

SENTIMEN ANALISIS CHATGPT DENGAN ALGORITMA NAÏVE BAYES DAN OPTIMASI PSO

Lestari Yusuf¹; Siti Masripah^{2*}

Program Studi Sistem Informasi¹
Universitas Nusa Mandiri Jakarta¹
www.nusamandiri.ac.id¹
lestari.lyf@nusamandiri.ac.id²

Program Studi Sistem Informasi Akuntansi Kampus Kota Bogor^{2*}
Universitas Bina Sarana Informatika²
www.bsi.ac.id²
siti.stm@bsi.ac.id^{2*}



Abstract— ChatGPT which is an OpenAI technology that responds to conversations between humans and machines. enabling users of all ages and backgrounds to communicate naturally in multiple languages without having prior knowledge or experience in programming or the computer world. However, a technology will always be at odds and has flaws on the human side, various assumptions about chatGPT are formed from many sides, such as in the world of education, chatGPT creates parallels for teachers and lecturers. When giving assignments, students/students can use chatGPT as material in answering assignments from teachers/lecturers. And that results in students/students not carefully reading the answers to these assignments, if that continues to happen, students/students will find it too easy to get something and then will lose interest in solving problems with their own efforts. This article aims to analyze sentiment analysis whose data is taken from Twitter using the keyword "CahtGPT OpenAI". With 2,000 data calculated using the naive Bayes algorithm and optimized using PSO, it is found that sentiment analysis for chatGPT itself has an accuracy of 69.23% with a positive class of 0.503 and a negative of 0.497 and obtains an AUC curve value of 0.68 +/- 0.55.

Keywords: chatGPT, Sentiment-analysis, naive bayes, PSO

Abstrak— ChatGPT yang merupakan teknologi OpenAI yang merespon percakapan antar manusia dan mesin. memungkinkan pengguna dari segala usia dan latar belakang untuk berkomunikasi secara alami dalam berbagai bahasa tanpa memiliki pengetahuan atau pengalaman sebelumnya dalam pemrograman atau dunia komputer. Tetapi sebuah teknologi akan selalu berseberangan dan memiliki kekurangan pada sisi kemanusiaan, berbagai asumsi mengenai chatGPT terbentuk dari banyak sisi seperti dalam dunia pendidikan chatGPT menjadi buat simalakama untuk para guru dan dosen. Saat pemberian tugas, siswa/mahasiswa bisa menggunakan chatGPT sebagai bahan dalam menjawab tugas-tugas dari guru/dosen. Dan itu mengakibatkan siswa/mahasiswa tidak membaca dengan seksama jawaban dari tugas-tugas tersebut, jika hal itu terus terjadi, siswa/mahasiswa akan merasa terlalu mudah mendapatkan sesuatu kemudian akan kehilangan minatnya dalam berusaha menyelesaikan masalah dengan usaha diri sendiri. Penelitian ini bertujuan menganalisa sentimen analisis terhadap Open AI ChatGPT yang datanya diambil dari twitter menggunakan kata kunci "CahtGPT OpenAI". Dengan 2000 data yang diolah menggunakan algoritma naive bayes dan di optimalisasi menggunakan PSO didapatkan bahwa sentimen analisis untuk chatGPT sendiri memiliki akurasi 69.23% dengan class positif sebanyak 0.503 dan negatif sebanyak 0.497 dan mendapatkan nilai kurva AUC sebesar 0.68 +/- 0.55.

Kata kunci: chatGPT, analisis sentimen, naive bayes, PSO

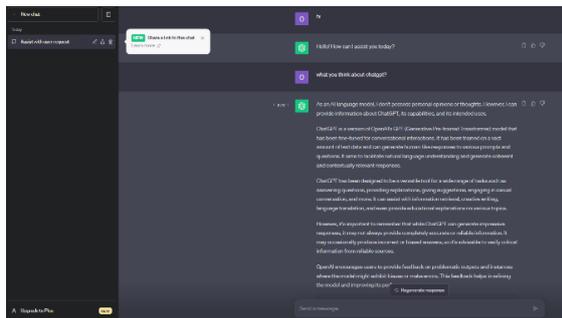
PENDAHULUAN

Teknologi AI yang saat ini berkaitan erat dengan dunia baru metaverse semakin menarik untuk dibahas, perkembangan AI semakin bertambah besar, dengan adanya teknologi baru yang hadir november tahun 2022 yang bernama chatGPT. ChatGPT yang merupakan teknologi

OpenAI yang merespon percakapan antar manusia dan mesin(Dowling & Lucey, 2023).

Gambar 1 merupakan tampilan chatGPT dimana chatGPT merupakan hasil penelitian machine learning yang memanfaatkan model deep neural network dari sebanyak 175 miliar parameter yang merupakan hasil pengembangan dari natural language processing(Leippold, 2023). ChatGPT

dapat di akses melalui <https://openai.com/blog/chatgpt>.



sumber: (Yusuf & Masripah, 2023)

Gambar 1. Tampilan ChatGPT

Dengan kemampuan mereka untuk memahami konteks, niat, sentimen, dan lainnya (Zhu et al., 2023), ChatGPT memungkinkan pengguna dari segala usia dan latar belakang untuk berkomunikasi secara alami dalam berbagai bahasa tanpa memiliki pengetahuan atau pengalaman sebelumnya dalam pemrograman atau dunia komputer (Zhou et al., 2023). ChatGPT dapat digunakan untuk memudahkan berbagai hal, mulai dari layanan pelanggan, hiburan, pendidikan, keuangan, *healthcare* dsb (Shaji George et al., 2023).

Tetapi sebuah teknologi akan selalu berseberangan dan memiliki kekurangan pada sisi kemanusiaan, berbagai asumsi mengenai chatGPT terbentuk dari banyak sisi seperti dalam dunia pendidikan chatGPT menjadi buah simalakama untuk para guru dan dosen (Savitri, 2023). Bukan hanya itu, lapangan pekerjaanpun terancam berkurang karena digantikan oleh openAI, begitupun dalam dunia pendidikan siswa dan mahasiswa dengan mudah menyelesaikan tugas dengan bantuan chatGPT dan itu menjadi dilema (Maulana, 2023).

Dalam artikel ini ingin menjelaskan mengenai penyelesaian masalah diatas menggunakan sentimen analisis chatGPT yang datanya diambil dari twitter, seperti yang pernah dilakukan (Syarifuddin, 2020) yang menggunakan twitter sebagai tempat mendapatkan data, pengambilan data melalui twitter juga dilakukan oleh (Fikri et al., 2020) yang melakukan pengambilan data "universitas", "muhammadiyah". Dan dalam penelitian ini data yang kami gunakan, menggunakan kata kunci "CahtGPT OpenAI". Sentimen analisis dapat membantu mengetahui positif dan negatif mengenai penggunaan chatGPT (Erfina, 2023).

Pada penelitian sebelumnya sentimen analisis digunakan untuk meneliti ulasan sebuah produk (Gunawan et al., 2018), meneliti sentimen analisis pada saat pandemi covid berlangsung (Samsir

et al., 2021) yang menggunakan metode naive bayes. Penelitian untuk sentimen analisis chatGPT pun pernah diteliti oleh (Erfina, 2023), metode Naive bayes pun baik untuk digunakan dalam masalah penelitian sentimen analisis seperti yang sebelumnya telah dibuat (Nugroho, 2018), begitupun (Duei Putri et al., 2022) ia pernah menggunakan metode naive bayes untuk menyelesaikan sentimen analisis kinerja dewan perwakilan rakyat yang menghasilkan akurasi yang baik, pada penelitian berikutnya (Astari et al., 2020) metode naive bayes disebutkan sangat stabil untuk digunakan sebagai metode yang menyelesaikan masalah sentimen analisis.

Maka dari itu penelitian ini ditujukan untuk menganalisis sentimen chatGPT menggunakan.

BAHAN DAN METODE

Pada penelitian ini, bahan yang digunakan dalam pengolahan data adalah bersumber dari Twitter. Tema yang diambil adalah tentang penggunaan AI ChatGPT. Komentar atau tweet dari twitter yang didapat, dilakukan pembersihan terlebih dahulu sehingga data dapat digunakan untuk diolah. Adapun tahapan pengolahan data yaitu Crawling data, Cleaning data, Labeling, Text processing, Klasifikasi dan Evaluasi, berikut penjelasannya :

1. Crawling data

Pada Crawling data yang digunakan dalam penelitian ini diawali dengan proses pengambilan data Twitter, yaitu menggunakan tool RapidMiner dengan menggunakan Retrieval Connection to Twitter dengan limit sebanyak 2000 data.

2. Cleaning

Proses Cleaning, yaitu menghapus kata-kata yang tidak digunakan seperti simbol, tagar, nama pengguna, URL, email (Undamayanti et al., 2022)

3. Labeling

Proses Labeling, yaitu memberikan label pada tweet apakah tweet bernada positif atau negative.

4. Text Processing

Tahap selanjutnya ada Text Processing agar data yang digunakan menjadi lebih baik untuk dilakukan Analisa selanjutnya.

- Proses Tokenize, untuk memisahkan satu kata dengan kata yang lainnya, dan satu kata akan menjadi atribut tersendiri.
- Transform Cases, untuk mengubah font dari tweet tersebut, maka semua tulisan akan diubah menjadi huruf kecil.

- c) Filter Stopwords, untuk menghapus kata-kata yang tidak penting seperti di, yang dan lain sebagainya.
- d) Filter tokens by length, akan membuang kata-kata yang terlalu pendek atau kata-kata yang terlalu Panjang yang tidak penting seperti 'yg, ya' dan lain sebagainya

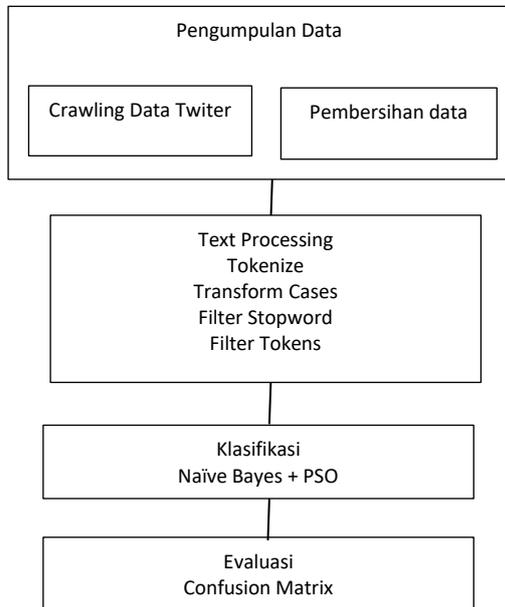
5. Klasifikasi

Tahap berikutnya adalah melakukan klasifikasi dengan algoritma Naïve bayes dan Optimasi PSO

6. Evaluasi

tahap evaluasi dilakukan menggunakan Confusion Matrix

Alur penelitian dijelaskan pada gambar 2 dibawah ini:



Sumber : (Yusuf & Masripah, 2023)
Gambar 2. Alur Penelitian

HASIL DAN PEMBAHASAN

A. Tahap Pengumpulan data

Pada tahap pengumpulan data, penulis mengambil data dari tweet masyarakat Indonesia terkait tentang maraknya penggunaan AI ChatGpt pada dunia Pendidikan. Pengambilan data pada bulan April 2023, sebanyak 2000 data. Crawling data dilakukan menggunakan tools Rapid miner. Pada tahap processing data yang didapat dari hasil crawling dibersihkan terlebih dahulu untuk menghilangkan data yang tidak diperlukan, sehingga data sebanyak 2000 setelah dibersihkan menghasilkan data sebanyak 949 data. Hasil dari hasil proses dokumen data dapat terlihat sejumlah

data 949 dan membentuk Hasil 2 special atribut dan 1.729 atribut regular seperti terlihat pada Tabel 1.

Tabel 1. Data hasil Crawling dan pembersihan data

n	sentim	text	abi	abisa	aca	aca
o	en		s	n	h	p	.
1	positif	contoh pengguna an chatgpt untuk mencari rujukan tulisan ilmiah	0	0	0	0
2	Positif	gpt update akan membawa perubahan revolusioner ini ke chatgpt	0	0	0	0
3	positif	buku panduan penggunaan an chatgpt artificial intelligence untuk membantu	0	0	0	0

Sumber : (Yusuf & Masripah, 2023)

Tabel 1 diatas merupakan hasil dari proses Crawling, cleaning dan labeling menggunakan tools rapidminer.

B. Tokenize

Untuk melakukan text processing pada penelitian ini, dilakukan tahapan yaitu tahapan tokenize, akan memisahkan kata dari kata-kata yang ada, dimana data yang dihasilkan dari proses tersebut dapat dilihat pada gambar 3. Kata "ChatGPT" muncul sebanyak 356 kali, muncul dalam 351 tweet. Pada perhitungan dengan Teknik TF-IDF yaitu Teknik untuk menghitung berapa kali suatu kata tersebut muncul dalam sebuah dokumen. Proses tokenize dijelaskan pada Tabel 2.

Tabel 2. Proses Tokenize

Word	Attribute name	Total Occurances	Document Occurrences
ChatGPT	ChatGPT	356.0	351.0
chatgpt	chatgpt	289.0	282.0
di	di	103.0	95.0
yang	yang	97.0	90.0
ini	ini	91.0	89.0
chatGPT	chatGPT	86.0	84.0

Sumber: (Yusuf & MSasripah, 2023)

C. Transform Cases

Melakukan perubah font dari tweet tersebut, maka semua tulisan akan diubah menjadi huruf kecil. Proses dijelaskan pada Tabel 3.

Tabel 3. Proses Transform Cases

Word	Attribute name	Total Occurances	Document Occurances
chatgpt	chatgpt	787.0	768.0
di	di	112.0	104.0
ini	ini	106.0	104.0
yang	yang	105.0	98.0
bisa	bisa	94.0	92.0
ai	ai	86.0	83.0

Sumber: (Yusuf & Masripah, 2023)

D. Filter Stopword

Melakukan penghapusan kata-kata tidak penting seperti di, yang dan lain sebagainya, sehingga akan terlihat seperti tabel 4 :

Tabel 4. Proses Filter Stopword

Word	Attribute name	Total Occurances	Document Occurances
chatgpt	chatgpt	787.0	768.0
ai	ai	86.0	83.0
yg	yg	73.0	63.0
pake	pake	54.0	54.0
ya	ya	53.0	51.0
kalo	kalo	41.0	39.0

Sumber:(Yusuf & Masripah, 2023).

Tabel 4 diatas menunjukkan proses filter stopwords setiap kata-kata pada setiap tweet yang sudah diberikan label.

E. Filter Tokens

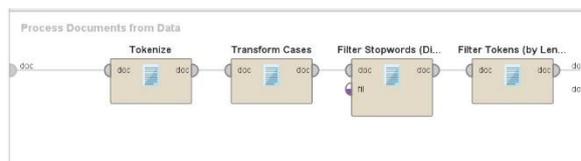
Akan melakukan pembuangan kata yang kurang dari 4 dan kata sebanyak lebih dari 25 kata, sehingga hasil yang didapatkan sebagai berikut, kata "yg", "ya" akan hilang, proses dijelaskan pada Tabel 5.

Tabel 5. Proses Filter Tokens

Word	Attribute name	Total Occurances	Document Occurances
chatgpt	chatgpt	787.0	768.0
pake	pake	54.0	54.0
kalo	kalo	41.0	39.0
bikin	bikin	34.0	33.0
pakai	pakai	33.0	33.0
banget	banget	29.0	28.0

Sumber: (Yusuf & Masripah, 2023)

Berikut langkah-langkah diatas yang diimplementasikan kedalam tools rapidminer, dijelaskan pada gambar 3:



Sumber : (Yusuf & Masripah, 2023)

Gambar 3. Langkah-langkah pre-processing data

Setelah semua pemrosesan teks dilakukan, maka kata-kata tersebut sudah siap digunakan untuk dilakukan analisis dengan algoritma yang akan digunakan, pada penelitian ini algoritma yang akan digunakan adalah Naïve bayes dan Optimasi PSO.

1. Naïve Bayes

Data yang sudah ada, diterapkan menggunakan algoritma klasifikasi Naïve Bayes, mendapatkan hasil akurasi sebesar 67.97%. Terlihat pada Tabel Accuracy Naïve Bayes, hasil dijelaskan pada table 6:

Tabel 6. Hasil perhitungan algoritma Naïve Bayes

	True positif	True negative	Class precision
Pred. positif	375	202	64.99%
Pred. negative	102	270	72.58%
Class recall	78.62%	57.20%	

Sumber: (Yusuf & Masripah, 2023)

Untuk nilai distribusi model pada atribut sentiment menghasilkan Class Positif 0.503 dan Class negative 0.497 dari 1729 atribut yang terbentuk terlihat pada Tabel Simple Distribution, dijelaskan pada Tabel 7 :

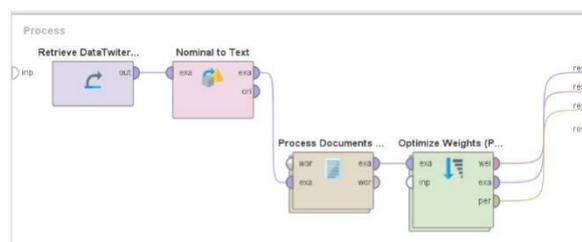
Tabel 7. Simple Distribution

Nilai	1729 distributions
Class Positif	0.503
Class Negatif	0.497

Sumber : (Yusuf & Masripah, 2023)

2. Naïve Bayes dengan optimasi PSO

Pengoptimalisasian PSO pada rapidminer digambarkan dalam bentuk yang dijelaskan pada gambar 4:



Sumber: (Yusuf & Masripah, 2023)

Gambar 4. Proses optimalisasi menggunakan PSO

Gambar 4 merupakan gambar penggunaan optimasi dengan PSO. Penggunaan optimasi dengan PSO meningkatkan nilai akurasi dimana mendapatkan nilai akurasi sebesar 69.23% terlihat pada Tabel Accuracy Naïve Bayes dengan Optimasi PSO :

Tabel 8. Accuracy Naïve Bayes dengan Optimasi PSO

	True positif	True negative	Class precision
Pred. positif	383	198	65.92%
Pred. negative	94	274	74.46%
Class recall	80.29%	58.05%	

Sumber: (Yusuf & Masripah, 2023)

Dengan nilai precision yang didapat sebesar : 74.81%. Berikutnya mencari nilai *precision* untuk perhitungan tersebut, digambarkan pada tabel 9 dibawah ini:

Tabel 9. Nilai Precision

	True positif	True negative	Class precision
Pred. positif	383	198	65.92%
Pred. negative	94	274	74.46%
Class recall	80.29%	58.05%	

Sumber : (Yusuf & Masripah, 2023)

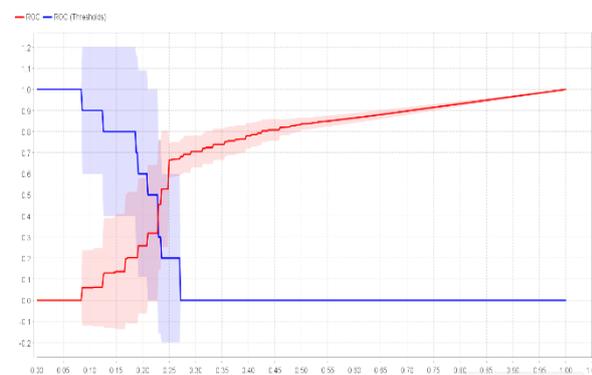
Dengan nilai Recall : 58.05%, terlihat pada tabel 10

Tabel 10. Nilai Recall

	True positif	True negative	Class precision
Pred. positif	383	198	65.92%
Pred. negative	94	274	74.46%
Class recall	80.29%	58.05%	

Sumber : (Yusuf & Masripah, 2023)

Dengan nilai tersebut kami juga mendapatkan nilai AUC dan kurva ROC sebesar AUC: 0.682 +/- 0.055 (micro average: 0.682) (positive class: negatif) dijelaskan oleh gambar 5:



Sumber : (Yusuf & Masripah, 2023)

Gambar 5. Curva AUC

KESIMPULAN

Dengan dibuatnya sentimen analisis openAI chatGPT menggunakan 1729 atribut bersih dari sebanyak 2000 data yang dihitung menggunakan metode naive bayes dan di optimalisasi menggunakan Pso didapatkan bahwa sentimen analisis untuk chatGPT sendiri memiliki akurasi 69.23% dengan class positif sebanyak 0.503 dan negatif sebanyak 0.497 dan mendapatkan nilai kurva AUC sebesar 0.68 +/- 0.55. dengan kesimpulan ini didapatkan sentimen analisis untuk chatGPT dengan nilai positif. Pada penelitian berikutnya disarankan agar dapat lebih detail pada proses normalisasi data, dikarenakan pada saat ini komentar-komentar yang ada di media sosial kadang menggunakan kata dengan arti negatif untuk mengutarakan hal positif akan suatu hal atau sebaliknya yang mempengaruhi makna dari setiap kalimat itu sendiri.

REFERENSI

Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27-29. <https://doi.org/10.30864/jsi.v15i1.332>

Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, 53(January), 103662. <https://doi.org/10.1016/j.frl.2023.103662>

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1), 34-40. <https://doi.org/10.23960/jitet.v10i1.2262>

Erfina, A. (2023). Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language. 2-11. <https://doi.org/10.56294/dm202345>

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71-76. <https://doi.org/10.32664/smatika.v10i02.455>

Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113.

- <https://doi.org/10.26418/jp.v4i2.27526>
Leippold, M. (2023). Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters*, 53(December 2022), 103617. <https://doi.org/10.1016/j.frl.2022.103617>
- Maulana, A. (2023). *Perguruan Tinggi Perlu Sikapi Peluang dan Risiko ChatGPT bagi Pembelajaran*. <https://www.unpad.ac.id/2023/02/perguruan-tinggi-perlu-sikapi-peluang-dan-risiko-chatgpt-bagi-pembelajaran/>
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 200. <https://doi.org/10.30645/j-sakti.v2i2.83>
- Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrighthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>
- Savitri, D. (2023). Praktisi AI Ungkap Peluang & Risiko Penggunaan ChatGPT di Perguruan Tinggi. *Detik.Com*. <https://www.detik.com/edu/perguruan-tinggi/d-6610414/praktisi-ai-ungkap-peluang--risiko-penggunaan-chatgpt-di-perguruan-tinggi>
- Shaji George, A., Hovan George, A., & Martin, Asg. (2023). *Partners Universal International Innovation Journal (PUIJ) A Review of ChatGPT AI's Impact on Several Business Sectors*. February, 9–23. <https://doi.org/10.5281/zenodo.7644359>
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Yusuf, L., & Masripah, S. (2023). *Pra-processing Text Sentimen analisis chatGPT dengan Algoritma Naive bayes dan Optimasi PSO*.
- Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023). *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*. 1–99. <http://arxiv.org/abs/2302.09419>
- Zhu, J. J., Jiang, J., Yang, M., & Ren, Z. J. (2023). ChatGPT and Environmental Research. *Environmental Science and Technology*, 1–4. <https://doi.org/10.1021/acs.est.3c01818>