Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS

https://doi.org/10.35870/jimik.v5i3.1016

Analisa Sentimen Pada Media Sosial "X" Pencarian Keyword ChatGPT Menggunakan Algoritma *K-Nearest Neighbors* (KNN)

Yuma Akbar [1], Anggit Nur Hannaa Regita [2*], Sugiyono [3], Tri Wahyudi [4]

[1,2*,3,4] Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: yumekhan@stikomcki.ac.id [1], anggitnurhannaregita@gmail.com [2*], inosoguy007@gmail.com [3], triwahyudi100390@gmail.com [4]

Histori Artikel:
Dikirim 29 Juli 2024; *Diterima dalam bentuk revisi* 18 Agustus 2024; *Diterima* 30 Agustus 2024; *Diterbitkan* 30 September 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Analisis sentimen terhadap penggunaan kecerdasan buatan (Artificial Intelligence/AI) semakin penting dalam pemahaman masyarakat terhadap teknologi yang berkembang pesat saat ini. Teknologi ini mempermudah manusia dalam menjalankan aktivitas sehari-hari. Salah satu wujud penerapannya adalah hadirnya ChatGPT, yaitu sebuah AI yang mampu berinteraksi dengan pengguna (user) melalui input dari pengguna, seperti menjawab berbagai pertanyaan yang diberikan. Topik ini memunculkan banyak pro dan kontra, sebagaimana banyak dibicarakan di media sosial. Penelitian diperlukan untuk menilai seberapa bijak masyarakat dalam menggunakan AI ini. Penelitian ini mengusulkan pendekatan menggunakan algoritma K-Nearest Neighbors (KNN) untuk menganalisis sentimen terkait AI. Algoritma KNN digunakan untuk mengklasifikasikan sentimen menjadi positif, negatif, atau netral berdasarkan kemiripan dengan kata terdekat dalam ruang fitur yang dihasilkan dari data teks. Metode ini memungkinkan pengelompokan sentimen secara efisien tanpa memerlukan model yang kompleks. Peneliti memilih analisis sentimen karena merupakan teknik yang tepat untuk pengolahan dataset. Dari 1000 opini pengguna media sosial "X" yang berhasil dikumpulkan, didapatkan 853 opini positif dan 147 opini negatif. Data tersebut kemudian diklasifikasikan menggunakan algoritma KNN, dan hasil klasifikasi menunjukkan akurasi sebesar 84,80%. Hasil analisis sentimen ini diharapkan dapat menjadi panduan bagi pengambil keputusan dalam mengembangkan dan menerapkan teknologi AI secara lebih bijak, sesuai dengan kebutuhan dan harapan masyarakat.

Kata Kunci: K-Nearest Neighbors; Artificial Intelligence; ChatGPT; X.

Abstract

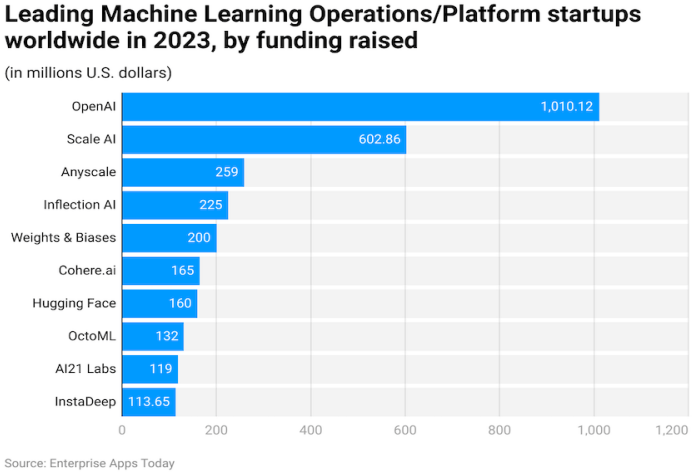
Sentiment analysis of the use of Artificial Intelligence (AI) is becoming increasingly important in public understanding of today's rapidly evolving technology, as it helps facilitate human activities. One of the key applications is the presence of ChatGPT, an AI capable of interacting with users through user input, such as answering various questions posed. This topic generates a lot of pros and cons, as widely discussed on social media. Research is needed to evaluate how wisely people use this AI. This study proposes an approach using the K-Nearest Neighbors (KNN) algorithm to analyze AI-related sentiment. The KNN algorithm is used to classify sentiment into positive, negative, or neutral, based on the similarity with the closest word in the feature space derived from text data. This method allows for efficient sentiment grouping without the need for complex models. Researchers chose sentiment analysis because it is an appropriate technique for data processing. Of the 1000 reviews collected from social media users on "X," 853 were positive, and 147 were negative. The data was classified using the KNN algorithm, followed by an accuracy evaluation yielding 84.80%. The results of this sentiment analysis are expected to guide decision-makers in developing and applying AI technology more intelligently, in line with societal needs and expectations.

Keyword: K-Nearest Neighbors; Artificial Intelligence; ChatGPT; X.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

1. Pendahuluan

Kecerdasan buatan (*Artificial Intelligence*/AI) telah berkembang pesat dalam beberapa tahun terakhir dan membawa berbagai inovasi yang memengaruhi berbagai aspek kehidupan manusia. Salah satu aplikasi AI yang semakin populer adalah sistem *chatbot*, yaitu *ChatGPT*, yang memungkinkan interaksi dengan pengguna melalui antarmuka suara atau teks. *ChatGPT* menggunakan teknologi *Natural Language Processing* (NLP) dan AI untuk menghasilkan respons terhadap pertanyaan yang diajukan oleh pengguna (Zhai, 2023). *ChatGPT*, yang dikembangkan oleh OpenAI, adalah model berbasis teks dengan kemampuan menghasilkan tanggapan yang menyerupai manusia terhadap pertanyaan pengguna. Teknologi ini telah menarik minat banyak pihak, mulai dari individu hingga perusahaan besar, untuk berbagai keperluan seperti dukungan teknis dan layanan pelanggan (Pakpahan, 2021).



Gambar 1. Grafik Pengguna *Platform Chat GPT*

Namun, untuk memahami perasaan, pendapat, dan tanggapan pengguna terhadap *ChatGPT*, penting untuk mengeksplorasi dampak negatif potensial, terutama ketika teknologi ini disalahgunakan oleh pihak yang tidak bertanggung jawab. Pemahaman mengenai interaksi pengguna dengan aplikasi ini diperlukan agar dapat mengurangi ketergantungan yang tidak sehat pada *ChatGPT*. Meskipun AI telah membantu menyelesaikan berbagai masalah, hal ini juga menimbulkan tantangan, seperti peningkatan persaingan di dunia kerja. Oleh karena itu, manusia perlu terus berkembang dan memperdalam pengetahuan tentang AI agar tidak sepenuhnya tergantikan oleh teknologi (Pakpahan, 2021).

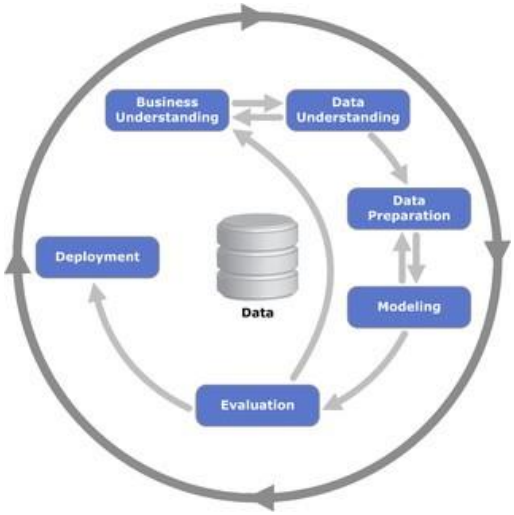
Berdasarkan situasi tersebut, analisis sentimen menjadi metode yang bermanfaat untuk memahami interaksi pengguna dengan *ChatGPT*. Algoritma *K-Nearest Neighbors* (KNN) adalah salah satu teknik yang dapat digunakan dalam analisis sentimen dan pembelajaran mesin, khususnya untuk klasifikasi dan regresi. Dalam analisis sentimen, KNN digunakan untuk mengklasifikasikan sentimen pengguna terhadap *ChatGPT* berdasarkan elemen-elemen seperti kata kunci dalam interaksi, durasi percakapan, atau pola komunikasi (Zhai, 2023).

Twitter, yang kini berganti nama menjadi X, adalah salah satu platform media sosial paling populer di Indonesia. Popularitas Twitter menjadikannya platform utama bagi publik untuk mengekspresikan pendapat terkait berbagai isu, termasuk penggunaan *ChatGPT* (Rahayu *et al.*, 2023). Berdasarkan uraian tersebut, maka peneliti ini bertujuan untuk mengetahui sentimen masyarakat terhadap *Chat* GPT pada media sosial "X" berupa sentimen positif dan negatif dengan menggunakan data "X" yang beredar dengan metode klasifikasi menggunakan algoritma *K- Nearest Neighbors* (KNN) hingga kemudian dapat dinilai apakah sentimen masyarakat cenderung kearah positif atau negatif.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

2. Metode Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah metode *Cross Industry Standard for Data Mining* (CRISP-DM). CRISP-DM merupakan kerangka kerja yang banyak diterapkan dalam pengembangan data mining dan analisis data, serta dirancang untuk membantu peneliti dan praktisi menyelesaikan masalah yang berhubungan dengan pengolahan data secara sistematis.



Gambar 2. Model Proses Penelitian

Metode ini terdiri dari enam tahap yang dijelaskan sebagai berikut:

2.1 Pemahaman Bisnis (*Business Understanding*)

Tahap pertama dalam CRISP-DM adalah pemahaman bisnis. Pada tahap ini, tujuan penelitian diarahkan pada pemahaman serta identifikasi kebutuhan analisis yang ingin dicapai. Dalam penelitian ini, fokus utama adalah untuk memahami sentimen publik terhadap penggunaan *ChatGPT* di media sosial "X" (sebelumnya dikenal sebagai Twitter). Tujuan penelitian adalah mengetahui bagaimana persepsi dan tanggapan masyarakat terkait *ChatGPT*, baik yang positif maupun negatif. Untuk memperoleh data, penelitian menggunakan API *Get Data* dari platform "X". Data mentah yang terkumpul kemudian diproses lebih lanjut menggunakan berbagai teknik seperti penggantian (*replace*) RT, tautan, *hashtag*, *mention*, simbol, serta menghapus duplikasi dan memangkas teks yang tidak relevan. Hasil akhir dari proses ini adalah data yang siap dianalisis menggunakan algoritma *K-Nearest Neighbors* (KNN). Proses ini memungkinkan klasifikasi sentimen berdasarkan kata kunci tertentu dalam dataset.

2.2 Pemahaman Data (*Data Understanding*)

Setelah tahap pengumpulan data selesai, langkah berikutnya adalah memahami data yang diperoleh. Pada tahap ini, peneliti mengevaluasi dan memastikan bahwa data yang terkumpul relevan dengan tujuan penelitian. Data diambil dari platform "X" melalui API *Get Data* dengan menggunakan teknik *crawling* untuk mengumpulkan berbagai tanggapan dan opini pengguna. Pada tahap ini juga dilakukan analisis deskriptif awal untuk mengevaluasi kualitas data, seperti memeriksa distribusi frekuensi dan mendeteksi potensi masalah dalam dataset, seperti nilai yang hilang atau duplikasi. Pemahaman data ini menjadi dasar untuk menentukan pendekatan analisis yang tepat serta langkah-langkah lebih lanjut yang diperlukan.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

2.3 Persiapan Data (*Data Preparation*)

Persiapan data merupakan salah satu tahapan paling penting dalam CRISP-DM karena mencakup proses transformasi data mentah menjadi bentuk yang siap untuk pemodelan. Data yang diperoleh dari platform "X" melalui API diproses melalui beberapa langkah untuk menghasilkan dataset yang bersih dan dapat digunakan. Berikut adalah tujuh langkah utama dalam proses persiapan data:

1) Pengumpulan Data
   Data dikumpulkan dari media sosial, termasuk TikTok dan platform "X", menggunakan kata kunci yang relevan seperti penggunaan *ChatGPT*. Data tersebut mencakup komentar, tanggapan, dan diskusi terkait dengan topik penelitian.
2) Pembersihan Data (*Data Cleansing*)
   Proses ini melibatkan pembersihan data dari elemen yang tidak diperlukan seperti karakter khusus, tanda baca, tautan, serta *stopwords* (kata-kata umum yang tidak memberikan informasi signifikan, seperti "dan", "atau", "juga"). Pembersihan ini bertujuan untuk meningkatkan kualitas data yang akan dianalisis.
3) Tokenisasi
   Teks yang dikumpulkan dipecah menjadi unit-unit lebih kecil yang disebut *token*. Proses tokenisasi ini sangat penting untuk mengubah data teks menjadi format yang dapat dianalisis lebih lanjut. Misalnya, dalam bahasa Inggris, tokenisasi dilakukan dengan memisahkan kata berdasarkan spasi.
4) Stemming atau Lemmatisasi
   Pada tahap ini, kata-kata direduksi menjadi bentuk dasarnya. *Stemming* dilakukan dengan menghilangkan akhiran kata, sedangkan *lemmatisasi* mengembalikan kata ke bentuk asalnya dengan mempertimbangkan struktur kalimat. Kedua teknik ini membantu menyederhanakan data teks.
5) Menghilangkan *Noise*
   Elemen yang tidak relevan, seperti angka, simbol, atau kata-kata yang sering muncul tetapi tidak berkaitan langsung dengan sentimen, dihilangkan dari data. Proses ini membantu menyaring informasi yang tidak diperlukan untuk analisis sentimen.
6) Normalisasi
   Teks dinormalisasi dengan menghapus tanda baca yang tidak penting atau mengganti frasa umum dengan istilah standar. Contohnya, frasa seperti "tidak baik" dapat diubah menjadi "buruk". Normalisasi ini memudahkan untuk mengidentifikasi pola dalam data.
7) Pelabelan Sentimen (*Sentiment Labeling*)
   Setelah pembersihan, data kemudian diberi label sesuai dengan kategorisasi sentimen, seperti "positif", "negatif", atau "netral". Pelabelan ini bisa dilakukan secara manual atau otomatis dengan menggunakan kamus sentimen atau model klasifikasi.

2.4 Pemodelan (*Modelling*)

Tahap pemodelan adalah saat data yang telah dipersiapkan digunakan untuk membangun model analisis sentimen. Dalam penelitian ini, model yang digunakan adalah algoritma *K-Nearest Neighbors* (KNN). Algoritma KNN dipilih karena keandalannya dalam mengklasifikasikan sentimen berdasarkan kemiripan antara data yang dianalisis. KNN bekerja dengan menghitung jarak antara titik data dalam ruang fitur dan memilih tetangga terdekat untuk menentukan sentimen. Dalam penelitian ini, model KNN digunakan untuk mengklasifikasikan data menjadi tiga kategori utama: sentimen positif, sentimen negatif, dan kategori netral. Data yang telah dibersihkan dan diproses dimasukkan ke dalam alat pemodelan seperti *RapidMiner* untuk menghasilkan prediksi klasifikasi.

2.5 Evaluasi (*Evaluation*)

Setelah model dibangun, langkah selanjutnya adalah evaluasi untuk memastikan bahwa model yang dikembangkan menghasilkan prediksi yang akurat. Pada tahap ini, dilakukan pengukuran performa model menggunakan berbagai metrik seperti akurasi, presisi, *recall*, dan F1-score. Metrik-

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

metrik ini digunakan untuk menilai seberapa baik model mampu mengklasifikasikan sentimen secara benar. Jika hasil evaluasi menunjukkan bahwa model belum memberikan performa yang diinginkan, peneliti dapat melakukan penyesuaian pada model atau metode yang digunakan. Proses iterasi ini bertujuan untuk mendapatkan model yang lebih optimal dan relevan dengan tujuan penelitian.

2.6 Penerapan (*Deployment*)

Tahap terakhir dari metode CRISP-DM adalah penerapan hasil penelitian. Pada tahap ini, model yang sudah dievaluasi diterapkan dalam praktik dan hasilnya disajikan dalam bentuk laporan atau presentasi. Hasil analisis sentimen yang dilakukan terhadap data dari media sosial "X" digunakan untuk memberikan gambaran mengenai persepsi publik terhadap penggunaan *ChatGPT*. Penerapan ini juga memungkinkan para pengambil keputusan untuk menggunakan hasil analisis dalam merancang strategi yang lebih sesuai dengan persepsi masyarakat. Dengan informasi yang diperoleh, dapat disusun langkah-langkah yang lebih tepat dalam mengembangkan teknologi atau layanan yang berkaitan dengan *ChatGPT*.

3. Hasil dan Pembahasan

3.1 Hasil

Hasil Pengujian ini dapat dilakukan dengan menggunakan metode *Cross – Industry Standard Process for Data Mining* (CRISP-DM) yang memiliki 6 tahapan yaitu:

3.1.1 Pemahaman Bisnis (*Bussiness Understanding*)

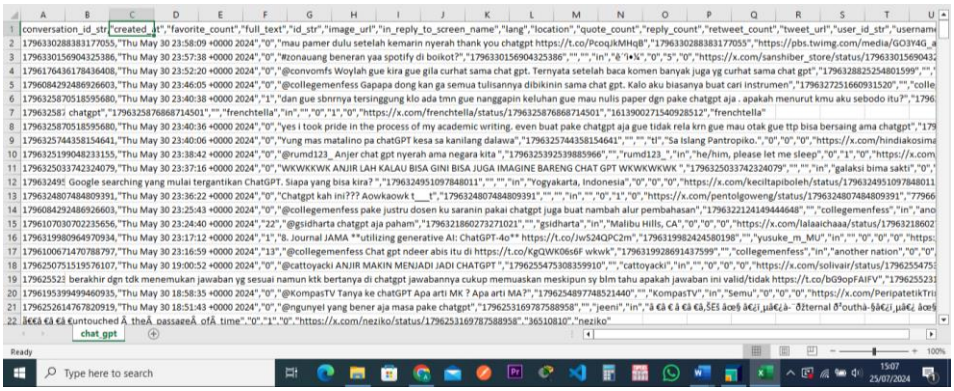
Chat GPT, yang dikembangkan oleh *OpenAI*, adalah model berbasis teks yang memiliki kemampuan untuk menciptakan tanggapan yang mirip dengan manusia untuk pertanyaan yang diberikan pengguna. *Chat GPT* telah berhasil membuat banyak orang, mulai dari individu hingga perusahaan besar, tertarik untuk menggunakan teknologi ini. Namun, untuk memahami perasaan, pendapat, dan tanggapan pengguna terhadap *Chat GPT* memunculkan dampak negatif yang seringkali disalah gunakan terhadap orang yang tidak bertanggung jawab yang menyebabkan ketergantungan terhadap *Chat GPT* ini. Tujuan dasar adalah untuk menentukan mencari nilai akurasi, *presisi* dan *recall* dan menganalisa sentimen masyarakat tentang *Chat GPT*.

3.1.2 Pemahaman Data (*Data Understanding*)

Memahami informasi data yang akan digunakan dalam penelitian, adapun beberapa tahap untuk melakukan penelitian ini yaitu pengumpulan data awal yang dibutuhkan untuk mendukung dalam melakukan pemahaman data. Adapun sumber data utama yang digunakan dalam penelitian ini adalah data dari media sosial X *Chat GPT* pada 1 Januari 2023 – 31 Mei 2023, 1 Januari 2022- 1 Januari 2024, 1 September 2022- 30 Desember 2024, 1 Januari 2021- 30 Maret 2022, 1 Juli 2023- 30 Oktober 2024, 1 Februari 2023- 30 Mei 2024.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

Gambar 3. Model Proses *Get* Data

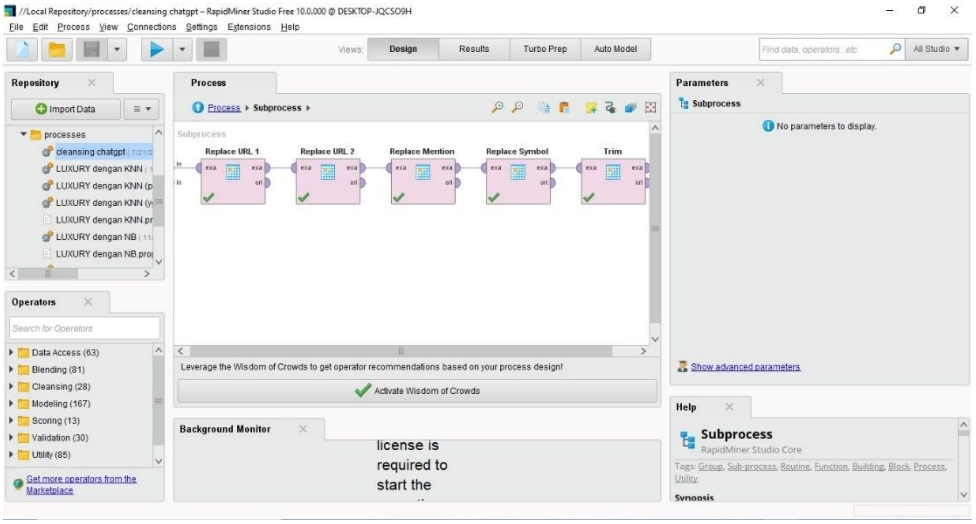


Gambar 4. Hasil Pengumpulan Crawling Data X

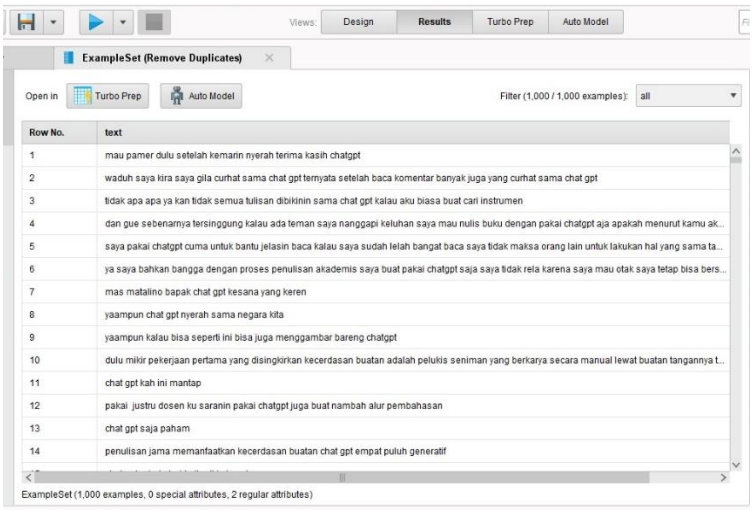
3.1.3 Data *Preparation*

Pada tahap ini, peneliti akan melakukan proses *cleansing* dengan menggunakan *RapidMiner* sebelum melakukan pemberian label pada sentimen, agar *dataset* pada csv terbaru terhindar dari duplikasi data dan tanda yang tidak diperlukan. Pada proses ini akan dilakukan pembersihan dari berbagai *noise* seperti menghilangkan *link* URL, *username*, *tag,* digit angka, dan karakter. Setelah selesai menjalankan proses *cleansing* dengan menggunakan *RapidMiner*. Awalnya terdapat 1.143 data X pada file csv, kemudian setelah melewati proses *cleansing*, maka data komentar berkurang menjadi 1000 data. Gambar 5 menunjukkan tahapan yang dilakukan dalam *RapidMiner* untuk menghapus *noise* dari dataset. Pada gambar tersebut, proses pembersihan data yang mencakup penghapusan URL, username, tag, dan elemen tidak relevan lainnya dapat dilihat dengan jelas. Proses ini penting untuk memastikan data yang digunakan dalam analisis hanya mencakup teks dari komentar pengguna yang relevan. Gambar 6 menggambarkan perbandingan antara jumlah data sebelum dan sesudah proses *cleansing*. Jumlah data awal yang berjumlah 1.143 berkurang menjadi 1.000 setelah penghapusan elemen-elemen yang tidak diperlukan. Penurunan jumlah data ini menunjukkan bahwa sejumlah besar *noise* berhasil dihapus, sehingga dataset menjadi lebih bersih dan siap untuk dianalisis.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016



Gambar 5. Proses Model *Cleansing* Data



Gambar 6. Hasil *Cleansing* Data

1) *Labeling* Data

Tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentimen berdasarkan masing-masing X. Selama proses penentuan sentimen berlangsung, peneliti melakukan pemberian sentimen secara manual. Berikut ini data yang sudah diberi label. Setelah proses pembersihan data selesai, tahap selanjutnya adalah pelabelan data atau klasifikasi sentimen. Pada tahap ini, peneliti menentukan apakah setiap entri dalam dataset memiliki sentimen positif, negatif, atau netral. Proses pelabelan ini sangat penting karena data yang telah diberi label nantinya akan digunakan sebagai acuan dalam pemodelan dan evaluasi sentimen. Pelabelan dilakukan secara manual oleh peneliti untuk memastikan akurasi dan konsistensi dalam interpretasi sentimen. Setiap entri dievaluasi berdasarkan isi teks yang disampaikan oleh pengguna, dan peneliti secara teliti menentukan apakah respons tersebut bernuansa positif, negatif, atau netral. Meskipun pelabelan otomatis dengan algoritma tertentu dapat dilakukan, pelabelan manual dalam konteks ini dipilih karena memungkinkan peneliti untuk lebih memahami konteks yang lebih kompleks dan variasi dalam bahasa yang digunakan pengguna media sosial.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

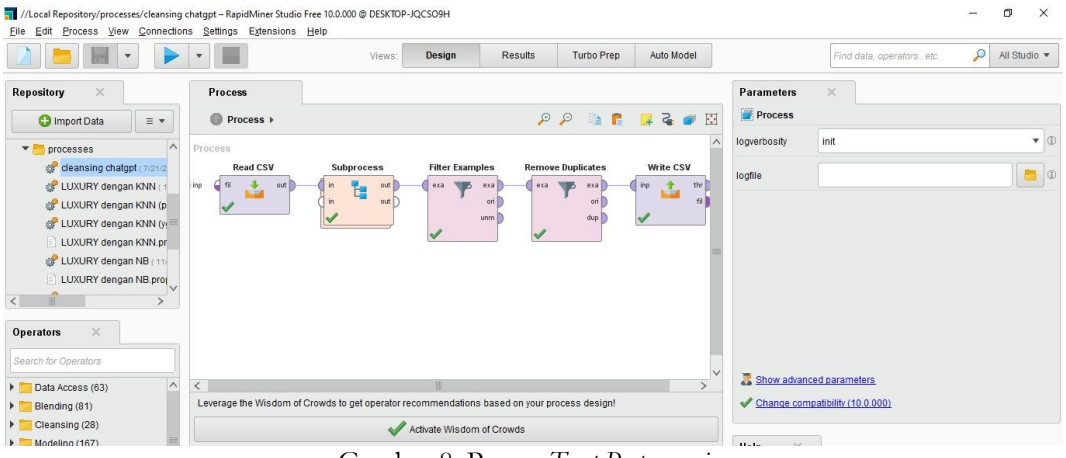
	A	B	C
1	no	text	label
2	1	mau pamer dulu setelah kemarin nyerah terima kasih chatgpt	positif
3	2	waduh saya kira saya gila curhat sama chat gpt ternyata setelah baca komentar banyak juga yang curhat sama chat gpt	positif
4	3	tidak apa apa ya kan tidak semua tulisan dibikinin sama chat gpt kalau aku biasa buat cari instrumen	positif
5	4	dan gue sebenarnya tersinggung kalau ada teman saya nanggapi keluhan saya mau nulis buku dengan pakai chatgpt aja apakah	positif
6	5	saya pakai chatgpt cuma untuk bantu jelasin baca kalau saya sudah lelah bangat baca saya tidak maksa orang lain untuk lakuk	positif
7	6	ya saya bahkan bangga dengan proses penulisan akademis saya buat pakai chatgpt saja saya tidak rela karena saya mau otak s	positif
8	7	mas matalino bapak chat gpt kesana yang keren	positif
9	8	yaampun chat gpt nyerah sama negara kita	negatif
10	9	yaampun kalau bisa seperti ini bisa juga menggambar bareng chatgpt	positif
11	10	dulu mikir pekerjaan pertama yang disingkirkan kecerdasan buatan adalah pelukis seniman yang berkarya secara manual lewat	positif
12	11	chat gpt kah ini mantap	positif
13	12	pakai justru dosen ku saranin pakai chatgpt juga buat nambah alur pembahasan	positif
14	13	chat gpt saja paham	positif
15	14	penulisan jama memanfaatkan kecerdasan buatan chat gpt empat puluh generatif	positif
16	15	chat gpt admin habis itu di ketawain	positif
17	16	yaampun makin jadi jadi chat gpt ini	positif
18	17	sebelum saya sudah coba untuk cari tahu mengapa pada sp dua ribu dua puluh tidak tersedia informasi tentang suku bangsa sa	positif
19	18	tanya ke chat gpt apa arti mahkama konstitusi apa arti mahkamah agung	positif
20	19	psikolog mahal jadi kalau curhat konsul ke chat gpt saja tanpa pamrih dia kecuali yang premium	positif
21	20	hobi saya ini curhat sama chat gpt	positif
22	21	ini mengingatkan lagi cium tidak ya tapi menarik chat gpt bisa seperti ini	positif
23	22	chatgpt serbaguna banget	positif
24	23	habis curhat sama chat gpt dan di kasih kepastian terima kasih ya chat gpt	positif
25	24	jadi dia bangga gimana dulu saya juga bangga bisa pakai chat gpt jadi permudah tidak banyak buang waktu lagian chat gpt tida	positif
26	25	kamu pakai chat gpt tidak lahat	positif

Gambar 7. Data Telah Dilabeli Sentimen

Gambar 7 menunjukkan hasil dari proses pelabelan sentimen. Setiap data dalam dataset telah diberi label yang mengklasifikasikan apakah komentar tersebut memiliki sentimen positif, negatif, atau netral. Pada gambar tersebut, tampak distribusi dari berbagai sentimen yang telah diberi label, memberikan gambaran tentang kecenderungan sentimen yang terkandung dalam dataset. Pelabelan manual memastikan bahwa klasifikasi sentimen dilakukan dengan cermat, mengingat nuansa dan konteks bahasa yang mungkin tidak selalu bisa ditangkap oleh algoritma otomatis. Hasil dari pelabelan ini akan digunakan sebagai dasar untuk membangun model klasifikasi yang lebih lanjut, termasuk analisis dengan algoritma *K-Nearest Neighbors* (KNN).

2) *Modelling*

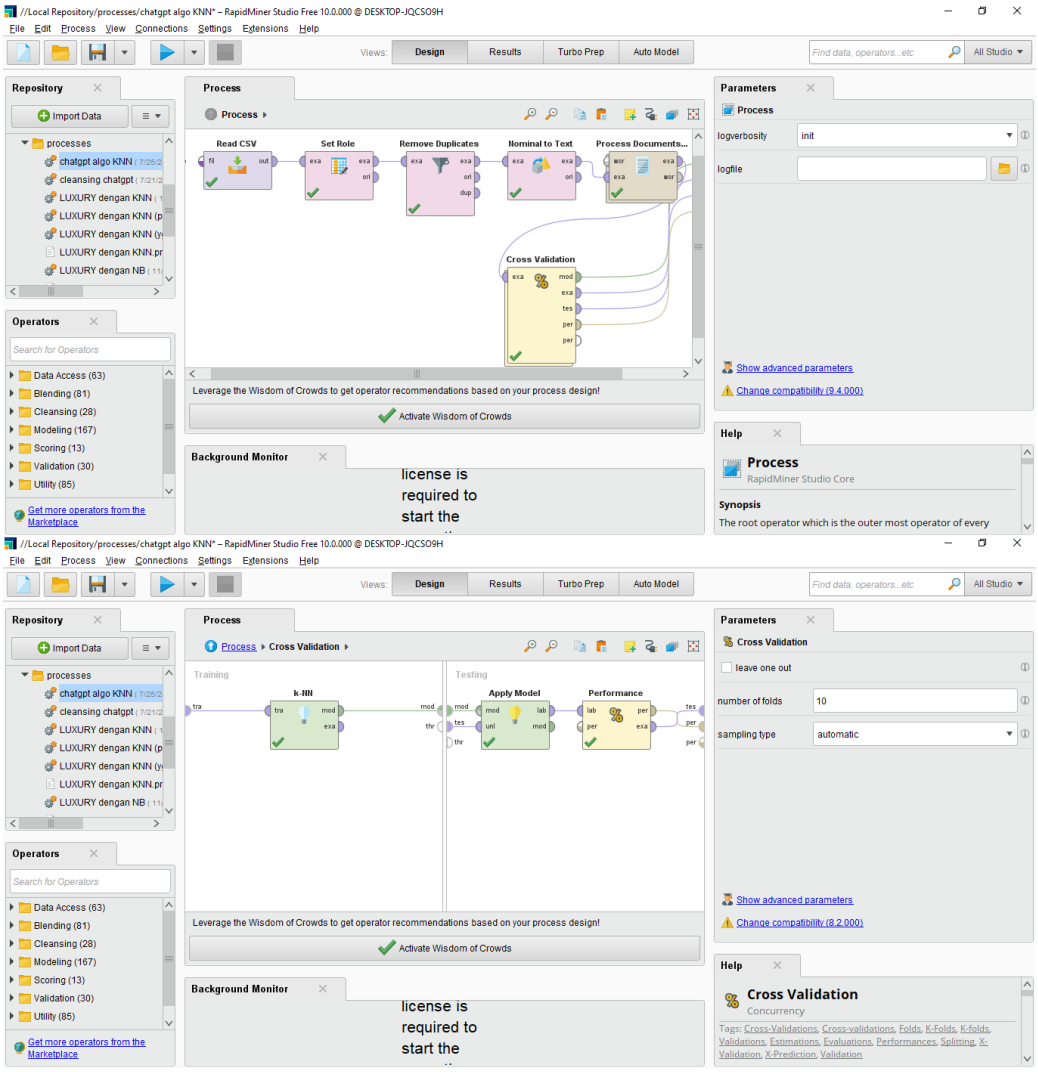
Pada tahapan ini peneliti akan melakukan *preprocessing* pada *dataset*. Hal ini ditunjukkan untuk menyiapkan data yang bersih dan data yang bebas dari *noise*. Pada proses ini juga untuk menghitung pembobotan kata untuk keperluan pada proses *modelling* nanti. Berikut proses *text preprocessing* dapat dilihat pada gambar 8. Setelah melawati tahapan *text preparation*, maka jumlah data X didapatkan menjadi 1000 data yang merupakan data bersih untuk dipakai pada tahap selanjutnya.



Gambar 8. Proses *Text Preprocessing*

Berikut proses *modelling* yang dilakukan di software *RapidMiner* yang ditunjukkan pada gambar 9.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E - ISSN: 2723-7079 | P - ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016



Gambar 9. Proses Modelling

3) Evaluasi

Pada tahap evaluasi, dilakukan pengujian terhadap model yang telah dibangun menggunakan algoritma *K-Nearest Neighbors* (K-NN). Evaluasi ini dilakukan untuk mengukur seberapa baik model dapat memprediksi sentimen berdasarkan dataset yang telah dibersihkan dan dilabeli. Salah satu metode evaluasi yang digunakan adalah *confusion matrix*, yang memberikan gambaran mengenai tingkat akurasi, presisi, *recall*, dan *specificity* dari model yang telah dihasilkan. *Confusion matrix* adalah matriks yang memperlihatkan hasil prediksi model berdasarkan empat metrik utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dalam penelitian ini, matriks ini digunakan untuk melihat bagaimana model K-NN mengklasifikasikan sentimen dalam dataset yang terdiri dari 1.000 data. Hasil pengujian menggunakan *confusion matrix* memberikan informasi mengenai seberapa baik model mengidentifikasi sentimen positif, negatif, atau netral

.

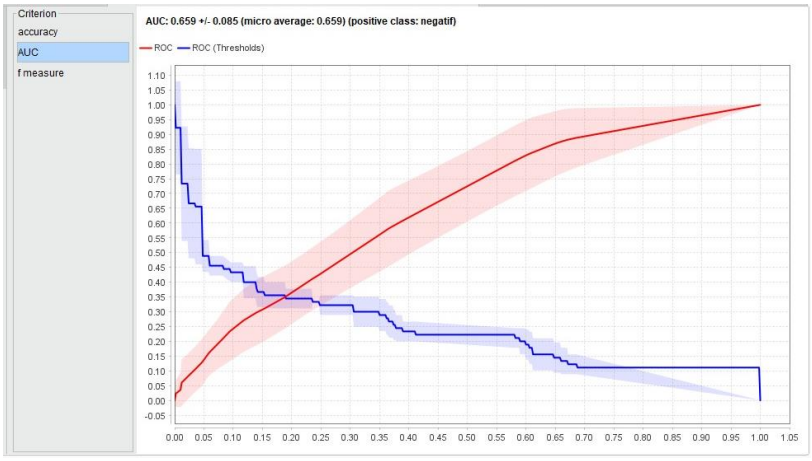
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

Gambar 10. Hasil Proses Akurasi

Gambar 10 menunjukkan tingkat akurasi dari model yang dibangun menggunakan algoritma K-NN, yaitu sebesar 84.80%. Nilai akurasi ini menunjukkan bahwa model dapat memprediksi dengan benar 84.80% dari total data dalam dataset. Akurasi yang tinggi untuk sentimen positif mengindikasikan bahwa model K-NN sangat efektif dalam mendeteksi sentimen yang dominan positif, namun menunjukkan kesulitan dalam mengklasifikasikan sentimen negatif. Dari tabel 1, terlihat bahwa model memiliki tingkat *precision* untuk kelas positif sebesar 85.80%, yang berarti 85.80% dari prediksi positif benar-benar positif. Untuk prediksi negatif, *precision* yang dihasilkan sebesar 38.10%, yang menunjukkan kelemahan dalam mengidentifikasi data negatif dengan benar. Nilai *recall* untuk sentimen positif mencapai 98.48%, sedangkan sentimen negatif memiliki *recall* sebesar 5.44%. *Recall* menunjukkan seberapa baik model mampu menemukan semua contoh yang benar untuk setiap kelas sentimen.

Tabel 1. Hasil Proses Akurasi

	True Positif	*True* Negatif	*Class Precision*
Pred. Positif	840	139	85.80%
Pred. Negatif	13	8	38.10%
Class Recall	98.48%	5.44%	



Gambar 11. Grafik *Area Under Curve* (AUC)

Berdasarkan pada Gambar 11 dapat dibuat kesimpulan bahwa *precision* menunjukkan tingkat ketepatan data yang diprediksi positif terhadap banyaknya data yang benar diprediksi positif yang menghasilkan *presentase* ketepatannya adalah 98.48% sedangkan untuk data bersentimen negatif memiliki *precision* sebesar 5.44%. Untuk sentimen negatif memiliki *recall (Specificity)* adalah 5.44% sehingga dapat disimpulkan bahwa model dapat menemukan kembali informasi atau data yang benar – benar negatif

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

dengan sangat baik, sedangkan pada sentimen positif memiliki *recall* sebesar 98.48%. Nilai *accuracy* yang dihasilkan menggunakan model *K- Nearest Neighbor* adalah 84.80%, sehingga dapat disimpulkan bahwa bahwa algoritma *K- Nearest Neighbor* dapat mengklasifikasi sentimen dengan baik menggunakan data *Chat* GPT.

3.1.4 *Deployment*

Tahap *deployment* merupakan langkah akhir dalam metode *Cross Industry Standard Process for Data Mining* (CRISP-DM). Pada tahap ini, hasil dari analisis data dan model yang telah dibangun diterapkan dan disajikan dalam bentuk laporan atau output yang bisa digunakan untuk pengambilan keputusan. Dalam konteks penelitian ini, *deployment* berfokus pada penyajian hasil analisis sentimen terhadap *ChatGPT* di media sosial "X", serta bagaimana hasil ini dapat dimanfaatkan oleh berbagai pemangku kepentingan. Tujuan utama dari tahap *deployment* adalah untuk memastikan bahwa hasil penelitian dapat digunakan secara efektif oleh para pengambil keputusan, pengembang teknologi, atau pemasar yang tertarik dengan persepsi masyarakat mengenai *ChatGPT*. Dengan menyelesaikan tahap *deployment*, hasil penelitian data mining dapat sepenuhnya dimanfaatkan untuk mendukung tujuan bisnis atau penelitian. Proses ini memastikan bahwa informasi yang dihasilkan dari analisis data benar-benar memberikan nilai dan membantu dalam pengambilan keputusan yang lebih efektif.

3.2 Pembahasan

Berdasarkan hasil penelitian ini, penggunaan algoritma *K-Nearest Neighbors* (KNN) dalam analisis sentimen terhadap *ChatGPT* di media sosial "X" memberikan hasil yang sejalan dengan beberapa penelitian sebelumnya, meskipun ada perbedaan dalam aspek metode dan hasil akhir. Dalam penelitian ini, KNN menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen positif dengan akurasi tinggi, sementara kemampuan untuk mendeteksi sentimen negatif masih terbatas. Temuan ini serupa dengan hasil penelitian Rifaldi *et al.* (2023), yang juga menggunakan data dari media sosial dan menemukan bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen positif secara efektif, tetapi kurang berhasil dalam mengidentifikasi sentimen negatif. Keterbatasan ini menunjukkan bahwa baik KNN maupun Naive Bayes menghadapi tantangan serupa dalam menangani data dengan sentimen negatif yang lebih kompleks.

Namun, studi yang menggunakan algoritma *Support Vector Machine* (SVM) seperti yang dilakukan oleh Pratama dan Hendry (2024) menunjukkan hasil yang lebih baik dalam mendeteksi sentimen negatif dibandingkan KNN. SVM, khususnya dengan pengoptimalan seperti yang diterapkan dalam penelitian Saepudin *et al.* (2024) dengan metode *Particle Swarm Optimization* (PSO), mampu meningkatkan akurasi secara keseluruhan, terutama dalam menangani sentimen yang lebih sulit terklasifikasikan. Hasil ini memberikan indikasi bahwa KNN memiliki keunggulan dalam kesederhanaan dan efisiensi untuk sentimen positif, namun SVM atau metode dengan optimasi tambahan bisa lebih tepat digunakan untuk dataset yang memiliki distribusi sentimen yang lebih beragam.

Dalam penelitian ini, mayoritas sentimen yang ditemukan terhadap *ChatGPT* bersifat positif, dan hal ini sejalan dengan penelitian yang dilakukan oleh Manurung *et al.* (2023) yang menemukan bahwa *ChatGPT* berdampak positif terhadap kemampuan berpikir mahasiswa di Universitas Atma Jaya. Temuan ini juga diperkuat oleh studi Darman (2024), di mana *ChatGPT* digunakan dalam penyelesaian masalah pertanahan dengan hasil yang positif. Kedua penelitian tersebut menunjukkan bahwa *ChatGPT* memberikan manfaat yang signifikan bagi penggunanya, baik dalam konteks akademik maupun profesional.

Meskipun demikian, beberapa penelitian lain, seperti Saraswati *et al.* (2023) dan Maula *et al.* (2023), mencatat adanya kekhawatiran terkait dampak negatif *ChatGPT*, terutama mengenai ketergantungan mahasiswa pada teknologi ini. Mereka menemukan bahwa penggunaan *ChatGPT* dapat memicu kemalasan dalam menyelesaikan tugas secara mandiri. Penelitian ini, meskipun tidak secara eksplisit menemukan banyak sentimen negatif, menunjukkan adanya presisi rendah pada

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

klasifikasi sentimen negatif, yang bisa jadi mencerminkan kesulitan dalam menangani nuansa negatif atau ketidakpuasan yang mungkin lebih tersirat dalam teks.

Penelitian memiliki beberapa kesamaan dengan studi lain dalam bidang pendidikan. Penelitian oleh Auna *et al.* (2023) dan Furqon Al Hadiq *et al.* (2023) mengkaji dampak *ChatGPT* dalam lingkungan pembelajaran dan menemukan bahwa penggunaannya dapat meningkatkan literasi digital dan efektivitas belajar. Meskipun konteks penelitian ini berbeda, dengan fokus pada media sosial "X", temuan yang menunjukkan dominasi sentimen positif terhadap *ChatGPT* mendukung pandangan bahwa teknologi ini dipersepsikan sebagai alat yang bermanfaat dalam berbagai bidang, termasuk pendidikan.

Namun, dalam aspek metodologi, penelitian ini menggunakan KNN yang lebih sederhana dibandingkan dengan algoritma Naive Bayes yang digunakan oleh Rifaldi *et al.* (2023) dan Purbayanto dan Suharsono (2024). Meskipun Naive Bayes menunjukkan performa yang baik dalam analisis sentimen, penelitian ini menunjukkan bahwa KNN juga mampu memberikan hasil yang memuaskan untuk sentimen positif, dengan tingkat akurasi yang sebanding. Perbedaan utama antara KNN dan Naive Bayes terletak pada cara masing-masing algoritma menangani distribusi data dan hubungan antara fitur, yang mungkin menjelaskan mengapa KNN kurang efisien dalam mengklasifikasikan sentimen negatif di penelitian ini. Meskipun KNN memberikan akurasi yang baik untuk sentimen positif, hasil penelitian menunjukkan perlunya mempertimbangkan algoritma lain, seperti SVM atau optimasi model, untuk meningkatkan performa dalam mengklasifikasikan sentimen negatif.

4. Kesimpulan

Berdasarkan hasil, instance yang diprediksi sebagai negatif (True Negative = 5,44%), sementara semua instance positif diprediksi dengan benar (True Positive = 98,48%). Nilai AUC sebesar 0,659 menunjukkan bahwa model ini memiliki kemampuan yang lebih baik daripada tebakan acak, tetapi tidak terlalu baik. Hasil sentimen masyarakat terhadap *ChatGPT* menunjukkan sentimen positif sebanyak 853 data, dan sentimen negatif sebanyak 147 data. Dari hasil tersebut, dapat disimpulkan bahwa sebagian besar masyarakat memiliki respons positif terhadap penggunaan *ChatGPT*. Berdasarkan klasifikasi model algoritma *K-Nearest Neighbor* terhadap dataset *ChatGPT*, didapatkan nilai akurasi sebesar 84,80%, sehingga dapat dikatakan bahwa algoritma K-NN dapat mengklasifikasikan kelas negatif dan positif secara baik dan benar. Penggunaan *Artificial Intelligence ChatGPT* semakin meningkat dan memberikan manfaat yang baik. Dengan menggunakan algoritma ini, dapat dikatakan bahwa algoritma K-NN dapat mengklasifikasikan data secara baik dan benar.

5. Ucapan Terima Kasih

Terima kasih kepada Sekolah Tinggi Komputer Cipta Karya Informatika serta semua pihak yang berjasa pada penelitian yang telah dilaksanakan sehingga bisa diselesaikan dengan baik dari awal persiapan hingga pelaksanaan penelitian ini.

6. Daftar Pustaka

Al Hadiq, M. F., & Ramadhan, C. U. (2023). Pengaruh model pembelajaran berbasis investigasi dengan dukungan ChatGPT terhadap keterampilan literasi digital siswa sekolah dasar. *COLLASE (Creative of Learning Students Elementary Education)*, *6*(6), 1187-1193. https://doi.org/10.22460/collase.v6i6.21673.

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

Amalia, S., Karomatan, T., & Ridwan, W. I. (2023). Pengaruh Penggunaan Chat Gpt Sebagai Content Creation Dalam Membangun Persepsi Konsumen Terhadap Strategi Pemasaran Umkm. *Jurnal Ilmiah Publika*, *11*(2), 525-533.

Anuar, S., Muda, M. S., & Mat, A. C. (2024). Penggunaan Kecerdasan Buatan Pelajar Kolej Komuniti Kemaman. *International Journal of Advanced Research in Education and Society*, *6*(1), 225-238. https://doi.org/10.55057/ijares.2024.6.1.19

Auna, H. S. A., & Hamzah, N. (2024). Studi Perspektif Siswa Terhadap Efektivitas Pembelajaran Matematika Dengan Penerapan Chatgpt. *HINEF: Jurnal Rumpun Ilmu Pendidikan*, *3*(1), 13-25. https://doi.org/10.37792/hinef.v3i1.1160.

Dani, R. W. (2023, December). Analisis Peran Artificial intelligence (AI) Dalam Penulisan karya Ilmiah Pada ChatGPT-3, 5 dari OpenAI. In *Proceeding of International Seminar on Adab and Humanities* (Vol. 5, No. 1, pp. 187-200).

Darman, R. (2024). Peran ChatGPT sebagai artificial intelligence dalam menyelesaikan masalah pertanahan dengan metode studi kasus dan black box testing. *Tunas Agraria*, 7(1), 18–46. https://doi.org/10.31292/jta.v7i1.256

Dewi, M. N., & Putra, R. E. (2024). Pengembangan Sistem Analisis Sentimen Masyarakat Terhadap Chatgpt Pada Twitter Dengan Perbandingan Metode Naive Bayes Classifier Dan K-Nearest Neighbors. *Journal of Informatics and Computer Science (JINACS)*, *5*(03), 344-357. https://doi.org/10.26740/jinacs.v5n03.p344-357.

Fahriza, M. N., & Riza, N. (2023). Analisis Sentimen pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer GPT menggunakan Metode Klasifikasi K-Nearest Neighbor (KNN). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(2), 1351-1358. https://doi.org/10.36040/jati.v7i2.6767.

Gaede, F. (2006). Marlin and LCCD—Software tools for the ILC. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *559*(1), 177-180. https://doi.org/10.1016/j.nima.2005.11.138

Handoyo, E. R., Sugiarto, J., Lolo, A., & Chai, K. (2023). Identifikasi Pengaruh Penggunaan ChatGPT terhadap Kemampuan Berfikir Mahasiswa di Universitas Atma Jaya Yogyakarta Prodi Sistem Informasi Angkatan 2021. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *3*(2), 342-352. https://doi.org/10.24002/konstelasi.v3i2.7241.

Irwansyah Saputra, D. A. K. (2022). Machine learning Untuk Pemula. *Informatika Bandung*.

Maula, S. R., Aprillian, S. D., Rachman, A. W., & Azman, M. N. M. (2023). Ketergantungan mahasiswa Universitas Jember terhadap artificial intelligence (AI). *ALADALAH: Jurnal Politik, Sosial, Hukum dan Humaniora*, 2(1), 1–14. https://doi.org/10.59246/aladalah.v2i1.608

Nur, M., Suryono, R. N., Bhagaskara, R. E., Pratama, M. A., & Pratama, A. (2023). *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2023 Surabaya*.

Oktavia, I., & Isnain, A. R. (2024). Analisis sentimen opini terhadap tools artificial intelligence (AI) berdasarkan Twitter menggunakan algoritma Naïve Bayes. *Jurnal Media Informatika Budidarma*. https://doi.org/10.30865/mib.v8i2.7524

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

Pakpahan, R. (2021). Analisa pengaruh implementasi artificial intelligence dalam kehidupan manusia. *Journal of Information System, Informatics and Computing Issue Period*, 5(2), 506–513. https://doi.org/10.52362/jisicom.v5i2.616

Pratama, A. D., & Hendry, H. (2024). Analisa sentimen masyarakat terhadap penggunaan ChatGPT menggunakan metode support vector machine (SVM). *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(1), 327–338. https://doi.org/10.29100/jipi.v9i1.4285

Pujiono, I. P., Rachmawanto, E. H., & Hana, F. M. (2024). Pengaruh Asisten Virtual Berbasis Artificial Intelligence Terhadap Integritas Sertifikasi Kompetensi Pemrograman secara Online. *JURNAL KRIDATAMA SAINS DAN TEKNOLOGI*, *6*(01), 34-46. https://doi.org/10.53863/kst.v6i01.1052.

Purbayanto, B., & Suharsono, T. N. (2023). Analisis Sentimen Pengguna X terhadap Chatgpt dengan Algoritme Naive Bayes. *Jurnal Telematika*, *18*(2), 63-71. https://doi.org/10.61769/telematika.v18i2.614.

Purnama, B., & Iwan, S. (2021). *Implementasi artificial intelligence dan machine learning*. Bandung: Informatika Bandung.

Purnama, B., & Sofana, I. (2021). *Implementasi artificial intelligence dan machine learning*. Bandung: Informatika Bandung.

Rahayu, S., Purnama, J. J., Hamid, A., & Hikmawati, N. K. (2023). Analisis Sentimen AicoGPT (Generative Pre-trained Transformer) Menggunakan TF-IDF. Jurnal Buana Informatika, 14(02), 97-106. https://doi.org/10.24002/jbi.v14i02.7039

Rifaldi, M. I., Ramadhan, Y. R., & Jaelani, I. (2023). Analisis Sentimen Terhadap Aplikasi Chatgpt Pada Twitter Menggunakan Algoritma Naïve Bayes. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, *7*(2), 802-814. http://dx.doi.org/10.30645/j-sakti.v7i2.687.

Risnina, N. N., Permatasari, S. T. I., Nurulhusna, A. Z., Anjelita, F. M., Wulaningtyas, C., & Rakhmawati, N. A. (2023). Pengaruh ChatGPT terhadap proses pembelajaran mahasiswa di Institut Teknologi Sepuluh Nopember. *Jurnal Pendidikan, Bahasa dan Budaya*, 2(4), 119–132. https://doi.org/10.55606/jpbb.v2i4.2364

Saepudin, A., Aryanti, R., Fitriani, E., Royadi, R., & Ardiansyah, D. (2024). Analisis Sentimen Pemanfaatan Artificial Intelligence di Dunia Pendidikan Menggunakan SVM Berbasis Particle Swarm Optimization. *Computer Science (CO-SCIENCE)*, *4*(1), 71-79. https://doi.org/10.31294/coscience.v4i1.2921

Sakti, E. M. S., & Herwanto, A. (2023). Pelatihan Penggunaan Chatgpt Terhadap Minat Baca Siswa SMA Muhammadiyah 1 Jakarta. *Jurnal Edukasi dan Multimedia*, *1*(3), 7-12. https://doi.org/10.37817/jurnaledukasidanmultimedia.v1i3

Saraswati, A. R., Karmina, V. A., Efendi, M. P., Candrakanti, Z., & Rakhmawati, N. A. (2023). Analisis pengaruh ChatGPT terhadap tingkat kemalasan berpikir mahasiswa ITS dalam proses pengerjaan tugas. *Jurnal Pendidikan, Bahasa dan Budaya*, 2(4), 40–48. https://doi.org/10.55606/jpbb.v2i4.2223

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.1016

Setiawan, D., Karuniawati, E. A. D., & Janty, S. I. (2023). Peran Chat Gpt (Generative Pre-Training Transformer) Dalam Implementasi Ditinjau Dari Dataset. *Innovative: Journal Of Social Science Research*, *3*(3), 9527-9539. https://doi.org/10.31004/innovative.v3i3.3286.

Sobari, A. A., & Syafrullah, M. (2023, October). PENERAPAN ALGORITMA K-NEAREST NEIGHBORS UNTUK MENGKLASIFIKASI SENTIEMEN MASYARAKAT TERHADAP KEBERADAAN CHAT GPT. In *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)* (Vol. 2, No. 2, pp. 836-845).

Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter. *Jambura Journal of Electrical and Electronics Engineering*, *3*(2), 78-83. https://doi.org/10.37905/jjeee.v3i2.10885

Wilie, D. P. (2023). Analisis sentimen opini publik terhadap ChatGPT di Twitter menggunakan metode Naive Bayes. Retrieved from https://www.kaggle.com/datasets/charunisa/ChatGPT-sentiment-analysis

Yaya, H., & Teguh, W. (2021). *Dasar-dasar deep learning dan implementasinya*. Yogyakarta: Gava Media Anggota IKAPI DIY.

Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN 4312418*. https://doi.org/10.2139/ssrn.4312418.