

PENGEMBANGAN APLIKASI DETEKSI TEKS MENGUNAKAN ALGORITMA REGRESI LOGISTIK

Angginy Akhirunnisa Siregar, Zulfahmi Indra, Arnita, Insan Taufik, Debi Yandra Niska

Ilmu Komputer, Universitas Negeri Medan
Jalan Willem Iskandar Medan, Indonesia
angginyans@mhs.unimed.ac.id

ABSTRAK

Penggunaan kecerdasan buatan khususnya ChatGPT semakin meningkat dalam penyelesaian tugas akademik oleh siswa. Fenomena ini menimbulkan tantangan baru bagi dunia pendidikan, terutama dalam memastikan keaslian dan integritas karya tulis siswa. Penelitian ini bertujuan untuk mengembangkan aplikasi deteksi teks berbasis algoritma Regresi Logistik guna membantu pendidik dalam memverifikasi teks yang dihasilkan oleh AI dan non-AI. Data penelitian diperoleh dari jawaban siswa yang dibagi menjadi dua kelompok, yaitu teks AI dan non-AI, serta didukung oleh data tambahan dari berita, buku, dan artikel. Evaluasi model menunjukkan bahwa algoritma Regresi Logistik mampu mencapai akurasi sebesar 87,2%, presisi 87,5%, recall 86,1%, dan F1-score 86,7%. Aplikasi ini dikembangkan menggunakan framework Streamlit agar dapat digunakan secara interaktif, baik untuk input manual maupun file. Meskipun aplikasi menunjukkan performa yang baik, masih terdapat keterbatasan dalam mendeteksi teks formal atau kompleks yang ditulis oleh pakar. Oleh karena itu, diperlukan pengembangan lebih lanjut untuk meningkatkan akurasi deteksi pada teks kompleks.

Kata kunci : ChatGPT, Deteksi Teks, Natural Language Processing, Regresi Logistik, Streamlit

1. PENDAHULUAN

Kecerdasan buatan semakin banyak digunakan untuk berinovasi, hal ini semakin tampak dalam pembuatan tugas siswa [1]. ChatGPT dan teknologi NLP lainnya memiliki potensi untuk meningkatkan penulisan akademik dan efisiensi penelitian. Namun, penggunaannya juga menimbulkan kekhawatiran tentang dampak pada keaslian dan kredibilitas karya akademik [2]

Berdasarkan hasil penelitian [3] rata-rata siswa menyampaikan empat alasan utama penggunaan ChatGPT untuk menyelesaikan tugas siswa. Alasan yang paling sering disebut adalah: untuk mencari konten yang menarik bagi mereka, untuk menganalisis topik secara lebih mendalam, dan di urutan ketiga (dipilih oleh hampir 50% responden) adalah untuk menulis tugas berupa teks (seperti esai, makalah seminar, ulasan, dll). Penggunaan yang terakhir bisa dianggap tidak etis, terutama jika hanya menyalin-tempel tanpa berpikir kritis [3]. Jurnal tersebut memaparkan hasil survei terkait sejauh mana penggunaan ChatGPT dianggap bisa diterima dalam mengerjakan tugas. Mayoritas siswa menganggap tidak pantas jika menyalin lebih dari setengah isi tugas secara langsung.

Departemen Pendidikan Kota New York mengumumkan larangan penggunaan ChatGPT. Namun, ChatGPT masih dapat diakses berdasarkan permintaan khusus untuk kelas yang mempelajari kecerdasan buatan. Jenna Lyle, juru bicara departemen, menjelaskan bahwa keputusan pelarangan ini didasari oleh kekhawatiran tentang dampak negatif pada pembelajaran siswa. Dalam pernyataan email yang dimuat di platform berita NBC News, Lyle menyampaikan, "*While the tool may be able to provide quick and easy answers to questions, it*

does not build critical-thinking and problem-solving skills, which are essential for academic and lifelong success," ("Meskipun alat tersebut mampu memberikan jawaban dari pertanyaan secara cepat dan mudah, alat tersebut tidak membangun keterampilan berpikir kritis dan pemecahan masalah, yang mana penting untuk kesuksesan akademik dan seumur hidup,") [4].

Seorang siswa Hingham High School menerima hukuman berupa nilai nol pada tugas esai kelas sejarah dan D untuk nilai akhir, yang menyebabkan nilai semesternya turun menjadi C setelah diduga menggunakan alat kecerdasan buatan untuk meneliti dan membuat *outline* esai kelas Sejarah [5]. Berdasarkan student handbook code of discipline yang disediakan oleh Hingham High School, penggunaan teknologi yang tidak sah, termasuk Kecerdasan Buatan (AI) pada penilaian dianggap melanggar Integritas Akademik: Kecurangan dan Plagiarisme [6].

Penerapan kecerdasan buatan di sekolah menimbulkan tantangan besar, terutama untuk penilaian tugas siswa. Siswa dari berbagai bidang saat ini sering menggunakan ChatGPT untuk menyelesaikan tugas, hal ini menjadi tantangan bagi para pendidik yang kesulitan memverifikasi keaslian tugas yang dikumpulkan [7]. Verifikasi keaslian tugas yang dikumpulkan siswa dibutuhkan guna mencapai nilai etika pada transparansi, kredibilitas, serta kekayaan intelektual tugas siswa.

Jurnal [8] mengungkapkan dalam dua studi percobaan bahwa guru pemula dan guru berpengalaman tidak dapat mengidentifikasi teks yang dihasilkan oleh ChatGPT dari teks yang ditulis oleh siswa. Namun, terdapat beberapa indikasi bahwa guru yang lebih berpengalaman membuat penilaian yang lebih bervariasi dan lebih akurat. [9] mengusulkan

model kecerdasan buatan untuk membedakan antara teks yang ditulis oleh manusia dan teks yang dihasilkan oleh AI dengan studi kasus teks dari Wikipedia dan ChatGPT. Jurnal tersebut menampilkan hasil algoritma Regresi Logistik dengan akurasi 95%, algoritma Regresi Logistik lebih unggul dibandingkan algoritma Naive Bayes, SVM, Random Forest, dan CNN yang mempunyai tingkat akurasi 89%, 93%, 88%, dan 61% pada model dataset.

Kesuksesan algoritma Regresi Logistik juga dapat terlihat pada jurnal [10] dalam mendeteksi *Cyberbullying*, jurnal tersebut menghasilkan akurasi sebesar 92,52%. Selain itu, pada jurnal [11] dalam kasus mendeteksi email sekali pakai, Regresi Logistik menghasilkan akurasi 97,1%. Berdasarkan jurnal-jurnal tersebut, penulis memutuskan untuk membangun aplikasi untuk mendeteksi teks menggunakan algoritma Regresi Logistik guna membantu tenaga pengajar dalam memverifikasi keaslian tugas yang dikumpulkan oleh siswa pada label AI dan non-AI.

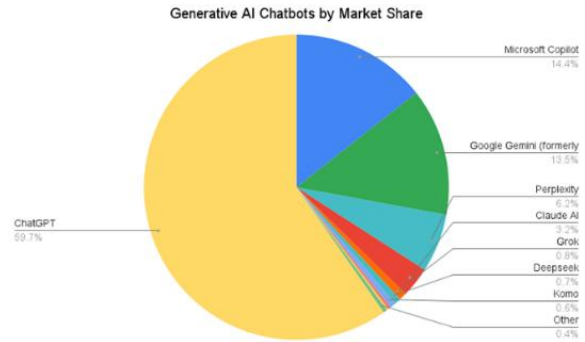
2. TINJAUAN PUSTAKA

2.1. Natural Language Processing (NLP)

Menurut [12] *Natural Language Processing* (NLP) telah muncul sebagai salah satu bidang penerapan kecerdasan buatan yang menjanjikan. *Natural Language Processing* (NLP) adalah bidang penelitian di mana suatu bahasa diproses untuk memahami aspek tata bahasa, makna, dan emosi. NLP telah menjadi bidang penelitian yang terus berkembang yang menghubungkan ilmu komputer dan linguistik.

Bahasa yang secara umum digunakan manusia dalam berkomunikasi satu sama lain disebut bahasa natural atau alami. Pada hakikatnya, bahasa selalu memiliki bentuk teks karena ketika manusia ingin menyampaikan gagasan, mereka akan menggunakan bentuk teks tertentu. Dalam proses menyampaikan gagasan tersebut, manusia akan berusaha agar teks yang digunakan mengandung bentuk-bentuk bahasa yang sesuai dengan maksud yang ingin disampaikan. Bentuk-bentuk bahasa tersebut merupakan sistem linguistik yang terdapat dalam teks, di mana bentuk teks dan pilihan unsur kebahasaan yang digunakan akan berbeda tergantung pada gagasan yang ingin diungkapkan [15].

Seiring perkembangan teknologi, kini muncul pula teks buatan mesin, yaitu teks yang dihasilkan oleh kecerdasan buatan (AI) dengan memanfaatkan teknologi *Natural Language Processing* (NLP) yang dilatih menggunakan data dari berbagai sumber.

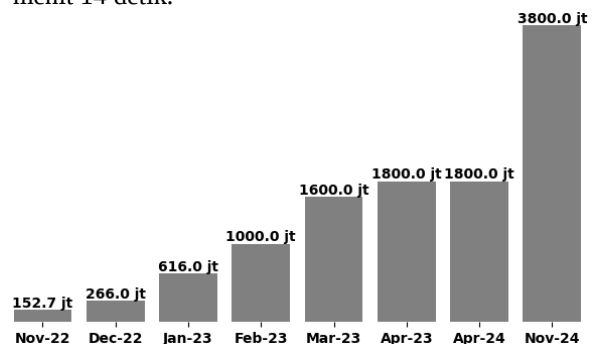


Gambar 1. Chatbot GenAI berdasarkan pangsa pasar

Grafik pada Gambar 1 memberikan gambaran mengenai pangsa pasar Chatbot GenAI. Dari data yang ada, ChatGPT menempati posisi teratas dengan pangsa pasar sebesar 59,7% [16]. Keberhasilan ChatGPT dalam meraih pangsa pasar yang signifikan ini menunjukkan bahwa kebanyakan pengguna lebih memilih menggunakan ChatGPT dibandingkan chatbot GenAI lainnya.

2.2. ChatGPT

ChatGPT pertama kali diperkenalkan pada tahun 2018 [17]. Pertumbuhan pengguna ChatGPT dapat dilihat pada grafik batang gambar 2, pertumbuhan tertinggi terjadi diantara bulan april 2024 dan bulan november 2024 dengan jumlah peningkatan pengguna sebesar 2000 juta. Berdasarkan artikel yang dimuat [18] rata-rata durasi pengunjung ChatGPT adalah 6 menit 14 detik.



Gambar 2. Pertumbuhan pengguna ChatGPT

Indonesia merupakan pengguna terbesar ke-lima setelah US, India, Brazil, dan UK dengan jumlah 4% [18].

2.3. Cross Validation

Cross-validation adalah salah satu metode pemisahan data yang digunakan untuk mengevaluasi kinerja umum model [19]. Data akan dipecah menjadi beberapa *fold* dan model akan dilatih untuk setiap *fold* yang tersedia. Sehingga, memungkinkan dataset digunakan secara efektif oleh model untuk pelatihan dan validasi.

2.4. Evaluasi Kinerja Model

Akurasi, presisi, recall, dan F1 score merupakan beberapa parameter untuk *performance evaluation* yang paling umum dan banyak digunakan [20], [21].

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

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

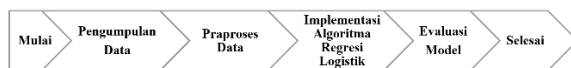
$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

2.5. Streamlit

Streamlit merupakan *framework* terbuka yang digunakan untuk membangun UI interaktif untuk aplikasi data *science* [22]. Streamlit dirancang untuk mempermudah pengembangan aplikasi web dengan menggunakan Python.

3. METODE PENELITIAN



Gambar 3. Alur Penelitian

3.1. Pengumpulan Data

Data merupakan hasil jawaban dari siswa yang dibagi menjadi dua kelompok dengan pertanyaan-pertanyaan yang sama. Pengumpulan data dilakukan menggunakan formulir daring Google untuk kedua kelompok. Kelompok pertama akan menjawab 15 soal menggunakan buku, website terpercaya, dan juga hasil pemikiran siswa tersebut, sedangkan kelompok kedua akan menjawab 15 soal menggunakan alat bantu ChatGPT.

Tabel 1. Distribusi Populasi dan Sampel

Kelas	Populasi	Sampel	
		AI	Manusia
X	52	5	5
XI	61	16	16
XII	43	5	7
Jumlah	156	26	28

Selain menggunakan data utama, peneliti juga menggunakan data tambahan berupa berita, dan juga buku sebagai data pendukung. Data pendukung sebanyak 34 dokumen teks untuk AI dan 35 dokumen teks untuk non-AI.

3.2. Praproses Data

Setelah melalui proses pembersihan, data yang sudah bersih selanjutnya diolah sesuai tahap berikut:

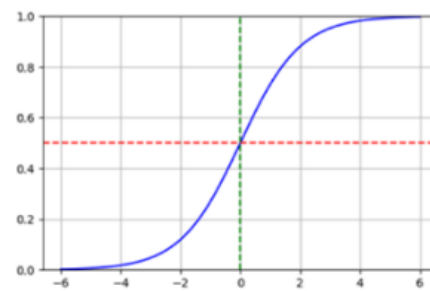
- Pembersihan data: Fokus proses ini adalah identifikasi dan memperbaiki kesalahan dan bias dengan mengisi atau melengkapi nilai yang hilang atau menghapus hal-hal yang tidak bisa

diperbaiki dan dilengkapi, menyelesaikan masalah ketidakkonsistenan dan standarisasi format, melakukan penanganan yang tepat untuk data yang tidak lazim [23].

- Normalisasi teks: Tahapan normalisasi teks dilakukan dengan memanfaatkan data kamus bahasa sebagai sekumpulan informasi untuk algoritma melakukan proses tersebut [24].
- Case Folding: Tujuan dari tahapan ini adalah untuk mempermudah komputer dalam membaca dan mengolah data [25].
- Tokenisasi: Tokenisasi bertujuan untuk mempermudah komputer memahami dan memproses informasi pada teks [26].
- Stopword removal: *Stopword* adalah kata-kata umum dalam bahasa yang tidak memberikan banyak makna kontekstual pada kalimat, seperti “dan”, “yang”, atau “adalah” [27].

3.3. Implementasi Algoritma Regresi Logistik

Tujuan algoritma regresi logistik biner adalah memprediksi salah satu dari dua kemungkinan hasil (biasanya 0 dan 1, atau negatif dan positif). Dengan kata sederhana, *variable terikat* (dependen) bersifat biner yang memiliki data yang dikodekan sebagai 1 (sukses/AI) atau 0 (gagal/non-AI).



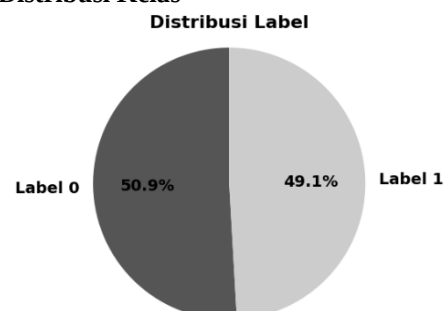
Gambar 4. Kurva sigmoid [28]

3.4. Evaluasi Model

Mengacu pada peluang dasar 50:50, model memiliki nilai akurasi awal (baseline) sebesar 50%. Model dianggap tidak memadai jika hasil evaluasi menunjukkan akurasi di bawah 70%, dan proses akan kembali ke tahap praproses. Sebaliknya, jika akurasi mencapai 70% atau lebih, model tersebut layak digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Distribusi Kelas



Gambar 5. Distribusi Kelas

Kolom ‘label’ digunakan untuk mengamati distribusi kelas AI (1) dan non-AI (0) dengan menggunakan `.value_counts()`.

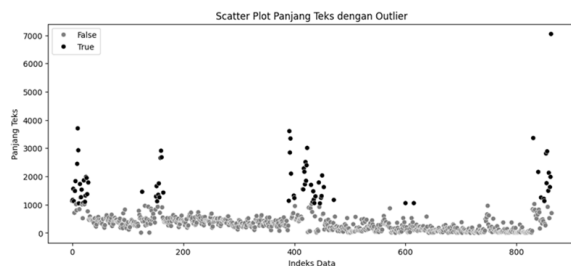
Setelah menghitung jumlah setiap label pada program, peneliti menggunakan grafik pie untuk menunjukkan distribusi kelas. Dari visualisasi ini, dapat dilihat bahwa:

- Teks non-AI (label 0) memiliki presentase 50.9 % atau dengan jumlah 440 sampel.
- Teks AI (label 1) memiliki presentase 49.1% atau dengan jumlah 424 sampel.

Perbedaan antara jumlah teks non-AI dan teks AI tidak terlalu signifikan, yang menunjukkan distribusi kelas yang relatif seimbang.

4.2. Mengetahui Permasalahan Data

- Missing Value: Tidak terdapat missing value pada seluruh label.
- Data Noise: Tidak terdapat noise pada seluruh label.
- Data Duplikat: Terdapat 42 data duplikat pada label teks, seluruh duplikat berasal dari label non-AI.
- Data Outlier



Gambar 6. Persebaran Outlier

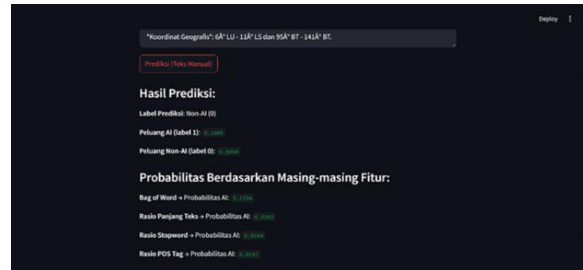
Terdapat 77 data outlier yang ditandai oleh titik hitam (True), pada teks AI 49 dan 28 pada teks non-AI, outlier terpanjang adalah 7048 huruf pada label non-AI dan outlier terpendek adalah 1047 huruf pada label non-AI.

4.3. Implementasi Model



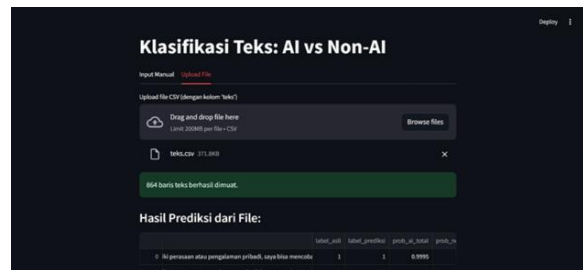
Gambar 7. Menu Utama

Pengguna dapat memilih antara menu input teks tunggal (secara manual) atau input teks sekaligus (file) pada menu utama.



Gambar 8. Deteksi Teks Manual (Tunggal)

Jika pengguna menggunakan halaman input manual, pengguna dapat langsung mengetik teks yang hendak diperiksa dan menekan tombol Prediksi (Teks Manual). Sistem akan melakukan pengolahan teks tersebut dan memberikan hasil prediksi berupa label AI dan non-AI. Selanjutnya, sistem akan memberikan nilai peluang setiap fitur berdasarkan teks yang pengguna berikan.



Gambar 9. Deteksi Teks File

Jika menggunakan halaman Upload file, pengguna dapat menginput file berupa csv dengan catatan file harus memiliki kolom dengan label “teks”. Sistem akan mengolah seluruh teks dan memberikan hasil yang dapat diunduh melalui tombol unduh yang telah disediakan.

4.4. Uji Teks Baru

Tabel 2. Hasil Prediksi Teks

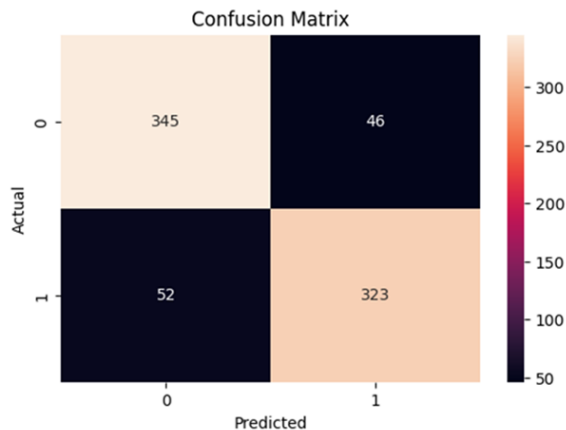
Teks	Keterangan	Label Asli	Hasil Prediksi
Teks 1	Wikipedia	Non-AI	Non-AI
Teks 2	Wikipedia	Non-AI	Non-AI
Teks 3	Berita	Non-AI	AI
Teks 4	Sosial Media	Non-AI	Non-AI
Teks 5	Buku	Non-AI	Non-AI
Teks 6	ChatGPT	AI	AI
Teks 7	ChatGPT	AI	Non-AI
Teks 8	Gemini	AI	AI
Teks 9	Gemini	AI	Non-AI
Teks 10	Microsoft Copilot	AI	AI

Sistem beberapa kali mengalami kesulitan dalam mendeteksi teks yang ditulis oleh pakar dengan gaya bahasa yang formal dan kompleks. Pada Teks 7 dan Teks 9, peneliti menggunakan ChatGPT dan Gemini untuk menghasilkan teks bernuansa emosional, yang kemudian diprediksi sebagai non-AI. Di mana teks

yang ditulis oleh manusia cenderung memiliki muatan emosional yang lebih kuat dibandingkan dengan teks yang dihasilkan oleh AI.

4.5. Evaluasi Kinerja Model

Evaluasi kinerja model digunakan untuk mengetahui model sudah layak digunakan atau masih belum layak digunakan.



Gambar 10. Confusion Matrix

Matriks tersebut digunakan untuk menunjukkan performa model dalam melakukan klasifikasi terhadap teks AI dan non-AI, model menunjukkan hasil TP = 323, FP = 46, TN = 345, dan FN = 52. Berdasarkan rumus setiap parameter pada 2.12., diketahui nilai akurasi, presisi, recall, dan F1-Score model sebagai berikut:

$$\text{Akurasi} = \frac{323+345}{323+345+46+52} = \frac{668}{766} \approx 0.872 \quad (5)$$

$$\text{Presisi} = \frac{323}{323+46} = \frac{323}{369} \approx 0.875 \quad (6)$$

$$\text{Recall} = \frac{323}{323+52} = \frac{323}{375} \approx 0.861 \quad (7)$$

$$F1 = 2 \times \frac{0.875 \times 0.861}{0.875 + 0.861} = 2 \times \frac{0.753}{1.736} \approx 0.867 \quad (8)$$

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa pengembangan aplikasi deteksi teks menggunakan algoritma Regresi Logistik berhasil membedakan teks yang dihasilkan oleh AI dan teks yang ditulis oleh manusia (non-AI) dengan tingkat akurasi sebesar 87,2%, presisi 87,5%, recall 86,1%, dan F1-score 86,7%. Aplikasi ini dapat membantu pendidik dalam memverifikasi keaslian tugas siswa, khususnya dalam menghadapi tantangan meningkatnya penggunaan alat kecerdasan buatan seperti ChatGPT dalam proses pembuatan tugas siswa. Dengan menerapkan *framework* Streamlit, aplikasi ini memberikan kemudahan penggunaan melalui dua metode input, baik secara manual (Tunggal) maupun melalui file.

Meski demikian, hasil penelitian juga mengungkapkan adanya keterbatasan dalam mendeteksi teks AI yang ditulis dengan gaya bahasa formal atau kompleks, karena pola kebahasaan yang digunakan oleh pakar sering kali menyerupai struktur teks yang dihasilkan oleh AI. Oleh karena itu, untuk meningkatkan akurasi di masa mendatang, dibutuhkan pengembangan lebih lanjut, terutama dalam penyempurnaan fitur linguistik dan peningkatan kompleksitas model yang digunakan.

DAFTAR PUSTAKA

- [1] M. M. Mariani, I. Machado, V. Magrelli, and Y. K. Dwivedi, "Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions," *Technovation*, vol. 122, Apr. 2023, doi: 10.1016/j.technovation.2022.102623.
- [2] I. Dergaa, K. Chamari, P. Zmijewski, and H. Ben Saad, "From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing," *Biol Sport*, vol. 40, no. 2, pp. 615–622, Mar. 2023, doi: 10.5114/BIOLOSPORT.2023.125623.
- [3] I. M. Krecar, M. Kolega, and L. Jurcec, "Perception of ChatGPT Usage for Homework Assignments: Students' and Professors' Perspectives," 2024.
- [4] K. Rosenblatt, "ChatGPT banned from New York City public schools' devices and networks," *NBC News*, Jan. 06, 2023.
- [5] K. Tenbarga, "Parents sue son's high school history teacher over AI 'cheating' punishment," *NBC News*.
- [6] Hingham Public Schools, "Hingham High School: Student Handbook Code of Discipline," Aug. 2024. Accessed: Apr. 12, 2025. [Online]. Available: https://core-docs.s3.us-east-1.amazonaws.com/documents/asset/uploaded_file/4900/HHS/4719901/Student_Handbook_Code_Discipline_2024_2025.pdf
- [7] M. S. Salim and S. I. Hossain, "An Applied Statistics dataset for human vs AI-generated answer classification," *Data Brief*, vol. 54, Jun. 2024, doi: 10.1016/j.dib.2024.110240.
- [8] J. Fleckenstein, J. Meyer, T. Jansen, S. D. Keller, O. Köller, and J. Möller, "Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100209.
- [9] A. Singh, D. Sharma, A. Nandy, and V. K. Singh, "Towards a large sized curated and annotated corpus for discriminating between human written and AI generated texts: A case study of text sourced from Wikipedia and ChatGPT," *Natural Language Processing Journal*, vol. 6, pp. 1–8, Mar. 2024, doi: 10.1016/j.nlp.2023.100050.

- [10] F. Akar, "Performance Analysis of NLP-Based Machine Learning Algorithms in Cyberbullying Detection," *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 17, no. 2, pp. 445–459, Aug. 2024, doi: 10.18185/erzifbed.1474112.
- [11] R. Alanazi and S. Alanazi, "A hybrid NLP and domain validation technique for disposable email detection," *Alexandria Engineering Journal*, vol. 102, pp. 200–210, Sep. 2024, doi: 10.1016/j.aej.2024.05.068.
- [12] M. Ayala-Chauvin and F. Avilés-Castillo, "Optimizing Natural Language Processing: A Comparative Analysis of GPT-3.5, GPT-4, and GPT-4o," *Data and Metadata*, vol. 3, Jan. 2024, doi: 10.56294/dm2024.359.
- [13] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A Survey of Text Representation and Embedding Techniques in NLP," *IEEE Access*, vol. 11, pp. 36120–36146, 2023, doi: 10.1109/ACCESS.2023.3266377.
- [14] H. Purwantoro and M. Y. Wibowo, *Natural Language Processing*, Solikin. Jakarta Selatan: Buletin Aptikom, 2020.
- [15] T. Wiranto and R. Santosa, "Bahasa, fungsi bahasa, dan konteks sosial," 2014.
- [16] Seo, "Top Generative AI Chatbots by Market Share – March 2025," FirstPageSage.
- [17] W. Rachbini, T. Evi, and Suyanto, *Pengenalan ChatGPT Tips dan Trik bagi Pemula*, 1st ed. Banten: CV. AA. Rizky, 2023.
- [18] F. Duarte, "Number of ChatGPT Users," <https://explodingtopics.com/blog/chatgpt-users>.
- [19] N. L. Amalia, "Analisis Sentimen pada Perpindahan Ibukota Indonesia dengan Algoritma Support Vector Machine Evaluasi Leksikon, Metode Ekstraksi Fitur, dan Kernel Trick," Universitas Islam Negeri Syarif Hidayatullah Jakarta, 2023.
- [20] V. Rupapara, F. Rustam, A. Amaar, P. B. Washington, E. Lee, and I. Ashraf, "Deepfake tweets classification using stacked Bi-LSTM and words embedding," *PeerJ Comput Sci*, vol. 7, 2021, doi: 10.7717/peerj-cs.745.
- [21] F. Rustam *et al.*, "Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model," *IEEE Access*, vol. 9, pp. 33675–33688, 2021, doi: 10.1109/ACCESS.2021.3061592.
- [22] DQLab, "Mengenal Streamlit, Tools Favorit Data Scientist," DQLab AI-Powered Learning.
- [23] R. F. Putra *et al.*, *Data Mining Algoritma dan Penerapannya*, 1st ed. Sonpedia Publishing Indonesia, 2023.
- [24] Y. Asri, D. Kuswardani, L. F. M. Horhoruw, and A. A. Sari, *Text Similarity*, 1st ed. Uwais Inspirasi Indonesia, 2024.
- [25] Y. Asri and D. Kuswardani, *Text Similarity*. Uwais Inspirasi indonesia, 2024.
- [26] S. Sofiana, *Konsep Bert pada Natural Language Processing*. CV. Eureka Media Aksara, 2024.
- [27] C. R. A. Widiawati and A. M. Wahid, *Pemrosesan Bahasa Alami Konsep, Algoritma & Implementasi*. Zahira Media Publisher, 2025.
- [28] W. M. Baihaqi, B. A. Kusuma, and L. H. Honimah, *Level dasar pembelajaran mesin dengan python dan iot dengan c*, 1st ed. Zahira Media Publisher, 2022. [Online]. Available: <https://books.google.co.id/books?id=4v2hEAAAQBAJ>