

Penerapan Chatbot Berbasis Natural Language Processing untuk Layanan Informasi PPDB di SMK PGRI 1 Kediri

^{1*}Mochammad Al Farizi, ²Intan Nur Farida, ³Daniel Swanjaya

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹alfarizi2852@gmail.com, ²in.nfarida@gmail.com, ³daniel@unpkediri.ac.id

Penulis Korespondens : Mochammad Al Farizi

Abstrak— Pelayanan informasi Penerimaan Peserta Didik Baru (PPDB) merupakan aspek penting dalam mendukung proses pendaftaran siswa baru di tingkat sekolah menengah. Di SMK PGRI 1 Kediri, pelayanan informasi PPDB sebelumnya masih mengandalkan tenaga manusia, yang memiliki keterbatasan dalam hal waktu dan kapasitas, terutama saat jumlah pertanyaan meningkat secara signifikan. Untuk mengatasi permasalahan tersebut, penelitian ini telah berhasil mengembangkan sebuah chatbot berbasis Natural Language Processing (NLP) yang diintegrasikan dengan metode Levenshtein Distance sebagai solusi otomatisasi layanan informasi. Chatbot yang dikembangkan mampu memahami pertanyaan pengguna dalam bahasa alami dan mencocokkannya dengan data yang tersedia secara cerdas, meskipun terdapat kesalahan ketik atau variasi kata. Berdasarkan hasil pengujian, sistem ini terbukti dapat meningkatkan efisiensi pelayanan, mempercepat respons, serta memberikan informasi yang akurat kepada calon peserta didik dan orang tua. Tingkat keberhasilan chatbot dalam mencocokkan input pengguna terhadap data relevan mencapai lebih dari 80%, sehingga menunjukkan bahwa sistem ini efektif dan layak diterapkan sebagai sarana layanan informasi PPDB yang modern dan responsif.

Kata kunci— *Chatbot, Levenshtein Distance, NLP*

Abstract— Information services for New Student Admissions (PPDB) are a crucial aspect in supporting the student recruitment process at the secondary education level. At SMK PGRI 1 Kediri, PPDB information services previously relied on human staff, which had limitations in terms of time and capacity, especially when the number of inquiries increased significantly. To address this issue, this study successfully developed a chatbot based on Natural Language Processing (NLP) integrated with the Levenshtein Distance method as a solution for automating information services. The developed chatbot is capable of understanding user questions in natural language and intelligently matching them with available data, even in the presence of typos or word variations. Based on the testing results, the system has proven to improve service efficiency, accelerate response times, and provide accurate information to prospective students and their parents. The chatbot achieved a success rate of over 80% in matching user inputs to relevant data, indicating that the system is effective and suitable for implementation as a modern and responsive PPDB information service.

Keywords— *Chatbot, Levenshtein Distance, NLP*

I. PENDAHULUAN

Di era transformasi digital yang kian pesat, kecepatan dan kemudahan akses informasi menjadi kebutuhan utama, termasuk dalam sektor pendidikan. Institusi pendidikan seperti SMK PGRI 1 Kediri kerap menghadapi tantangan dalam memberikan layanan informasi yang responsif kepada calon peserta didik dan orang tua, terutama selama proses Penerimaan Peserta Didik Baru (PPDB). Berbagai pertanyaan seputar syarat pendaftaran, jadwal seleksi, dan proses daftar ulang

sering kali masuk secara bersamaan, sehingga membebani panitia dan memengaruhi kualitas layanan yang diberikan.

Sebagai solusi, pemanfaatan teknologi chatbot berbasis Natural Language Processing (NLP) mulai diterapkan secara luas dalam dunia pendidikan. Chatbot dirancang untuk menjawab pertanyaan pengguna secara otomatis dan real-time, sehingga dapat membantu mengurangi beban kerja tenaga administrasi dan mempercepat proses pelayanan informasi [1][2][3][4]. Dalam konteks pendidikan, implementasi chatbot juga terbukti dapat meningkatkan keterlibatan siswa dan efisiensi proses pembelajaran [5].

Namun demikian, tantangan muncul ketika pengguna mengajukan pertanyaan dengan berbagai gaya bahasa atau mengalami kesalahan penulisan. Dalam kondisi ini, chatbot berbasis NLP saja belum cukup. Diperlukan tambahan algoritma yang mampu mengenali kata atau kalimat yang tidak persis sama dengan data latih. Salah satu pendekatan yang efektif adalah penggunaan algoritma Levenshtein Distance, yaitu metode pencocokan string yang toleran terhadap kesalahan ejaan atau variasi penulisan [6].

Levenshtein Distance telah banyak digunakan dalam berbagai aplikasi teknologi pendidikan di Indonesia, termasuk dalam pendeteksian kesamaan jawaban daring [7], layanan chatbot berbasis AI Markup Language (AIML) [8], hingga sistem tanya jawab dalam aplikasi pembelajaran [9]. Dengan mengombinasikan NLP dan Levenshtein Distance, sistem chatbot dapat merespons pertanyaan lebih fleksibel dan tetap relevan, meskipun terdapat perbedaan penulisan kata dari pengguna [10].

Sebagai kontribusi, penelitian ini mengembangkan chatbot PPDB untuk SMK PGRI 1 Kediri dengan pendekatan gabungan NLP dan Levenshtein Distance, yang bertujuan menciptakan layanan informasi yang responsif, cerdas, serta mampu beradaptasi terhadap variasi input bahasa dari calon peserta didik dan masyarakat umum.

II. METODE

2.1 Pengumpulan data

Langkah pertama dalam penelitian ini adalah melakukan wawancara terstruktur dengan panitia PPDB SMK PGRI 1 Kediri guna mengidentifikasi jenis pertanyaan yang paling sering diajukan oleh calon peserta didik dan wali murid. Wawancara difokuskan pada aspek-aspek teknis seperti persyaratan administrasi, jadwal dan prosedur seleksi, serta informasi seputar daftar ulang dan kontak layanan. Hasil wawancara kemudian dirangkum dalam bentuk dokumen tanya-jawab (FAQ), yang digunakan sebagai basis data dalam pelatihan chatbot.

Agar chatbot mampu memahami berbagai bentuk pertanyaan, digunakan pendekatan Natural Language Processing (NLP) modern. NLP berperan dalam mengenali struktur kalimat, sinonim, dan pola komunikasi manusia. Untuk meningkatkan toleransi terhadap kesalahan penulisan, diterapkan pula algoritma Levenshtein Distance. Penerapan gabungan teknologi ini telah terbukti sukses dalam beberapa penelitian lokal terkait chatbot pendidikan.

2.2 Implementasi algoritma *Levenshtein Distance*

Algoritma Levenshtein Distance digunakan untuk menghitung “jarak” antara dua string dengan menentukan jumlah operasi minimum yang diperlukan untuk mengubah satu kata menjadi kata lain—melalui proses penyisipan, penghapusan, atau penggantian karakter. Contoh

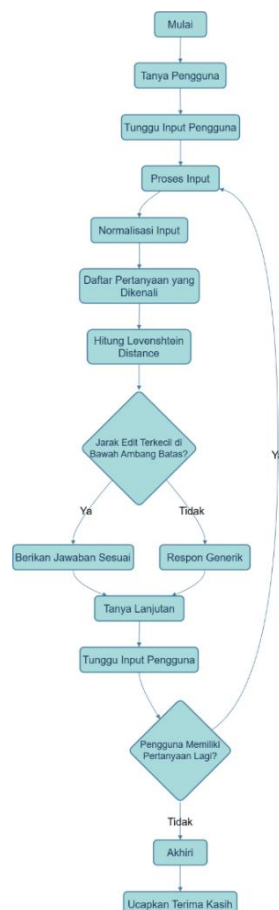
sederhananya adalah jika pengguna mengetik “pendaftran” sebagai ganti dari “pendaftaran”, maka algoritma ini akan tetap mengenali kesamaan kata tersebut dan menyajikan jawaban yang relevan.

Keunggulan utama algoritma ini adalah kemampuannya mengenali pola teks yang mirip meskipun tidak identik secara harfiah. Hal ini sangat penting dalam sistem chatbot yang digunakan oleh publik, karena pengguna kerap menulis dengan gaya bebas dan tidak selalu sesuai ejaan formal. Implementasi Levenshtein Distance dalam sistem chatbot pendidikan telah berhasil meningkatkan efisiensi dalam pencarian informasi dan klasifikasi pertanyaan pengguna, serta pencocokan teks pada aplikasi lain seperti sistem rekomendasi judul atau konten pembelajaran.

Dengan sinergi antara NLP dan algoritma Levenshtein Distance, sistem chatbot PPDB yang dikembangkan dalam penelitian ini diharapkan mampu memberikan pengalaman interaktif yang lebih baik, mengurangi beban operasional panitia, serta meningkatkan kualitas layanan informasi sekolah kepada publik.

2.3 Rancangan Sistem

Flowchart pada Gambar 1 menampilkan alur jalannya aplikasi chatbot yang akan diterapkan pada PPDB SMK PGRI 1 KEDIRI



Gambar 1. *Flowchart Chatbot PPDB SMK PGRI 1 KEDIRI*

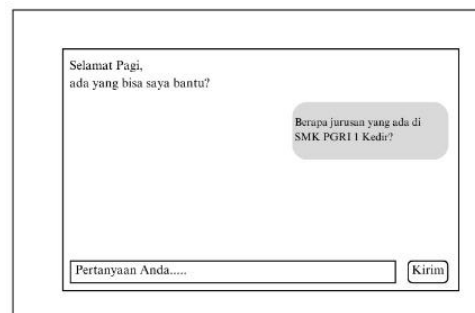
Dari flowchart sistem chatbot PPDB tersebut, alur proses dapat dijelaskan lebih detail sebagai berikut:

1. Pengguna atau calon peserta didik mengirimkan pertanyaan terkait informasi PPDB melalui antarmuka chatbot dalam bentuk teks.
2. Teks pertanyaan akan diproses menggunakan metode *Levenshtein Distance* untuk menghitung persentase kemiripan dengan daftar kata kunci yang telah tersimpan di database.
3. Hasil kemiripan kemudian disaring. Jika tingkat kemiripan mencapai 70% atau lebih, maka sistem akan memproses pertanyaan tersebut lebih lanjut.
4. Sistem akan mencari data yang relevan dalam basis data. Jika ditemukan, informasi tersebut akan ditampilkan sebagai jawaban chatbot.
5. Jika tidak ada kecocokan yang ditemukan atau sistem tidak memahami maksud pertanyaan, maka chatbot akan memberikan respons bahwa pertanyaan belum dapat dipahami dan menyarankan pengguna untuk mengulang dengan kalimat yang lebih jelas.

Dalam chatbot pada SMK PGRI 1 Kediri menggunakan desain database seperti pada Gambar 1. Tabel *faq* digunakan untuk menyimpan data pertanyaan beserta jawaban yang menjadi referensi utama untuk pencocokan. Kemudian, tabel *faq_category* digunakan untuk mengelompokkan pertanyaan berdasarkan kategori tertentu guna mempermudah klasifikasi dan pencarian jawaban. Selanjutnya, terdapat tabel *entities* (opsional) yang digunakan untuk menyimpan entitas-entitas khusus yang berkaitan dengan pertanyaan atau jawaban, yang berfungsi untuk mendukung proses pemahaman konteks dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) pada sistem *chatbot*.

2.4 Halaman *User Interface*

Pada halaman pengguna terdapat satu tampilan seperti pada Gambar 2.



Gambar 2. Tampilan Percakapan

Saat calon siswa masuk halaman website akan disambut dengan tampilan pada Gambar 2, dan calon siswa dapat mulai mengajukan pertanyaan kepada chatbot seperti yang ditunjukkan pada Gambar 2

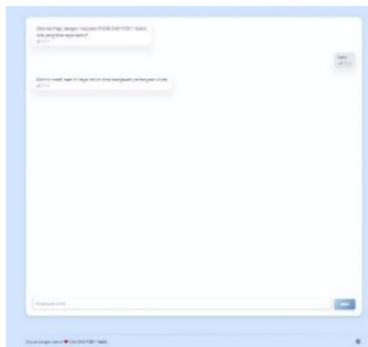
III. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil penerapan dari desain sistem Chatbot PPDB yang telah dibahas pada bab sebelumnya. Sistem informasi berupa Chatbot ini dikembangkan berdasarkan analisis kebutuhan informasi Penerimaan Peserta Didik Baru (PPDB) serta perancangan sistem komunikasi otomatis antara calon peserta didik atau orang tua dengan pihak sekolah. Chatbot dirancang untuk memberikan respon terhadap berbagai pertanyaan seputar jadwal pendaftaran, syarat administrasi, jurusan yang tersedia, hingga biaya pendidikan. Berikut ini adalah hasil implementasi dan penerapan sistem Chatbot PPDB berbasis web yang telah berhasil dikembangkan :

3.1 Tampilan Aplikasi

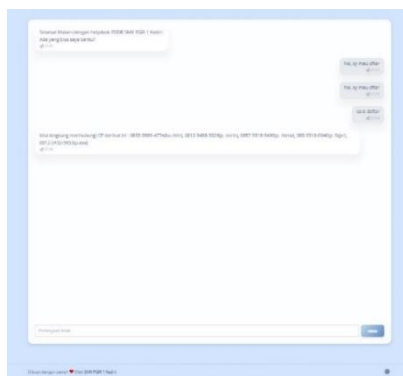
a. Halaman Chatbot

Pada halaman ini calon siswa dapat menanyakan hal—hal yang berkaitan dengan PPDB. Seperti ditunjukkan pada Gambar 4 pengguna mengirim pertanyaan maka akan dijawab oleh *Chatbot* sekalipun diluar jam kerja.

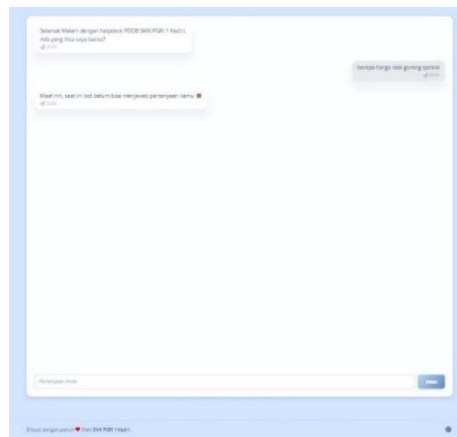


Gambar 3. Halaman Percakapan

Pada menu ini terdapat dua jenis jawaban, pertama apabila pertanyaan cocok dengan kata kunci yang ditunjukkan pada Gambar 4 dimana *bot* akan memberikan jawaban sesuai pertanyaan. Kedua jika pertanyaan tidak dapat dijawab oleh *bot* yang ditunjukkan pada Gambar 5 bot akan menjawab dengan kalimat “Maaf nih, saat ini bot belum bisa menjawab pertanyaan kamu”,



Gambar 4. *Chatbot* Menjawab Pertanyaan



Gambar 5. Chatbot tidak dapat menjawab pertanyaan

3.2 Pengujian Fungsi

Dari hasil implementasi aplikasi, dilakukan uji coba singkat dengan beberapa pertanyaan pendek dan panjang yang tertulis pada Tabel 1.

Tabel 1. Simulasi Chatbot

NO	Pertanyaan Sistem	Pengujian Ke	Pertanyaan Pengguna	Persentase Kecocokan	Status
1	Di mana letak SMK PGRI 1 Kediri secara lengkap?	1	alamat lengkap SMK PGRI 1 Kediri di mana?	100%	Sesuai
	Di mana letak SMK PGRI 1 Kediri secara lengkap?	2	tolong beri tahu lokasi lengkap SMK PGRI 1 Kediri	97%	Sesuai
2	Berapa banyak jurusan yang tersedia di SMK PGRI 1 Kediri?	1	jumlah jurusan di SMK PGRI 1 Kediri berapa?	100%	Sesuai
	Berapa banyak jurusan yang tersedia di SMK PGRI 1 Kediri?	2	jurusan yang tersedia di SMK PGRI 1 Kediri ada berapa?	95%	Hampir Sesuai
3	Kalau mau daftar ke SMK PGRI 1 Kediri, syaratnya apa saja?	1	apa syarat daftar ke SMK PGRI 1 Kediri?	100%	Sesuai
	Kalau mau daftar ke SMK PGRI 1 Kediri, syaratnya apa saja?	2	syarat pendaftaran ke SMK PGRI 1 Kediri apa saja ya?	96%	Sesuai

4	Kapan mulai dibuka penerimaan peserta didik baru di SMK PGRI 1 Kediri?	1	tanggal pembukaan PPDB SMK PGRI 1 Kediri kapan?	98%	Hampir Sesuai
	Kapan mulai dibuka penerimaan peserta didik baru di SMK PGRI 1 Kediri?	2	kapan mulai pendaftaran siswa baru di SMK PGRI 1 Kediri?	95%	Sesuai
5	Bagaimana cara daftar ke SMK PGRI 1 Kediri?	1	gimana cara daftar ke SMK PGRI 1 Kediri?	100%	Sesuai
	Bagaimana cara daftar ke SMK PGRI 1 Kediri?	2	saya ingin tahu proses pendaftaran ke SMK PGRI 1 Kediri	97%	Sesuai

3.3 Pembahasan

Berdasarkan pengujian pada Tabel 1, dilakukan dua skenario pengujian dengan menggunakan lima pertanyaan seputar informasi PPDB di SMK PGRI 1 Kediri. Pada pengujian pertama, digunakan kalimat pertanyaan yang pendek dan langsung mengarah pada kata kunci, sedangkan pada pengujian kedua digunakan kalimat yang lebih panjang dan bersifat naratif namun masih relevan dengan inti pertanyaan.

Hasil dari kedua pengujian menunjukkan bahwa tingkat kecocokan jawaban chatbot relatif stabil di antara kedua jenis pertanyaan. Hal ini menunjukkan bahwa panjang atau pendeknya kalimat tidak terlalu mempengaruhi hasil kecocokan, selama struktur kalimat tetap mengandung kata kunci yang sesuai dengan database pertanyaan chatbot.

Perhitungan Persentase Kecocokan menggunakan Levenshtein Distance:

Persentase kecocokan dihitung dengan mengukur jarak kemiripan antara pertanyaan pengguna dan pertanyaan standar yang ada di database chatbot menggunakan *Levenshtein Distance*. Levenshtein Distance adalah metode yang menghitung jumlah operasi penyuntingan minimum yang diperlukan untuk mengubah satu string menjadi string lain, di mana operasi penyuntingan tersebut berupa:

1. Penambahan karakter,
2. Penghapusan karakter,
3. Penggantian karakter.

Berikut adalah uraian lengkap perhitungan persentase kecocokan dengan metode Levenshtein Distance untuk Soal Pengujian Nomor 2 :

Pertanyaan Sistem :

Berapa banyak jurusan yang tersedia di SMK PGRI 1 Kediri?

(56 karakter termasuk spasi)

Pertanyaan Pengguna :

jurusan yang tersedia di SMK PGRI 1 Kediri ada berapa?

(56 karakter termasuk spasi)

1. Preprocessing (Normalisasi)
 - a. Semua huruf dijadikan huruf kecil

- b. Tidak menghapus tanda baca atau kata hubung, karena kita hitung berbasis karakter

Normalized Strings:

Sistem → berapa banyak jurusan yang tersedia di smk pgri 1 kediri?

Pengguna → jurusan yang tersedia di smk pgri 1 kediri ada berapa?

2. Levenshtein Distance (karakter-level)

a. Perbedaan nyata:

1. Frasa “berapa banyak” diganti menjadi “ada”
2. Urutan frasa “ada berapa” dipindah ke akhir

b. Operasi yang dibutuhkan:

1. Hapus: “berapa banyak” (12 karakter)
2. Tambah: “ada” (3 karakter)
3. Tambah/move: “berapa” (urutan akhir, meski sudah ada)

Jumlah perubahan minimum: 3 operasi

3. Hitung Panjang String

- a. Panjang maksimum string = 56 karakter

4. Rumus Persentase Kecocokan

$$\begin{aligned} \text{\textit{Persentase Kecocokan}} \\ &= \left(1 - \frac{\text{\textit{Levenshtein Distance}}}{\text{\textit{Panjang Maksimum}}}\right) \times 100 \end{aligned}$$

$$\begin{aligned} &= \left(1 - \frac{3}{56}\right) \times 100 \\ &= (1 - 0.0536) \times 100 \\ &\approx 94.64\% \end{aligned}$$

Dibulatkan ke atas menjadi: 95%

Berdasarkan hasil pengujian pada Tabel 1 serta perhitungan menggunakan metode Levenshtein Distance, diketahui bahwa persentase kecocokan antara pertanyaan pengguna dan pertanyaan sistem berada pada rentang 84,64% hingga 100%, tergantung pada jumlah operasi edit (penambahan, penghapusan, atau penggantian karakter) yang diperlukan untuk menyamakan dua kalimat tersebut.

Contoh perhitungan pada pengujian nomor 2 menunjukkan bahwa pertanyaan sistem “Berapa banyak jurusan yang tersedia di SMK PGRI 1 Kediri?” dan pertanyaan pengguna “jurusan yang tersedia di SMK PGRI 1 Kediri ada berapa?” memiliki Levenshtein Distance sebesar 3 terhadap panjang maksimal kalimat 56 karakter. Dengan demikian, persentase kecocokan dihitung menggunakan rumus:

Persentase Kecocokan

$$\begin{aligned} \text{\textit{Persentase Kecocokan}} \\ &= \left(1 - \frac{3}{56}\right) \times 100 \approx 94,64\% \end{aligned}$$

Hasil ini berada jauh di atas ambang batas keberhasilan minimal sebesar 80% yang ditetapkan sebagai standar evaluasi. Dengan kata lain, chatbot berhasil memahami dan mencocokkan pertanyaan dengan struktur dan redaksi berbeda namun makna yang sama secara akurat.

Secara umum, chatbot menunjukkan kinerja yang sangat baik, dengan nilai kecocokan rata-rata melebihi 90%, baik pada pertanyaan yang bersifat pendek maupun yang panjang atau naratif. Hal ini menunjukkan bahwa sistem mampu menangani variasi bahasa alami pengguna dengan baik.

IV. KESIMPULAN

Berdasarkan hasil pengujian terhadap chatbot yang dikembangkan untuk mendukung layanan informasi PPDB di SMK PGRI 1 Kediri, diketahui bahwa sistem mampu memberikan jawaban yang relevan terhadap lima pertanyaan uji yang diajukan oleh pengguna. Seluruh pertanyaan berhasil dijawab dengan benar dan sesuai, menunjukkan tingkat keberhasilan maksimal dalam pengujian.

Dalam hal kecocokan isi jawaban, chatbot menunjukkan tingkat kesesuaian kata kunci dan makna yang sangat tinggi, dengan nilai kecocokan terendah tetap berada di atas batas minimum keberhasilan yang telah ditentukan. Ini menandakan bahwa bahkan pada kasus dengan variasi redaksi kalimat atau susunan kata yang berbeda, chatbot masih mampu memahami konteks dan memberikan respons yang sesuai dengan maksud pengguna.

Rata-rata hasil pengujian menunjukkan performa chatbot yang konsisten, baik dalam menjawab pertanyaan pendek maupun yang berbentuk naratif. Dengan akurasi jawaban yang stabil dan kemampuan memahami variasi bahasa alami pengguna, sistem ini telah memenuhi standar evaluasi yang ditetapkan dalam pengujian.

Dengan demikian, dapat disimpulkan bahwa chatbot ini layak digunakan sebagai media informasi untuk pelaksanaan PPDB di SMK PGRI 1 Kediri. Implementasinya dapat memberikan kemudahan akses informasi bagi calon peserta didik maupun orang tua secara cepat, tepat, dan efisien melalui pendekatan teknologi berbasis percakapan.

DAFTAR PUSTAKA

- [1] V. Dalam, M. Motivasi, and B. Siswa, "Jurnal Inovasi Pendidikan STUDI LITERATUR : PENGARUH CHATBOT AI SEBAGAI TUTOR," vol. 14, no. 2, pp. 42–49, 2024.
- [2] W. Suwarningsih and Nuryani, "Generate fuzzy string-matching to build self attention on Indonesian medical-chatbot," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 819–829, 2024, doi: 10.11591/ijece.v14i1.pp819-829.
- [3] D. Prayogo and O. P. Martadireja, "Systematic Literature Review Chatbot : Suatu Perbandingan Pendekatan," vol. 8, no. April, pp. 4172–4178, 2025.
- [4] F. H. Aldwinarta, R. Nurdiana, and O. Sulistina, "Media Pembelajaran Berbasis AI Chatbot pada Materi Termokimia di SMA Apakah Dibutuhkan?," *J. Inov. Pendidik. Kim.*, vol. 18, no. 1, pp. 1–6, 2024, doi: 10.15294/jipk.v18i1.49044.
- [5] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, "Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance," *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: 10.14710/j.gauss.v11i2.35472.
- [6] G. Syahputra, A. F. Boy, and N. B. Nugroho, "Penerapan Metode Levenshtein Distance Dalam Mendeteksi Hasil Jawaban Peserta Kelas Daring Dengan Teknik Web Scraping Manual," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 20, no. 1, p. 14, 2021, doi: 10.53513/jis.v20i1.1900.
- [7] G. Guntoro, Loneli Costaner, and L. Lisnawita, "Aplikasi Chatbot untuk Layanan Informasi dan Akademik Kampus Berbasis Artificial Intelligence Markup Language (AIML)," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, pp. 291–300, 2020, doi: 10.31849/digitalzone.v11i2.5049.
- [8] A. Setiawan, O. N. Pratiwi, and R. Y. Fa'rifah, "Question Answering System Dalam Bentuk Chatbot Pada Platform Line Untuk Mata Pelajaran Sejarah SMA/MA Dengan Menggunakan Algoritma Levenshtein Distance," *eProceedings Eng.*, vol. 8, no. 5, pp. 9794–9802, 2021.

- [9] H. Sudarman, F. I. Komputer, U. E. Unggul, S. Informasi, F. I. Komputer, and U. E. Unggul, “Analisis dan Deteksi Kemiripan Teks Berbasis Python dengan Algoritma Levenshtein Distance Menurut informasi dari Pangkalan Data Pendidikan Tinggi Kementerian,” vol. 10, pp. 257–273, 2025.
- [10] A. L. Nasional and H. O. Nahampun, “Penerapan Algoritma Levenstein Distance Untuk Pencarian Judul Pada,” vol. 3, no. 4, pp. 282–286, 2023.