau layanan mereka, serta meningkatkan efisiensi strategi pemasaran. Berikut beberapa metode dalam analisis, seperti pendekatan berbasis aturan dan pendekatan berbasis mesin [3].

Langkah-langkah umum dalam analisis sentimen dan klasifikasi teks melibatkan:

1. *Crawling data*: Mengumpulkan dataset yang relevan dengan studi kasus, seperti *review* tentang pembahasa *Chat GPT*.
2. *Pre-processing*: Tahapan merupakan tahapan penting dalam pengolahan teks yang melibatkan beberapa proses esensial. Tahapan ini mencakup tiga langkah utama *Tokenization*: Proses ini melibatkan pemecahan kalimat menjadi token atau kata-kata individual. Dengan melakukan tokenization, teks dapat dipecah menjadi lebih kecil, mempermudah analisis dan pemrosesan lebih lanjut. *Stopwords Removal*: Langkah ini bertujuan untuk menghilangkan kata-kata umum yang tidak memiliki makna penting dalam konteks analisis teks. Kata-kata semacam itu, yang sering disebut sebagai stopwords, tidak memberikan kontribusi signifikan terhadap pemahaman isi teks dan dapat diabaikan agar fokus dapat diberikan pada informasi yang lebih relevan. *Stemming*: Proses ini melibatkan penghapusan akhiran kata untuk mendapatkan akar kata. Dengan melakukan stemming, variasi kata yang mungkin muncul dalam teks dapat disederhanakan menjadi bentuk dasar atau akar kata yang sama. Hal ini membantu mengurangi kompleksitas dan konsistensi dalam analisis teks. Melalui tahapan pre-processing ini, teks yang awalnya kompleks dapat diubah menjadi representasi yang lebih sederhana dan mudah diolah, memungkinkan aplikasi metode analisis teks seperti pengelompokan atau klasifikasi untuk memberikan hasil yang lebih akurat dan bermakna.
3. Transformasi: Memberikan nilai bobot pada data teks, seringkali dengan menggunakan teknik seperti *TF-IDF (Term Frequency-Inverse Document Frequency)*.
4. Seleksi Fitur: Mengurangi jumlah data yang dianalisis dengan memilih fitur-fitur yang paling relevan.
5. Klasifikasi: Penggunaan metode klasifikasi *Naive Bayes* untuk melakukan klasifikasi teks.
6. Interpretasi/Penilaian: Evaluasi dilaksanakan untuk mengukur tingkatan akurasi dan nilai *AUC (Area Under the Curve)*, yang mengindikasikan seberapa efektif model dalam memprediksi sentimen.
7. Opini Publik: Melibatkan pendapat publik yang berkaitan dengan isu-isu yang menyangkut



- kepentingan umum, mencakup berbagai pendapat yang berkumpul sekitar isu tertentu.
8. youtube: Memanfaatkan data dari media social yang memungkinkan pengguna mengirim dan membaca pesan singkat (*youtube*) yang mencerminkan opini dan sentimen.
  9. ChatGPT: Melibatkan analisis sentimen terkait pandemi virus corona, memahami reaksi dan pendapat masyarakat terhadap situasi ini.
  10. Pertambangan Data: Menurut Yuli Mardi, data mining didefinisikan sebagai proses penggunaan teknik atau metode khusus untuk menemukan informasi dari koleksi data yang besar. Tahapan-tahapan kunci dalam proses ini melibatkan *cleaning* data, *integration* data, *selection* data, *transformation* data, *evaluation* pola, dan *presentation* pola.
  11. *Text Mining*: Sebagai sebagian dari data mining, text mining melibatkan analisis data yang berbentuk teks atau dokumen untuk mengungkap pola dan informasi berharga. Tahapan awal pemrosesan atau *preprocessing* data teks melibatkan langkah-langkah seperti tokenization, pembersihan, penyederhanaan huruf, perbaikan kata tidak baku, penyaringan, dan stemming.
  12. *Transformasi Case*: Merubah huruf dalam teks menjadi huruf kecil atau besar.
  13. *Pemfilteran StopWords*: Eliminasi kata-kata yang tidak memberikan kontribusi signifikan dalam menghasilkan pola atau data. *Tokenization*: Memisahkan teks menjadi potongan-potongan yang disebut token untuk analisis lebih lanjut.
  14. *Naïve Bayes Algorithm*: Sebuah algoritma klasifikasi yang berdasarkan teorema Bayes dengan mengasumsikan kemandirian, digunakan untuk melakukan prediksi pada data. Dalam konteks analisis sentimen, Naïve Bayes adalah metode yang efisien dan tepat.
  15. *Performance* : menentukan sejauh mana ukuran sistem atau algoritma data mining mampu menjalankan tugasnya dengan efisien dan efektif.

Semua langkah-langkah ini penting untuk menghasilkan analisis sentimen yang akurat dan dapat diandalkan, memungkinkan pemahaman mendalam terhadap pendapat dan sentimen masyarakat terhadap berbagai isu.

**C. Tujuan Analisis Sentimen**

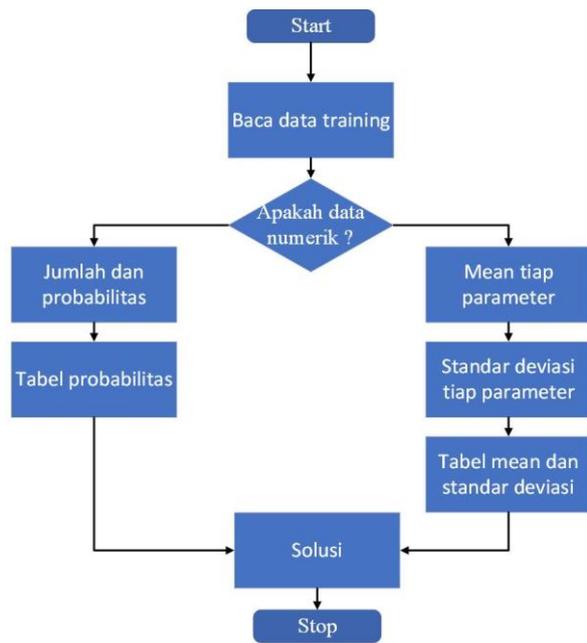
Tujuan analisis sentimen pada penelitian ini adalah:

1. Mengidentifikasi dan memahami persepsi serta pandangan masyarakat terhadap interaksi dengan chatbot GPT di platform YouTube.
2. Menilai tingkat penerimaan chatbot GPT oleh pengguna YouTube.
3. Mengukur harapan masyarakat terhadap peran chatbot GPT sebagai masa depan pekerjaan di media sosial.

4. Mendapatkan wawasan mendalam tentang faktor-faktor yang mempengaruhi sentimen pengguna terhadap kecerdasan buatan.

**D. Naïve Bayes**

Navie Bayes adalah suatu keputusan, dengan menggunakan perhitungan matematika dengan syarat nilai keputusan adalah benar, berdasarkan informasi obyek. Tujuan dari penelitian ini adalah melakukan prediksi klasifikasi analisis sentiment pada chtgpt dengan memanfaatkan algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* berperan dalam mengidentifikasi pengetahuan atau pola kesamaan karakteristik dalam suatu kelompok atau kelas tertentu. [4] Berikut skema algoritma *naïve bayes*:



**Gambar 1** Algoritma *Naïve Bayes*

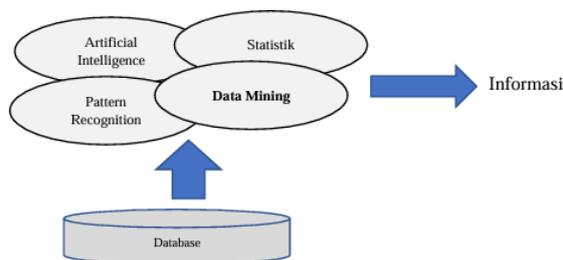
Gambar 1 mengilustrasikan langkah-langkah algoritma Naïve Bayes yang dimulai dengan pembacaan data pelatihan. Setelah itu, algoritma ini membedakan data menjadi parameter yang digunakan untuk menentukan probabilitas setiap parameter. Selanjutnya, algoritma mencari jumlah kemungkinan pada tabel yang telah diacukan, yang mencerminkan distribusi probabilitas dari setiap parameter. Dengan menggunakan informasi ini, algoritma Naïve Bayes dapat menghitung solusi atau prediksi berdasarkan perbandingan probabilitas antara dua dataset.

**E. Data Mining**

Data *mining* merupakan proses eksplorasi data yang bertujuan menciptakan korelasi, pola, serta tren yang signifikan dengan menganalisis set data besar yang tersimpan dalam *repository*. Data mining dapat diartikan sebagai upaya untuk mengidentifikasi pola dan pengetahuan sejumlah data. [5]. Berdasarkan uraian diatas dapat disimpulkan, maka data *mining* merupakan informasi tersembunyi dalam basis data diproses untuk



mengungkap pola, dan teknik statistik, kecerdasan buatan, dan *machine learning* digunakan untuk mengambil atau memisahkan data serta menetapkan informasi pengetahuan dari basis data tersebut.



**Gambar 2.** Proses data *mining*

Pada Gambar 2.2, terlihat bahwa data diambil dari database dan masuk ke dalam proses data mining. Proses ini tetap terkait erat dengan kecerdasan buatan, pengenalan pola, dan statistika, karena merupakan cabang dari ilmu kecerdasan buatan yang memfokuskan pada pengelompokan data dari database menggunakan pendekatan numerik. Data mining bertujuan untuk menggali pola atau informasi yang tersembunyi dalam dataset tersebut. Dengan menggunakan teknik-teknik seperti machine learning dan analisis statistik, data mining dapat menghasilkan output berupa informasi yang telah diprediksi sebelumnya. Ini menciptakan keterkaitan antara data mining, kecerdasan buatan, pengenalan pola, dan statistika, memungkinkan pengambilan keputusan yang lebih baik dan pemahaman yang lebih mendalam terhadap data yang ada.

Berikut tahap permodelan dan analisis data :

1. Pra-pemrosesan data (atau *pre-processing* data) adalah tahap awal dalam analisis data di mana data yang telah dikumpulkan dipersiapkan untuk analisis lebih lanjut. Tujuan dari pra-pemrosesan adalah untuk memastikan bahwa data dalam bentuk yang tepat dan bersih sebelum dianalisis atau dimasukkan ke dalam algoritma pembelajaran mesin. Proses ini melibatkan beberapa tugas, termasuk:
2. Pembersihan Data (*Data Cleaning*): Mendeteksi dan mengatasi nilai-nilai yang hilang atau data yang tidak valid. Tindakan ini melibatkan pengisian nilai-nilai yang tidak muncul, penghapusan baris atau kolom yang tidak lengkap, atau penerapan interpolasi data.
3. Transformasi Data (*Data Transformation*): Merubah bentuk data agar lebih cocok untuk analisis. Proses ini bisa mencakup normalisasi (menyesuaikan nilai variabel ke skala yang seragam) atau transformasi logaritmik atau eksponensial, bergantung pada distribusi data.
4. Encoding Pemberian Kode pada Variabel Kategorikal (*Categorical Variable Encoding*): Apabila

data mencakup variabel kategorikal (contohnya, jenis kelamin, kategori produk), variabel-variabel tersebut harus diubah menjadi bentuk numerik agar dapat diaplikasikan pada algoritma pembelajaran mesin. Salah satu metode untuk melaksanakannya adalah dengan menerapkan teknik seperti *one-hot encoding*.

5. Penghapusan *Noise* (*Noise Removal*): Data sering kali bisa mengandung noise, yang dapat mengganggu analisis. Teknik statistik atau filter digital digunakan untuk menghilangkan noise dari data.
6. Pengaturan Fitur (*Feature Scaling*): Beberapa algoritma pembelajaran mesin, seperti regresi linear atau K-means clustering, memerlukan skala yang seragam pada semua fitur. Oleh karena itu, fitur-fitur yang memiliki skala yang berbeda mungkin perlu diubah agar memiliki skala yang seragam.
7. Ekstraksi Fitur (*Feature Extraction*): Kadang-kadang, fitur-fitur baru yang lebih informatif dapat diekstraksi dari data yang ada. Ini melibatkan teknik-teknik seperti Analisis Komponen Utama (PCA) atau pengekstrakan fitur dari teks.

#### F. Text Preprocessing

Tahap *Text Preprocessing* adalah tahapan kunci dalam analisis teks di mana aplikasi mengidentifikasi dan mempersiapkan data dari setiap dokumen teks. Proses *preprocessing* ini melibatkan beberapa langkah penting untuk membersihkan dan mengubah teks mentah menjadi bentuk yang dapat diolah dengan lebih baik. [6] Langkah-langkah ini mencakup:

1. *Case Folding*: Proses ini melibatkan konversi semua huruf dalam teks menjadi huruf kecil atau huruf besar, bertujuan untuk memastikan konsistensi dalam analisis. Contohnya, mentransformasikan "TeKS ini" menjadi "teks ini"
2. Tokenisasi: Tokenisasi merupakan langkah membagi teks menjadi token-token kecil atau kata-kata individual. Setiap kata atau frasa dalam teks dianggap sebagai token terpisah. Sebagai contoh, kalimat "Analisis teks adalah penting" dapat dipecah menjadi token-token seperti ["Analisis", "teks", "adalah", "penting"].  
Filtering: Pada tahap ini, token-token yang tidak relevan atau tidak diinginkan dapat dihapus dari teks. Ini termasuk tanda baca, karakter khusus, dan kata-kata umum yang tidak memberikan makna tambahan dalam analisis. Proses ini membantu memperkecil dimensi data dan meningkatkan fokus pada kata-kata yang lebih penting.
3. stemming: Stemming merupakan langkah menghapus afiks dari kata untuk mendapatkan akar kata atau bentuk dasarnya. Sebagai contoh, dalam bahasa Inggris, kata-kata seperti "running", "runs", dan "ran" akan distem menjadi bentuk dasar "run". Tindakan ini membantu mengurangi



variasi kata yang memiliki akar kata yang sama, sehingga mempermudah analisis dan pemahaman.

**G. Text Mining**

Text mining merupakan proses memilah-milah bacaan yang tidak terstruktur serta menciptakan data. Tujuannya buat mengenali bentuk, gaya, serta pengetahuan yang tercantum pada bacaan yang besar serta lingkungan. Tata cara text mining mengaitkan pemakaian teknik-teknik analisis linguistik, statistik, serta machine learning buat memproses, mengorganisasi, serta menganalisis bacaan. [7] Proses tahapan *text mining*, antara lain:

- a. Pra-pemrosesan: Melakukan membersihkan dan normalisasi teks, seperti menghapus aksentuasi, menghilangkan kata-kata yang tidak diperlukan.
- b. Tokenisasi: Proses memecah teks menjadi unit terkecil yang dikenal sebagai token.

- c. Representasi Teks: Mengubah teks menjadi numerik yang diolah oleh algoritma. Metode yang digunakan *TF-IDF* atau *Latent Dirichlet Allocation*
- d. Analisis Pemodelan: Melakukan analisis, seperti analisis frekuensi kata, analisis sentimen, klasifikasi bacaan, ataupun analisis topik, buat mendapatkan pengetahuan yang berharga dari bacaan.
- e. Evaluasi: Proses mengevaluasi terkait temuan untuk keputusan yang valid.

*Teks mining* dapat dimanfaatkan dalam beberapa sektor, yaitu sentimen, opini, penambahan, entitas, proses bahasa alami, serta bermacam aplikasi yang lain. Text mining menjadi kunci dalam menggali pengetahuan dari data teks yang melimpah, membantu dalam mengambil keputusan berdasarkan informasi yang terkandung dalam teks tersebut

**3. Hasil dan Pembahasan**

A. Tahapan *Project*

1. *Crawling Data*

Crawling data Mengumpulkan dataset yang relevan dengan studi kasus, seperti review tentang pembahasa Chat GPT menggunakan netlytic

**Tabel 1** Data yang sudah di dapat

Text
Intelegent kecerdasan
Chatgpt apa sudah ada buat tanadingbot
Intinya kita jangan jadi walle
Terima Kasih
Jadi kuli tetep aman ya

2. *Pre-Processing*

Pre-processing data merupakan fase kritis dalam siklus analisis data, di mana langkah-langkah penting diambil untuk memastikan bahwa data yang telah terkumpul bersih, konsisten, dan siap untuk dianalisis. Proses ini melibatkan beberapa tahapan, yang pertama adalah pembersihan data, di mana data yang tidak lengkap, duplikat, atau memiliki nilai yang tidak valid diidentifikasi dan dihapus untuk memastikan keakuratan analisis yang akan datang. Setelah itu, transformasi data dilakukan untuk mengubah format atau struktur data agar lebih sesuai dengan kebutuhan analisis. Encoding variabel kategorikal merupakan langkah selanjutnya, di mana variabel kategori diubah menjadi bentuk yang dapat diinterpretasikan oleh model atau algoritma. Penghapusan *noise* dilakukan untuk menghilangkan gangguan atau variabilitas yang tidak relevan dari data. Pengaturan fitur dan ekstraksi fitur kemudian dilibatkan untuk

memastikan bahwa data yang digunakan memiliki representasi yang optimal dalam konteks analisis tertentu. Selama proses pembersihan data, *Google Colab* digunakan sebagai perangkat lunak yang andal untuk memastikan integritas dan kebersihan data, membantu membangun dasar yang kuat untuk analisis data yang efektif dan akurat.

**Tabel 2** Data yang sudah di cleaning

Text	Sentimen
Intelegent kecerdasan	Positif
Chatgpt apa sudah ada buat tanadingbot	Negatif
Intinya kita jangan jadi walle	Negatif
Terima Kasih	Positif
Jadi kuli tetep aman ya	Negatif

3. *TF-IDF*

Metode *Term Frequency Inverse Document Frequency (TF-IDF)* adalah pendekatan untuk menilai seberapa penting suatu kata (*term*) dalam suatu dokumen pada masing-masing kata tersebut. Proses perhitungan bobot menggunakan *TF-IDF* melibatkan dua aspek utama, yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*.

Pertama-tama, nilai *TF* per kata dihitung untuk menunjukkan seberapa sering kata tersebut muncul dalam suatu dokumen. Rumusnya dinyatakan sebagai berikut:

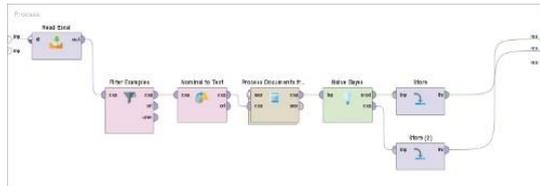
$$(1) \quad TF_{kata} = \frac{\text{Jumlah kemunculan kata dalam dokumen}}{\text{Jumlah total kata dalam dokumen}}$$

**Rumus 1** Perhitungan nilai TF





*Bayes Classifier* mengasumsikan bahwa semua atribut yang diperhitungkan bersifat independen atau tidak saling ketergantungan, sebuah asumsi yang menjadi dasar bagi kemudahan implementasi dan kecepatan perhitungan. Hal ini sangat relevan dalam konteks analisis sentiment data training, di mana algoritma ini mampu memberikan hasil yang cukup akurat dengan mempertimbangkan nilai variabel kelas untuk mengklasifikasikan data dengan tepat.. [11]



Gambar 5. Desain proses *naive bayes*

*Tokenize* proses dokumen to data



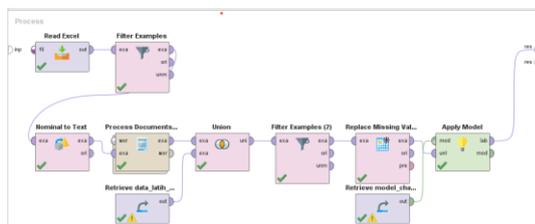
Gambar 6. Desain proses *tokenize*

Hasil data training

Row No.	sentimen	text	asasin	about	abstrak	accidentally	accurred	acih
1	positif	kawankawan bingung.	0	0	0	0	0	0
2	positif	intelligent kecerdasan	0	0	0	0	0	0
3	positif	intelligence kecerdas.	0	0	0	0	0	0
4	positif	pandangan pengal...	0	0	0	0	0	0
5	positif	chatgpt tadingbot	0	0	0	0	0	0
6	negatif	mantap lirik link lan...	0	0	0	0	0	0
7	negatif	indrya walle	0	0	0	0	0	0
8	negatif	heboh media menge...	0	0	0	0	0	0
9	negatif	ngalahin google tahu...	0	0	0	0	0	0
10	negatif	apalibrya linal lutan	0	0	0	0	0	0
11	positif	mantap informasi ter...	0	0	0	0	0	0
12	positif	terima kasih	0	0	0	0	0	0
13	positif	menjelajahi pidato in...	0	0	0	0	0	0
14	negatif	mengumpulkan kecer...	0	0	0	0	0	0

Gambar 7. Desain proses hasil training

Data analisis sentiment prediksi



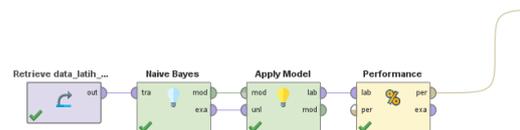
Gambar 8. Desain proses prediksi dari data training

Row No.	text	sentimen	prediction...	confidence...	confidence...	about	about	adanya	adalah
1	terminator	?	negatif	0	1	0	0	0	0
2	lambat banget	?	positif	1	0	0	0	0	0
3	emang updat.	?	positif	1	0	0	0	0	0
4	lmu murah	?	negatif	0	1	0	0	0	0
5	chatgpt benar.	?	positif	1	0	0	0	0	0
6	universal bas.	?	negatif	0	1	0	0	0	0
7	aman ehem...	?	negatif	0	1	0	0	0	0
8	udah pale ch.	?	positif	1	0	0	0	0	0
9	blanks intena	?	positif	1	0	0	0	0	0
10	blanyou aha.	?	positif	1	0	0	0	0	0
11	terngayu	?	negatif	0	1	0	0	0	0
12	channel faur.	?	negatif	0	1	0	0	0	0
13	samsama	?	negatif	0	1	0	0	0	0
14	reparat	?	positif	1	0	0	0	0	0

Gambar 9. Desain proses hasil prediksi

### 7. Performance

Kinerja (*performance*) dalam data mining adalah ukuran sejauh mana suatu sistem atau algoritma data mining mampu menjalankan tugasnya dengan efisien dan efektif. Kinerja ini dapat diukur dengan berbagai metrik yang sesuai dengan tujuan analisis data yang dilakukan. Beberapa faktor yang dapat memengaruhi kinerja dalam data mining meliputi akurasi hasil, kecepatan eksekusi, kemampuan untuk mengekstrak pola yang berarti, serta kemampuan untuk menangani data yang besar dan kompleks. Dalam mengevaluasi performa algoritma Naive Bayes Classifier, langkah selanjutnya melibatkan pencatatan nilai dan analisis menggunakan *Confusion Matrix*. *Confusion Matrix* memberikan gambaran yang terperinci tentang sejauh mana algoritma dapat mengklasifikasikan data dengan benar, dengan memperhitungkan *True Positives*, *True Negatives*, *False Positives*, dan *False Negatives*. Setelah mendapatkan *Confusion Matrix* dari berbagai percobaan, evaluasi dilakukan dengan menghitung nilai rata-rata dari berbagai metrik yang dapat diambil dari matriks tersebut, seperti akurasi, presisi, recall, dan *F1-score*. Proses ini memberikan pemahaman holistik tentang kinerja algoritma dalam berbagai konteks dan kondisi data. Nilai rata-rata yang dihasilkan dari evaluasi ini tidak hanya memberikan indikasi tentang seberapa baik model dapat memprediksi, tetapi juga membantu dalam mengidentifikasi percobaan yang dapat dijadikan acuan untuk penggunaan model algoritma yang dipilih. Dengan pendekatan ini, pengguna dapat membuat keputusan yang lebih informasional dan tepat terkait penerapan *Naive Bayes Classifier* pada data spesifik, meningkatkan keahlian dan efektivitas model yang dikembangkan. [12] Berikut adalah *Confussion matrix*.



Gambar 10. Desain proses *performance* di rapidminer



accuracy: 95.88%			
	true positif	true negatif	class precision
pred positif	582	0	100.00%
pred negatif	33	186	84.93%
class recall	94.63%	100.00%	

**Gambar 11.** Hasil *performance confusion metrik*

Pada gambar 3. 9 Hasil confusion matrix menyediakan landasan untuk menghitung metrik-metrik di atas. Dengan mengetahui nilai TP, TN, FP, dan FN, kita dapat menghitung akurasi, presisi, recall, dan F1-Score. Evaluasi model dengan menggunakan kriteria-kriteria ini membantu dalam pemahaman yang lebih baik tentang kinerja model pada data yang tidak terlihat selama pelatihan. Presisi recall data benar positif akurasi sebesar 94% sedangkan presisi recall benar negatif 100%.

#### 4. Kesimpulan

Berdasarkan penelitian tentang analisis cht gpt sebagai masa depan pekerjaan dapat disimpulkan adalah

1. Metode *naive Bayes* memakai data training buat menciptakan peluang tiap kriteria dalam kelas yang berbeda. Maka, nilai probabilitas dari kriteria tersebut bisa dimaksimalkan dalam konteks penggunaan ChatGPT sebagai fokus masa depan pekerjaan.
2. Berdasarkan pemanfaatan ChatGPT sebagai data training untuk metode *naive Bayes*, terdapat 1119 data yang dihasilkan, dan setelah proses ini, jumlahnya berkurang menjadi 1086. Oleh karena itu, metode *naive Bayes* berhasil memprediksi ChatGPT sebagai masa depan pekerja dengan tingkat akurasi sebesar 95,88%.
3. Algoritma *naive Bayes* mengandalkan prinsip probabilitas dalam memanfaatkan data petunjuk untuk mendukung proses pengklasifikasian keputusan

#### 5. Daftar Pustaka

- [1] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naive Bayes," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v4i3.1368>
- [2] A. P. O. Dewi, "Kecerdasan Buatan sebagai Konsep Baru pada Perpustakaan," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 4, no. 4, pp. 453–460, 2020, doi: 10.14710/anuva.4.4.453-460.
- [3] F. Kurniawan dan Q. Al Qorni, "Exploring Sentimen Analysis Using Machine Learning : A Case Study on Partai Demokrasi Indonesia Perjuangan ( PDIP ) in the 2024 General Election," *sinomics journal*, vol. 2, no. 4, hal. 911–920, 2023.
- [4] Rayuwati, Husna Gemasih, and Irma Nizar, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.
- [6] L. Hermawan and M. Bellanar Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *J. Transform.*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.
- [7] M. N. Fahriza and N. Riza, "Analisis Sentimen Pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer Gpt Menggunakan Metode Klasifikasi K-Nearest Neighbor(Knn) Systematic Literature Review," *J. Mbs. Tek. Inform.*, vol. 7, no. 2, pp. 1351–1358, 2023.
- [8] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/796>
- [9] E. Purnaningrum and I. Ariqoh, "Google Trends Analytics Dalam Bidang Pariwisata," *Maj. Ekon.*, vol. 24, no. 2, pp. 232–243, 2019, doi: 10.36456/majeko.vol24.no2.a2069.
- [10] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>
- [11] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naive Bayes Classifier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [12] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Di Twitter," *J. TEKNO KOMPAK*, vol. 17, no. 2, pp. 25–38, 2022.

