

Published online on the page : <https://journal.nagastra.org/>**Journal of Pattern Recognition and Intelligent Data**

| ISSN (Print) xxxx-xxxx | ISSN (Online) xxxx-xxxx |



Chatbot AI Menggunakan Metode TF-IDF Dan Cosine Similarity Untuk Layanan Pelanggan Toko Bangunan

Irhaz Salim^{1,*}, Nofal Fauzan², Mutiara Rahmadina³, Aulia Maharani Desyanta⁴, Raisyafa Ihsan⁵

¹Program Studi Pendidikan Teknik Informasi dan Komunikasi, UIN Sjech M. Djamil Djambek Bukittinggi, Indonesia

Abstrak

Perkembangan teknologi kecerdasan buatan mendorong berbagai sektor usaha untuk memanfaatkan chatbot sebagai solusi layanan pelanggan otomatis. Toko bangunan sebagai salah satu sektor ritel menghadapi tantangan dalam memberikan layanan informasi produk yang cepat dan akurat kepada pelanggan. Penelitian ini bertujuan merancang dan mengimplementasikan chatbot berbasis kecerdasan buatan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity untuk layanan pelanggan toko bangunan. Sistem dibangun menggunakan framework Laravel dengan integrasi database produk secara real-time. Metode TF-IDF digunakan untuk pembobotan kata dari pesan masukan pengguna terhadap korpus pengetahuan yang telah disiapkan, sedangkan Cosine Similarity digunakan untuk mengukur kemiripan antara vektor pertanyaan pengguna dan vektor jawaban yang tersedia. Pengujian dilakukan dengan 50 skenario percakapan yang mencakup pertanyaan tentang harga produk, stok barang, jam operasional, metode pembayaran, dan informasi pengiriman. Hasil pengujian menunjukkan akurasi chatbot sebesar 88% dalam mengenali dan menjawab pertanyaan pelanggan dengan tepat. Penelitian ini membuktikan bahwa kombinasi TF-IDF dan Cosine Similarity efektif digunakan dalam sistem chatbot untuk domain layanan pelanggan toko bangunan.

Kata kunci : chatbot, TF-IDF, Cosine Similarity, layanan pelanggan, toko bangunan

Abstract

The advancement of artificial intelligence technology encourages various business sectors to utilize chatbots as automated customer service solutions. Building material stores, as a retail sector, face challenges in providing fast and accurate product information to customers. This study aims to design and implement an artificial intelligence-based chatbot using the Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity methods for customer service in building material stores. The system is built using the Laravel framework with real-time product database integration. The TF-IDF method is used for word weighting from user input messages against a prepared knowledge corpus, while Cosine Similarity is used to measure the similarity between the user query vector and available answer vectors. Testing was conducted with 50 conversation scenarios covering questions about product prices, stock availability, operating hours, payment methods, and delivery information. Test results show a chatbot accuracy of 88% in correctly recognizing and answering customer queries. This study proves that the combination of TF-IDF and Cosine Similarity is effective for chatbot systems in the customer service domain of building material stores.

Keywords: chatbot, TF-IDF, Cosine Similarity, customer service, building material store

1. Introduction

Perkembangan teknologi informasi yang pesat mendorong transformasi digital di berbagai sektor usaha, termasuk toko bangunan. Pelanggan kini mengharapkan layanan yang cepat, responsif, dan tersedia kapan saja tanpa harus menunggu tenaga penjual secara langsung [1]. Kebutuhan ini mendorong pemanfaatan kecerdasan buatan, khususnya sistem chatbot, sebagai solusi layanan pelanggan otomatis yang dapat beroperasi dua puluh empat jam sehari [2].

Chatbot merupakan sistem perangkat lunak yang mampu mensimulasikan percakapan manusia menggunakan teknik pemrosesan bahasa alami (Natural Language Processing/NLP) [3]. Dalam konteks layanan pelanggan toko bangunan, chatbot dapat menjawab pertanyaan tentang harga produk, ketersediaan stok, metode pembayaran, jam operasional, dan informasi pengiriman secara otomatis dan

real-time [4]. Hal ini tidak hanya meningkatkan kepuasan pelanggan, tetapi juga mengurangi beban kerja staf toko.

Penelitian sebelumnya menunjukkan bahwa metode berbasis pembobotan kata seperti TF-IDF (Term Frequency-Inverse Document Frequency) terbukti efektif dalam sistem tanya-jawab otomatis [5]. Kombinasi TF-IDF dengan Cosine Similarity memungkinkan sistem menghitung kemiripan antara pertanyaan pengguna dengan basis pengetahuan yang tersedia sehingga menghasilkan respons yang relevan dan akurat [6]. Namun, penerapannya pada domain spesifik toko bangunan dengan integrasi database produk masih belum banyak dikaji.

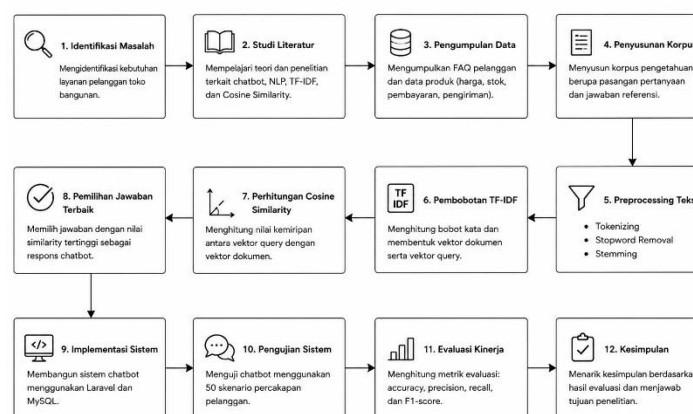
Penelitian ini mengidentifikasi gap bahwa belum terdapat sistem chatbot yang secara khusus dirancang untuk layanan pelanggan toko bangunan dengan kemampuan mengakses data produk secara real-time menggunakan metode TF-IDF dan Cosine Similarity [7] [8]. Novelty dari penelitian ini terletak pada integrasi metode TF-IDF dan Cosine Similarity dengan database produk dinamis dalam framework Laravel untuk menghasilkan chatbot yang mampu memberikan informasi harga dan stok produk toko bangunan secara akurat dan kontekstual [9].

Penelitian ini memberikan kontribusi dalam pengembangan layanan pelanggan berbasis kecerdasan buatan pada sektor ritel toko bangunan melalui pemanfaatan metode TF-IDF dan Cosine Similarity yang terintegrasi dengan basis data produk secara real-time. Pendekatan yang digunakan memiliki kompleksitas komputasi yang relatif rendah sehingga dapat diimplementasikan pada lingkungan usaha skala kecil dan menengah tanpa memerlukan sumber daya perangkat keras yang tinggi. Implementasi sistem diharapkan mampu meningkatkan kecepatan akses informasi serta kualitas pelayanan kepada pelanggan secara lebih efektif dan efisien

Penelitian ini bertujuan untuk merancang, mengimplementasikan, dan mengevaluasi sistem chatbot berbasis TF-IDF dan Cosine Similarity yang terintegrasi dengan database produk toko bangunan, serta mengukur tingkat akurasi sistem dalam menjawab pertanyaan pelanggan secara otomatis.

2. Method

Penelitian ini menggunakan metode Research and Development (R&D) dengan model pengembangan Software Development Life Cycle (SDLC) menggunakan pendekatan Waterfall. Tahapan penelitian meliputi identifikasi masalah, studi literatur, pengumpulan data, penyusunan korpus pengetahuan, preprocessing teks, pembobotan TF-IDF, perhitungan Cosine Similarity, pemilihan jawaban terbaik, implementasi sistem, pengujian sistem, evaluasi kinerja, dan kesimpulan.



Gambar 1. Tampilan Antar Muka Chatbot

Berdasarkan Gambar 1, penelitian diawali dengan identifikasi masalah dan studi literatur terkait chatbot, TF-IDF, dan Cosine Similarity. Selanjutnya dilakukan pengumpulan data serta penyusunan korpus pengetahuan yang berisi pasangan pertanyaan dan jawaban. Data kemudian diproses melalui tahap preprocessing, pembobotan TF-IDF, dan perhitungan Cosine Similarity untuk menentukan jawaban yang paling relevan. Sistem diimplementasikan menggunakan Laravel dan MySQL, kemudian dilakukan pengujian dan evaluasi untuk mengukur kinerja chatbot yang dikembangkan.

2.1. Pengumpulan Data dan Persiapan Korpus

Data dikumpulkan melalui observasi langsung pada toko bangunan yang menjadi mitra penelitian. Korpus pengetahuan chatbot disusun berdasarkan 200 pertanyaan umum pelanggan yang dikelompokkan ke dalam lima kategori, yaitu informasi produk, harga, stok, jam operasional, dan metode pembayaran. Setiap pertanyaan dipasangkan dengan jawaban referensi yang kemudian digunakan sebagai basis pengetahuan sistem.

2.2. Metode TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen. Metode ini memberikan bobot yang lebih tinggi pada kata yang sering muncul dalam suatu dokumen tetapi jarang muncul pada dokumen lainnya, sehingga dapat merepresentasikan informasi yang lebih relevan [10]. Nilai TF-IDF dihitung menggunakan persamaan berikut.

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

di mana $(TF(t,d))$ merupakan frekuensi kemunculan term (t) pada dokumen (d) , (N) adalah jumlah total dokumen dalam korpus, dan $(df(t))$ adalah jumlah dokumen yang mengandung term (t) .

Sebelum proses pembobotan dilakukan, data teks terlebih dahulu melalui tahap preprocessing yang meliputi tokenisasi, penghapusan stopword bahasa Indonesia, dan stemming menggunakan algoritma Sastrawi. Hasil pembobotan TF-IDF kemudian digunakan untuk membentuk representasi vektor dari setiap dokumen dan pertanyaan pengguna yang selanjutnya diproses menggunakan metode Cosine Similarity untuk menentukan tingkat kemiripan antar teks.

2.3. Metode Cosine Similarity

Cosine Similarity digunakan untuk mengukur kemiripan antara vektor TF-IDF dari pertanyaan pengguna dengan vektor TF-IDF dari setiap dokumen dalam basis pengetahuan [11]. Nilai Cosine Similarity dihitung menggunakan persamaan.

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} \quad (3)$$

$$\text{Cosine Similarity}(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

Nilai Cosine Similarity berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan tingkat kemiripan yang tinggi. Dokumen dengan nilai Cosine Similarity tertinggi dipilih sebagai respons chatbot

[12]. Ambang batas (threshold) sebesar 0,3 ditetapkan sebagai nilai minimum kemiripan untuk memberikan respons relevan; jika tidak ada dokumen yang memenuhi threshold, sistem memberikan respons default

2.4. Arsitektur Sistem

Sistem chatbot dibangun menggunakan framework Laravel 10 dengan database MySQL. Arsitektur sistem terdiri dari tiga lapisan utama, yaitu lapisan antarmuka pengguna (frontend), lapisan logika bisnis (backend), dan lapisan data. Frontend diimplementasikan sebagai floating chatbot pada halaman landing page menggunakan HTML, CSS, dan JavaScript dengan komunikasi asinkron melalui AJAX. Backend mengimplementasikan logika TF-IDF dan Cosine Similarity pada ChatbotController. Lapisan data terdiri dari tabel products (id, name, price, stock, category_id), tabel categories, dan tabel chat_logs untuk pencatatan riwayat percakapan.

2.5. Pengujian Sistem

Pengujian sistem dilakukan menggunakan 50 skenario pertanyaan yang mencakup lima kategori layanan, yaitu harga produk, stok barang, jam operasional, metode pembayaran, dan informasi pengiriman. Setiap kategori terdiri atas 10 pertanyaan yang digunakan untuk mengukur kemampuan chatbot dalam memberikan respons yang sesuai dengan pertanyaan pengguna.

Kinerja sistem dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score [13] [11]. Hasil respons chatbot kemudian dibandingkan dengan jawaban yang diharapkan untuk menentukan tingkat ketepatan sistem dalam menjawab pertanyaan pelanggan.

3. Results and Discussion

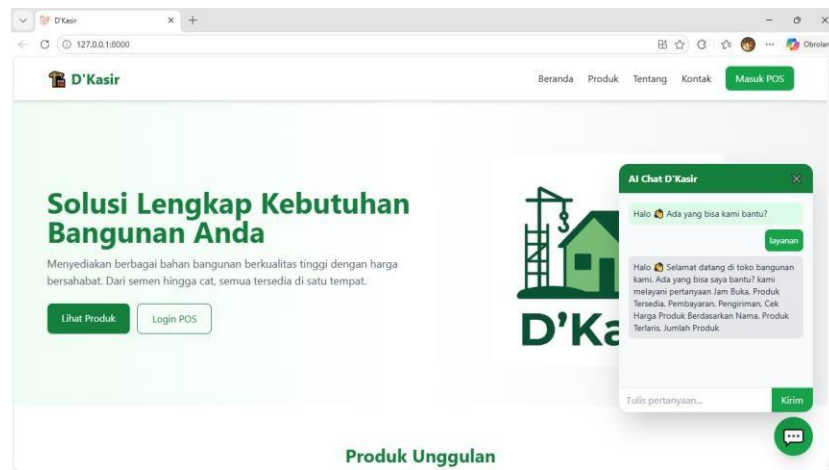
Sistem chatbot berhasil dikembangkan dan diimplementasikan pada platform web berbasis Laravel dengan antarmuka floating chatbot yang responsif. Chatbot dapat diakses langsung dari halaman landing page toko bangunan dan mampu merespons pertanyaan pelanggan secara real-time.

3.1. Hasil Implementasi Antarmuka Chatbot

Implementasi chatbot berhasil dilakukan pada platform web menggunakan framework Laravel. Antarmuka chatbot dirancang dalam bentuk floating bubble yang ditempatkan pada bagian kanan bawah halaman sehingga mudah diakses oleh pengguna tanpa mengganggu tampilan utama website. Ketika ikon chatbot dipilih, sistem menampilkan jendela percakapan yang memungkinkan pengguna berinteraksi secara langsung melalui masukan teks.

Antarmuka percakapan menampilkan pesan pengguna dan respons chatbot secara terpisah untuk meningkatkan keterbacaan selama proses interaksi. Sistem juga dilengkapi dengan typing indicator yang ditampilkan ketika chatbot sedang memproses pertanyaan pengguna. Fitur tersebut memberikan umpan balik visual sehingga pengguna dapat mengetahui bahwa sistem sedang menghasilkan respons.

Integrasi chatbot dengan basis data produk memungkinkan sistem menampilkan informasi terkait harga, ketersediaan stok, metode pembayaran, dan layanan pengiriman secara real-time. Hasil implementasi antarmuka chatbot ditunjukkan pada Gambar 2.



Gambar 2. Tampilan Antar Muka Chatbot

Berdasarkan Gambar 2, antarmuka chatbot dirancang untuk memberikan kemudahan interaksi antara pengguna dan sistem. Pengguna dapat mengajukan pertanyaan terkait informasi produk, harga, ketersediaan stok, metode pembayaran, maupun layanan pengiriman melalui jendela percakapan yang tersedia. Sistem kemudian memproses masukan pengguna menggunakan metode TF-IDF dan Cosine Similarity untuk menentukan respons yang paling relevan berdasarkan basis pengetahuan yang telah disusun. Hasil implementasi menunjukkan bahwa chatbot dapat memberikan respons secara cepat dan terstruktur sehingga mendukung peningkatan kualitas layanan pelanggan pada toko bangunan.

3.2. Hasil Pengujian Skenario Percakapan

Pengujian skenario percakapan dilakukan untuk mengevaluasi kemampuan chatbot dalam memberikan respons yang relevan terhadap berbagai pertanyaan pelanggan. Hasil pengujian pada masing-masing kategori pertanyaan disajikan pada Tabel 1.

Tabel 1. Hasil Pengujian Skenario Percakapan

Kategori	Contoh Pertanyaan	Respons Chatbot	Cosine Sim.
Harga Produk	Berapa harga semen?	Semen Portland 50kg – Rp 65.000, stok: 200 sak	0,87
Stok Barang	Cat putih masih ada?	Cat Dulux Putih 5kg tersedia, stok: 8 pcs	0,82
Jam Operasional	Toko buka jam berapa?	Kami buka 08:00–17:00 WIB setiap hari	0,91
Pembayaran	Bisa bayar QRIS?	Kami menerima Cash, Transfer Bank, dan QRIS	0,85

Pengiriman	Bisa diantar ke rumah?	Pengiriman tersedia sekitar toko	0,79 area
------------	------------------------	--	--------------

Dari tabel di atas, terlihat bahwa chatbot mampu memberikan respons yang relevan untuk berbagai kategori pertanyaan dengan nilai Cosine Similarity yang tinggi. Pertanyaan tentang jam operasional memperoleh nilai kemiripan tertinggi (0,91) karena frasa tersebut sangat spesifik dan konsisten dalam korpus pengetahuan.

3.3. Hasil Evaluasi Akurasi

Setelah dilakukan pengujian skenario percakapan, tahap berikutnya adalah mengevaluasi kinerja chatbot dalam memberikan jawaban yang sesuai dengan pertanyaan pengguna. Evaluasi dilakukan berdasarkan jumlah prediksi yang benar dan salah pada setiap kategori pertanyaan, kemudian dihitung nilai akurasi dan F1-Score untuk mengetahui tingkat performa sistem. Hasil evaluasi akurasi chatbot disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Akurasi per Kategori

Kategori	Benar	Salah	Akurasi	F1-Score
Harga Produk	9	1	90%	0,89
Stok Barang	8	2	80%	0,81
Jam Operasional	10	0	100%	1,00
Pembayaran	9	1	90%	0,90
Pengiriman	8	2	80%	0,80
Total/Rata-rata	44	6	88%	0,88

Hasil evaluasi menunjukkan akurasi keseluruhan sebesar 88% dengan F1-Score rata-rata 0,88. Kategori jam operasional mencapai akurasi sempurna (100%) karena pertanyaan dalam kategori ini memiliki pola yang sangat seragam. Sebaliknya, kategori stok barang dan pengiriman memperoleh akurasi lebih rendah (80%) karena variasi frasa yang lebih beragam dari pengguna.

3.4. Pembahasan

Penerapan metode TF-IDF dan Cosine Similarity pada chatbot layanan pelanggan toko bangunan terbukti efektif dengan akurasi 88%. Hasil ini konsisten dengan penelitian sebelumnya yang menunjukkan efektivitas metode berbasis vektor ruang untuk sistem tanya-jawab domain-spesifik. Keunggulan utama pendekatan ini adalah kemampuannya menangani variasi frasa pertanyaan tanpa memerlukan pencocokan kata yang eksak.

Integrasi dengan database produk Laravel memungkinkan chatbot memberikan informasi harga dan stok secara dinamis dan real-time, yang merupakan keunggulan signifikan dibandingkan chatbot berbasis aturan statis. Sistem juga mencatat seluruh riwayat percakapan yang dapat digunakan untuk analisis kebutuhan pelanggan lebih lanjut.

Kelemahan sistem saat ini terletak pada kemampuan memahami pertanyaan dengan ejaan tidak baku atau penggunaan bahasa gaul yang umum dalam percakapan informal. Peningkatan korpus pengetahuan

dan penambahan modul koreksi ejaan dapat meningkatkan performa sistem secara signifikan pada skenario penggunaan nyata.

4. Conclusion

Penelitian ini berhasil merancang dan mengimplementasikan sistem chatbot berbasis TF-IDF dan Cosine Similarity untuk layanan pelanggan toko bangunan menggunakan framework Laravel. Sistem mampu menjawab pertanyaan pelanggan secara otomatis terkait harga produk, ketersediaan stok, jam operasional, metode pembayaran, dan informasi pengiriman. Pengujian terhadap 50 skenario percakapan menghasilkan akurasi keseluruhan sebesar 88% dengan F1-Score rata-rata 0,88, membuktikan bahwa kombinasi metode TF-IDF dan Cosine Similarity efektif untuk sistem chatbot pada domain layanan pelanggan toko bangunan. Penelitian selanjutnya disarankan untuk mengembangkan kemampuan chatbot menggunakan pendekatan deep learning dan memperluas basis pengetahuan dengan data percakapan nyata dari toko bangunan untuk meningkatkan akurasi dan generalisasi sistem.

References

- [1] M. T. P. Adam, M. Wessel, and A. Benlian, "AI-based chatbots in customer service and their effects on user compliance," *Electron. Mark.*, vol. 31, no. 2, pp. 427–445, 2021, doi: 10.1007/s12525-020-00414-7.
- [2] V. R. Prasetyo, N. Benarkah, and V. J. Chrisintha, "Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya," *Teknika*, vol. 10, no. 2, pp. 114–121, 2021, doi: 10.34148/teknika.v10i2.370.
- [3] R. Martins and A. Silva, "Transforming Customer Service: The Role of AI-Powered Chatbots in Modern Retail and Business Sectors," *J. Inf. Syst. Technol. Manag.*, vol. 21, pp. 1–18, 2024, doi: 10.4301/S1807-1775202421002.
- [4] A. Setiyadi and W. Nugroho, "Implementasi Natural Language Processing (NLP) pada Chatbot untuk Meningkatkan Efisiensi Layanan Pelanggan," *J. Tekno Kompak*, vol. 17, no. 1, pp. 55–66, 2023, doi: 10.33365/jtk.v17i1.2412.
- [5] G. H. Setiawan and I. M. A. Adnyana, "Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 252–257, 2023, doi: 10.30871/jaic.v7i2.5611.
- [6] R. Shrivastava, A. Pujahari, S. P. Singh, and T. Bhowmik, "Efficient Question Answering in Chatbot Using TF-IDF and Cosine Similarity," *Innov. Inf. Commun. Technol.*, pp. 25–31, 2023, doi: 10.1007/978-981-19-3796-5_3.
- [7] F. R. Chan, R. Yanti, A. Hidayat, A. Ramadhanu, and F. Annas, "Deep Learning-Based Forecasting of Significant Wave Height in Padang Coastal Waters: A Comparative Study of RNN, LSTM, and GRU Architectures," *J. Comput. Res. Innov.*, vol. 11, no. 1, 2026, doi: 10.24191/jcrinn.v11i1.587.
- [8] Z. Mustafa, A. Rosyid, and S. Handayani, "Implementation of a WhatsApp-based Canteen Menu Chatbot using TF-IDF and Cosine Similarity: A Case Study with Laravel--MySQL Application Server," *J. Pendidik. Teknol. Kejuru.*, vol. 10, no. 1, pp. 45–56, 2026, [Online]. Available: <https://jptk.pj.unp.ac.id/index.php/jptk/article/view/485>
- [9] D. P. Putra and F. Ramadhan, "Pengembangan Sistem Informasi Inventory Toko Bangunan Real-Time Berbasis Web Menggunakan Framework Laravel," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 3, pp. 312–321, 2023, doi: 10.35870/jtik.v7i3.914.
- [10] V. Oktaviani, B. Warsito, H. Yasin, R. Santoso, and Suparti, "Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method," *J. Phys. Conf. Ser.*, vol. 1943, no. 1, 2021, doi: 10.1088/1742-6596/1943/1/012147.
- [11] F. R. Chan and A. Ramadhanu, "IMPLEMENTASI HYBRID INTELLIGENCE SYSTEM UNTUK KLASIFIKASI BIJI-BIJIAN DENGAN ALGORITMA PCA DAN KNN," *INTI Nusa Mandiri*, vol. 19, no. 2, pp. 240–250, 2025, doi: 10.33480/inti.v19i2.6397.
- [12] R. Gunawan and B. Setiawan, "Sistem Tanya Jawab Otomatis Berbasis Chatbot Menggunakan Metode TF-IDF dan Cosine Similarity," *J. Teknoinfo*, vol. 14, no. 2, pp. 89–96, 2020, doi: 10.33365/jti.v14i2.612.

- [13] F. R. Chan, F. Annas, Y. E. Yuspita, and G. Darmawati, "Implementation of Convolutional Neural Networks (CNN) in An Emotion Detection System for Measuring Learning Concentration Levels," *Knowbase Int. J. Knowl. Database*, vol. 04, no. 01, pp. 49–61, 2024, doi: 10.0.121.7/knowbase.v4i1.8429.