

Pengembangan chatbot informasi PPDB KB & TK As Salam menggunakan TF-IDF dan Cosine Similarity

Ikhwan Ningrattama^{1*}, Ari Nurul Alfian²

^{1,2}Program Studi Teknik Informatika, Universitas Bina Insani, Indonesia

^{1*}ikhwanningrattama@gmail.com, ²arin@binainsani.ac.id

Abstrak: Perkembangan teknologi informasi mendorong lembaga pendidikan menyediakan layanan informasi yang cepat, mudah diakses, dan interaktif. Chatbot pada website KB & TK As Salam sebelumnya masih menggunakan conditional statement dan pilihan menu, sehingga calon wali murid belum dapat mengajukan pertanyaan menggunakan kalimat sendiri. Selain itu, dataset pertanyaan dan jawaban tersimpan di dalam source code serta percakapan pengguna belum dicatat. Penelitian ini bertujuan mengembangkan chatbot informasi Penerimaan Peserta Didik Baru (PPDB) menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity, dilengkapi halaman pengelolaan dataset dan log percakapan. Metode pengembangan sistem yang digunakan adalah Rapid Application Development (RAD), sedangkan pengumpulan data dilakukan melalui observasi, wawancara, dan studi pustaka. Pertanyaan pengguna diproses melalui case folding, tokenisasi, stopword removal, dan stemming, kemudian diberikan bobot TF-IDF dan dibandingkan dengan 77 pasangan pertanyaan dan jawaban pada basis pengetahuan menggunakan Cosine Similarity. Hasil hold-out testing terhadap 32 pertanyaan uji terpisah menghasilkan accuracy 93,75%, precision 92,86%, recall 100%, dan F1-Score 96,30%. Pengujian fungsional menggunakan Black-Box Testing menunjukkan seluruh 16 skenario memperoleh hasil sesuai. Uji coba pengguna terhadap 10 responden dengan 12 pernyataan menghasilkan nilai rata-rata 4,31 dan persentase penerimaan 86,17% dalam kategori Sangat Setuju. Dengan demikian, chatbot berhasil dikembangkan untuk memberikan informasi PPDB secara lebih fleksibel serta membantu administrator mengelola basis pengetahuan dan mengevaluasi percakapan pengguna.

Kata Kunci: Chatbot; Cosine Similarity; PPDB; TF-IDF; Website

Abstract: The development of information technology encourages educational institutions to provide fast, accessible, and interactive information services. The chatbot on the KB & TK As Salam website previously relied on conditional statements and menu-based questions, preventing prospective parents from submitting questions in their own words. In addition, the question-and-answer dataset was stored in the source code, and user conversations were not recorded. This study aims to develop a New Student Admission (PPDB) information chatbot using Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity, supported by dataset management and conversation log features. The system was developed using Rapid Application Development (RAD), while data were collected through observation, interviews, and literature review. User questions were processed through case folding, tokenization, stopword removal, and stemming, weighted using TF-IDF, and compared with 77 question-and-answer pairs in the

knowledge base using Cosine Similarity. Hold-out testing on 32 separate test questions produced an accuracy of 93.75%, precision of 92.86%, recall of 100%, and an F1-Score of 96.30%. Functional testing using Black-Box Testing showed that all 16 scenarios produced the expected results. User testing involving 10 respondents and 12 statements obtained an average score of 4.31 and an acceptance percentage of 86.17%, categorized as Strongly Agree. Therefore, the chatbot was successfully developed to provide more flexible PPDB information services while supporting administrators in managing the knowledge base and evaluating user conversations.

Keywords: Chatbot; Cosine Similarity; PPDB; TF-IDF; Website

1. PENDAHULUAN

Pemanfaatan teknologi informasi pada lembaga pendidikan tidak hanya mendukung pembelajaran, tetapi juga mempercepat penyampaian informasi kepada masyarakat. Chatbot menjadi salah satu bentuk layanan otomatis yang dapat digunakan untuk memberikan respons secara konsisten dan mudah diakses. Ula et al. menunjukkan bahwa chatbot dapat meningkatkan efektivitas layanan informasi akademik [1], sedangkan Anbiya et al. melaporkan bahwa otomatisasi respons berbasis kecerdasan buatan dapat meningkatkan kualitas layanan informasi pada pendidikan tinggi [2]. Dalam konteks penerimaan peserta didik baru, digitalisasi informasi juga penting karena PPDB berkaitan langsung dengan proses pendaftaran, persyaratan, jadwal, biaya, dan komunikasi awal antara sekolah dengan calon orang tua atau wali murid [15], [16].

KB & TK As Salam telah memiliki website dan widget Pusat Bantuan untuk menyampaikan informasi PPDB. Namun, berdasarkan observasi dan wawancara dengan Huriyah Syahla selaku Staf Administrasi pada 29 Juli 2026, chatbot yang berjalan masih menggunakan conditional statement dan daftar pertanyaan yang harus dipilih pengguna. Chatbot belum menyediakan masukan teks bebas, sehingga calon wali murid tidak dapat mengajukan pertanyaan dengan susunan kalimat sendiri. Selain itu, pertanyaan dan jawaban tersimpan di dalam source code sehingga pembaruan informasi masih bergantung pada pihak yang memahami kode program. Sistem juga belum menyimpan log percakapan, sehingga sekolah belum memiliki data untuk menilai pertanyaan yang sering muncul atau pertanyaan yang belum dapat dijawab.

Berbagai penelitian telah mengembangkan chatbot pendidikan dengan metode dan platform yang berbeda. Apriliani et al. mengimplementasikan pemrosesan bahasa alami pada chatbot sekolah berbasis website [3]. Nur Aziza et al. menggunakan Cosine Similarity pada chatbot informasi akademik dan melaporkan akurasi 90% [4]. Al Farizi et al. menerapkan Levenshtein Distance pada chatbot layanan PPDB dan memperoleh tingkat keberhasilan lebih dari 80% [5]. Setiawan dan Adnyana menggunakan TF-IDF dan Cosine Similarity untuk meningkatkan performa chatbot helpdesk [6]. Amirulloh et al. mengembangkan chatbot PPDB melalui WhatsApp [7], sedangkan Santoso et al. mengembangkan chatbot PPDB berbasis website dengan pendekatan sentence similarity [8]. Di luar chatbot, Tanuwijaya et al. menunjukkan bahwa kombinasi TF-IDF dan Cosine Similarity juga dapat digunakan untuk mengukur relevansi dokumen teks [9].

Kajian tersebut menunjukkan bahwa pencocokan berbasis kemiripan teks dapat diterapkan pada berbagai layanan informasi. Namun, penelitian yang ditinjau memiliki perbedaan objek, platform, metode, dan fitur. Kebutuhan pada KB & TK As Salam bukan sekadar membangun chatbot baru, melainkan mengembangkan widget yang sudah tersedia dari mekanisme pilihan menu berbasis conditional statement menjadi pencocokan teks bebas, sekaligus menyediakan basis pengetahuan yang dapat dikelola administrator dan pencatatan percakapan untuk evaluasi. Oleh karena itu, celah penelitian terletak pada

integrasi metode pencocokan teks dengan kebutuhan operasional pengelolaan informasi PPDB pada website sekolah yang sudah berjalan.

Pendekatan yang diusulkan menggunakan praproses teks berupa case folding, tokenisasi, stopword removal, dan stemming, dilanjutkan pembobotan TF-IDF dan pengukuran Cosine Similarity. TF-IDF digunakan untuk memberikan bobot pada term sesuai tingkat kepentingannya, sedangkan Cosine Similarity membandingkan sudut antardua vektor teks [6], [9], [10], [17]. Sistem memilih pertanyaan pada basis pengetahuan dengan nilai kemiripan tertinggi dan memberikan jawaban jika nilai tersebut memenuhi threshold. Kontribusi penelitian ini bukan algoritma baru, melainkan pengembangan sistem yang menggabungkan pencocokan TF-IDF–Cosine Similarity dengan pengelolaan dataset secara dinamis dan log percakapan pada widget PPDB berbasis website.

Penelitian ini bertujuan menganalisis kebutuhan pengembangan chatbot PPDB, menerapkan TF-IDF dan Cosine Similarity untuk mencocokkan variasi pertanyaan calon wali murid, mengembangkan halaman administrator untuk mengelola dataset serta log percakapan, dan mengevaluasi keberhasilan fungsi serta ketepatan jawaban chatbot. Secara praktis, sistem diharapkan memberi kebebasan kepada calon wali murid untuk bertanya menggunakan kalimat sendiri, membantu administrator memperbarui informasi tanpa mengubah source code, dan menyediakan data percakapan yang dapat digunakan sebagai dasar pengembangan basis pengetahuan berikutnya.

2. METODE PENELITIAN

A. Desain Penelitian dan Pengumpulan Data

Penelitian menggunakan Rapid Application Development (RAD) yang terdiri atas Requirements Planning, User Design, Construction, dan Cutover. RAD dipilih karena mendukung pengembangan sistem berbasis kebutuhan pengguna melalui proses iteratif yang relatif cepat [11]. Pada tahap Requirements Planning dilakukan observasi dan wawancara pada 29 Juli 2026 untuk mengidentifikasi keterbatasan chatbot lama serta kebutuhan informasi PPDB. Wawancara dilakukan dengan Huriyah Syahla selaku Staf Administrasi KB & TK As Salam. Tahap User Design mencakup rancangan alur chatbot, praproses teks, basis data, antarmuka pengguna dan administrator. Tahap Construction mengimplementasikan fungsi pencocokan teks, pengelolaan dataset, dan log percakapan. Pada tahap Cutover, widget diintegrasikan ke website dan diuji secara fungsional serta berdasarkan ketepatan jawaban.

B. Basis Pengetahuan dan Alur Sistem

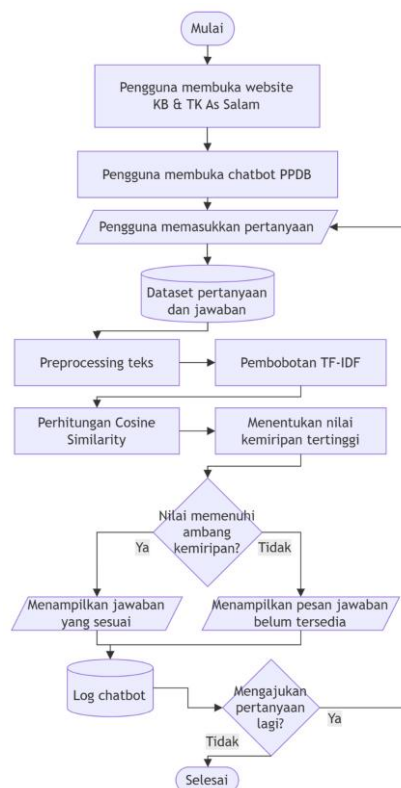
Basis pengetahuan terdiri atas 77 pasangan pertanyaan dan jawaban yang berasal dari informasi resmi KB & TK As Salam dan dikonfirmasi oleh Staf Administrasi. Data mencakup informasi PPDB, biaya, fasilitas, kegiatan, profil, informasi umum, dan kontak. Distribusi dataset ditunjukkan pada Tabel 1. Seluruh 77 data tetap digunakan sebagai basis pengetahuan karena sistem tidak melalui proses pelatihan model seperti pada algoritma machine learning berbasis parameter.

Tabel 1. Distribusi dataset chatbot

Kategori	Jumlah	Persentase
Kegiatan	20	25,97%
Profil	13	16,88%
Biaya	10	12,99%
Fasilitas	10	12,99%
PPDB	10	12,99%

Kategori	Jumlah	Persentase
Umum	9	11,69%
Kontak	5	6,49%
Total	77	100,00%

Alur sistem dimulai ketika pengguna membuka widget chatbot dan memasukkan pertanyaan. Sistem mengambil pertanyaan pada basis pengetahuan, melakukan praproses pada pertanyaan pengguna dan data referensi, membentuk representasi TF-IDF, menghitung Cosine Similarity, memilih nilai kemiripan tertinggi, memeriksa threshold, menampilkan jawaban atau fallback, lalu menyimpan percakapan ke log. Gambar 1 menunjukkan alur pemrosesan tersebut.



Gambar 1. Alur pemrosesan chatbot informasi PPDB

C. Praproses, TF-IDF, dan Cosine Similarity

Praproses hanya diterapkan pada pertanyaan pengguna dan kolom pertanyaan dalam basis pengetahuan. Kolom jawaban tidak dipraproses karena jawaban disimpan dalam bentuk asli dan hanya ditampilkan setelah pertanyaan referensi terbaik ditemukan. Tujuan praproses adalah menyeragamkan bentuk teks sehingga variasi huruf, tanda baca, kata umum, dan kata berimbuhan tidak terlalu memengaruhi proses pembentukan vektor.

Tahapan praproses terdiri atas case folding, pembersihan tanda baca dan tokenisasi, stopword removal, serta stemming. Case folding mengubah seluruh huruf menjadi huruf kecil. Pembersihan tanda baca menghapus karakter yang tidak diperlukan, kemudian tokenisasi memecah kalimat menjadi kata-kata tunggal. Stopword removal menghapus kata umum seperti "apa", "saja", "untuk", "ke", "as", dan "salam". Kata "kb" dan "tk" tetap dipertahankan karena keduanya membedakan jenjang pendidikan. Tahap stemming selanjutnya mengubah kata berimbuhan menjadi bentuk dasar, misalnya "persyaratan"

menjadi "syarat" dan "mendaftarkan" menjadi "daftar". Contoh transformasi pertanyaan ditunjukkan pada Tabel 2.

Tabel 2. Contoh tahapan praproses teks

Tahapan	Hasil
Teks awal	Apa saja persyaratan untuk mendaftar anak ke KB As Salam?
Case folding	apa saja persyaratan untuk mendaftar anak ke kb as salam?
Tokenisasi	[apa, saja, persyaratan, untuk, mendaftar, anak, ke, kb, as, salam]
Stopword removal	[persyaratan, mendaftar, anak, kb]
Stemming	[syarat, daftar, anak, kb]

Hasil praproses digunakan untuk membentuk kosakata (vocabulary) dan menghitung bobot Term Frequency-Inverse Document Frequency (TF-IDF). Penelitian ini menggunakan TF mentah, yaitu jumlah kemunculan term dalam sebuah dokumen. Document Frequency (DF) dihitung hanya dari dokumen referensi pada basis pengetahuan; pertanyaan pengguna tidak dimasukkan sebagai dokumen ketika menghitung DF dan IDF. Nilai IDF yang diperoleh dari korpus referensi kemudian digunakan untuk membentuk vektor pertanyaan pengguna dan seluruh pertanyaan dalam basis pengetahuan. Formulasi TF-IDF ditunjukkan pada (1).

$$w(t,d) = tf(t,d) \times \log_{10}\left(\frac{N}{df(t)}\right) \quad (1)$$

Pada (1), $w(t,d)$ adalah bobot term t pada dokumen d , $tf(t,d)$ adalah frekuensi kemunculan term, N adalah jumlah dokumen referensi, dan $df(t)$ adalah jumlah dokumen referensi yang mengandung term. Kata yang muncul pada sedikit dokumen memperoleh IDF lebih tinggi dan menjadi lebih diskriminatif, sedangkan kata yang muncul pada seluruh dokumen memperoleh IDF 0. Pendekatan ini digunakan untuk menonjolkan term yang lebih spesifik dalam proses pencocokan [6], [9], [17].

Setelah setiap pertanyaan direpresentasikan sebagai vektor TF-IDF, kemiripan antara pertanyaan pengguna dan setiap pertanyaan referensi dihitung menggunakan Cosine Similarity sebagaimana ditunjukkan pada (2) [10].

$$\cos(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} \quad (2)$$

Nilai Cosine Similarity berada pada rentang 0 sampai 1. Sistem menghitung similarity terhadap seluruh 77 pertanyaan referensi, kemudian memilih kandidat dengan nilai tertinggi. Pada proses pemeringkatan, nilai similarity diberi penyesuaian berdasarkan bobot prioritas data. Bobot prioritas tidak menggantikan nilai similarity, melainkan hanya membantu menentukan kandidat ketika terdapat nilai yang berdekatan. Rumus penyesuaian nilai ditunjukkan berikut.

$$nilai_{akhir} = similarity \times (1 + bobot \times 0,01)$$

Setelah kandidat terbaik diperoleh, nilai similarity dibandingkan dengan threshold 0,1. Jika $similarity \geq 0,1$, jawaban yang terhubung dengan kandidat ditampilkan; jika $similarity < 0,1$, sistem menampilkan jawaban fallback.

D. Contoh Perhitungan TF-IDF dan Cosine Similarity

Perhitungan manual digunakan untuk memperjelas proses pembentukan bobot TF-IDF dan pengukuran kemiripan yang diterapkan pada sistem. Contoh menggunakan satu pertanyaan pengguna dan tiga pertanyaan referensi agar setiap tahap perhitungan dapat ditelusuri secara langsung. Pertanyaan pengguna adalah "Bagaimana cara daftar PPDB?" yang setelah praproses menghasilkan $Q = [daftar, ppdb]$. Data referensi yang digunakan adalah $D1 = [daftar, ppdb]$, $D2 = [biaya, masuk, ppdb]$, dan $D3 = [jadwal, masuk, ppdb]$. Pertanyaan Q tidak dihitung sebagai dokumen referensi, sehingga jumlah dokumen untuk perhitungan IDF adalah $N = 3$.

Vocabulary dibentuk dari seluruh term yang terdapat pada Q dan ketiga dokumen referensi, yaitu $V = \{\text{daftar, ppdb, biaya, masuk, jadwal}\}$. Urutan vocabulary tersebut digunakan secara konsisten ketika membentuk seluruh vektor. Penelitian menggunakan TF mentah, sehingga nilai TF adalah 1 apabila suatu term muncul satu kali pada dokumen dan 0 apabila term tidak muncul. Document Frequency (DF) dihitung hanya dari D1, D2, dan D3.

1) Perhitungan DF dan IDF

Term "daftar" muncul hanya pada D1 sehingga $DF = 1$, term "ppdb" muncul pada D1, D2, dan D3 sehingga $DF = 3$, term "biaya" hanya muncul pada D2 sehingga $DF = 1$, term "masuk" muncul pada D2 dan D3 sehingga $DF = 2$, dan term "jadwal" hanya muncul pada D3 sehingga $DF = 1$. Nilai IDF dihitung menggunakan $\log_{10}(N/df)$. Contoh perhitungan IDF ditunjukkan pada persamaan berikut.

$$IDF(\text{daftar}) = \log_{10}\left(\frac{3}{1}\right) = 0,477$$

$$IDF(\text{ppdb}) = \log_{10}\left(\frac{3}{3}\right) = 0$$

$$IDF(\text{masuk}) = \log_{10}\left(\frac{3}{2}\right) = 0,176$$

Term "ppdb" memperoleh bobot 0 karena muncul pada seluruh dokumen referensi dan tidak dapat membedakan satu dokumen dari dokumen lainnya.

2) Perhitungan Bobot TF-IDF

Bobot TF-IDF diperoleh dengan mengalikan TF dengan IDF. Contoh perhitungan bobot term "daftar" pada Q dan D1 ditunjukkan berikut. Pada D2 dan D3, TF term "daftar" bernilai 0 sehingga bobot TF-IDF-nya 0. Hasil pembobotan seluruh term ditunjukkan pada Tabel 3.

$$TF - IDF(\text{daftar}, Q) = 1 \times 0,477 = 0,477$$

$$TF - IDF(\text{daftar}, D1) = 1 \times 0,477 = 0,477$$

Tabel 3. Hasil perhitungan DF, IDF, dan bobot TF-IDF

Term	DF	IDF	TF-IDF Q	TF-IDF D1	TF-IDF D2	TF-IDF D3
daftar	1	0,477	0,477	0,477	0	0
ppdb	3	0	0	0	0	0
biaya	1	0,477	0	0	0,477	0
masuk	2	0,176	0	0	0,176	0,176
jadwal	1	0,477	0	0	0	0,477

Berdasarkan Tabel 3, vektor dibentuk menggunakan urutan term [daftar, ppdb, biaya, masuk, jadwal]. Vektor pertanyaan pengguna dan ketiga dokumen referensi adalah sebagai berikut.

$$Q = [0,477; 0; 0; 0; 0]$$

$$D_1 = [0,477; 0; 0; 0; 0]$$

$$D_2 = [0; 0; 0,477; 0,176; 0]$$

$$D_3 = [0; 0; 0; 0,176; 0,477]$$

Vektor tersebut selanjutnya digunakan pada perhitungan Cosine Similarity.

3) Perhitungan Cosine Similarity

Kemiripan Q dengan D1 dihitung melalui dot product dan panjang vektor. Tahapan perhitungannya ditunjukkan pada persamaan berikut.

$$Q \cdot D_1 = (0,477121 \times 0,477121) = 0,227645$$

$$\|Q\| = \sqrt{(0,477121^2)} = 0,477121$$

$$\|D_1\| = \sqrt{(0,477121^2)} = 0,477121$$

$$\cos(Q, D_1) = \frac{0,227645}{0,477121 \times 0,477121} = 1,000$$

Pada Q dengan D2 dan D3 tidak terdapat term yang sama-sama memiliki bobot lebih dari 0, sehingga dot product bernilai 0 dan Cosine Similarity juga bernilai 0.

Tabel 4. Hasil perhitungan Cosine Similarity

Perbandingan	Dot Product	Q	D	Cosine Similarity
Q dengan D1	0,227645	0,477121	0,477121	1,000
Q dengan D2	0	0,477121	0,508579	0,000
Q dengan D3	0	0,477121	0,508579	0,000

D1 memiliki nilai Cosine Similarity tertinggi, yaitu 1,000, sehingga dipilih sebagai pertanyaan referensi yang paling mirip dengan Q. Sistem kemudian mengambil jawaban yang terhubung dengan D1. Contoh manual ini hanya memperlihatkan pembobotan TF-IDF dan Cosine Similarity utama. Pada implementasi sebenarnya, proses yang sama dilakukan terhadap seluruh 77 pertanyaan referensi, kemudian nilai similarity digunakan untuk memilih kandidat terbaik dan diperiksa terhadap threshold 0,1. Jika nilai similarity tertinggi berada di bawah threshold, chatbot menampilkan jawaban fallback.

E. Prosedur Evaluasi

Evaluasi sistem dilakukan melalui tiga tahap, yaitu evaluasi ketepatan jawaban chatbot, pengujian fungsional, dan uji pengguna. Evaluasi ketepatan jawaban menggunakan 32 pertanyaan uji independen yang tidak dimasukkan ke dalam 77 data referensi. Ketiga puluh dua pertanyaan uji tersebut disusun dalam beberapa kategori untuk merepresentasikan variasi pertanyaan pengguna, yaitu pertanyaan relevan dengan susunan kalimat berbeda, bahasa informal, singkatan, kesalahan penulisan, dan pertanyaan di luar basis pengetahuan. Seluruh 77 data tetap digunakan sebagai dokumen referensi, sedangkan 32 pertanyaan uji digunakan sebagai query independen untuk mengevaluasi kemampuan chatbot dalam mencocokkan variasi pertanyaan pengguna.

Pertanyaan uji terdiri atas 32 pertanyaan yang dikelompokkan berdasarkan karakteristik input pengguna. Kategori pengujian meliputi pertanyaan relevan dengan variasi susunan kalimat, bahasa informal, singkatan, kesalahan penulisan, dan pertanyaan di luar basis pengetahuan. Pertanyaan relevan digunakan untuk menguji kemampuan sistem dalam menemukan jawaban yang sesuai pada basis pengetahuan, sedangkan pertanyaan di luar basis pengetahuan digunakan untuk menguji kemampuan sistem dalam menolak kecocokan yang tidak relevan dan memberikan respons fallback.

Hasil evaluasi diklasifikasikan menjadi true positive (TP), true negative (TN), false positive (FP), dan false negative (FN). TP terjadi ketika pertanyaan memiliki jawaban pada basis pengetahuan dan sistem memberikan jawaban relevan. TN terjadi ketika pertanyaan tidak memiliki jawaban yang sesuai dan sistem menampilkan fallback. FP terjadi ketika pertanyaan seharusnya tidak dijawab tetapi sistem memberikan jawaban yang tidak relevan. FN terjadi ketika pertanyaan sebenarnya memiliki jawaban pada basis pengetahuan tetapi sistem gagal memberikan jawaban relevan. Berdasarkan klasifikasi tersebut dihitung accuracy, precision, recall, dan F1-Score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Selain threshold utama sebesar 0,1, dilakukan perbandingan menggunakan threshold 0,3; 0,5; 0,7; dan 0,8 terhadap 32 pertanyaan uji yang sama. Perbandingan digunakan untuk melihat perubahan TP, TN, FP, dan FN serta trade-off antara kemampuan menerima

pertanyaan relevan dan kemampuan menolak pertanyaan yang memiliki kesamaan kata tetapi berbeda konteks. Analisis ini digunakan untuk mengevaluasi threshold yang diterapkan pada sistem.

Pengujian fungsional dilakukan menggunakan Black-Box Testing untuk memastikan keluaran sistem sesuai dengan kebutuhan tanpa memeriksa struktur kode internal [12]. Sebanyak 16 skenario diuji, meliputi fungsi widget, penerimaan dan pencocokan pertanyaan, fallback, penyimpanan log, autentikasi administrator, penambahan, perubahan, dan penghapusan dataset, penggunaan dataset baru, serta pengelolaan log. Setiap skenario dinyatakan Sesuai apabila hasil aktual sama dengan hasil yang diharapkan.

Uji pengguna dilakukan terhadap 10 responden yang telah mencoba widget chatbot. Kuesioner terdiri atas 12 pernyataan yang mencakup kemudahan menemukan dan menggunakan widget, fleksibilitas mengajukan pertanyaan, kecepatan respons, kesesuaian dan kejelasan jawaban, kelengkapan informasi, kejelasan fallback, tampilan antarmuka, dan kepuasan keseluruhan. Penilaian menggunakan Skala Likert lima tingkat [13], [14] sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Skala penilaian uji pengguna

No.	Kategori Penilaian	Singkatan	Bobot
1	Sangat Setuju	SS	5
2	Setuju	S	4
3	Netral	N	3
4	Tidak Setuju	TS	2
5	Sangat Tidak Setuju	STS	1

$$\text{Rata - rata} = \frac{\text{Total skor jawaban}}{\text{Jumlah responden}}$$

$$\text{Persentase penerimaan} = \frac{\text{Total skor diperoleh}}{\text{Skor maksimum}} \times 100\%$$

Nilai rata-rata setiap pernyataan dihitung dengan membagi total skor jawaban dengan jumlah responden. Persentase penerimaan keseluruhan dihitung dengan membagi total skor yang diperoleh dengan skor maksimum, kemudian dikalikan 100%. Skor maksimum pada penelitian ini dihitung sebagai berikut.

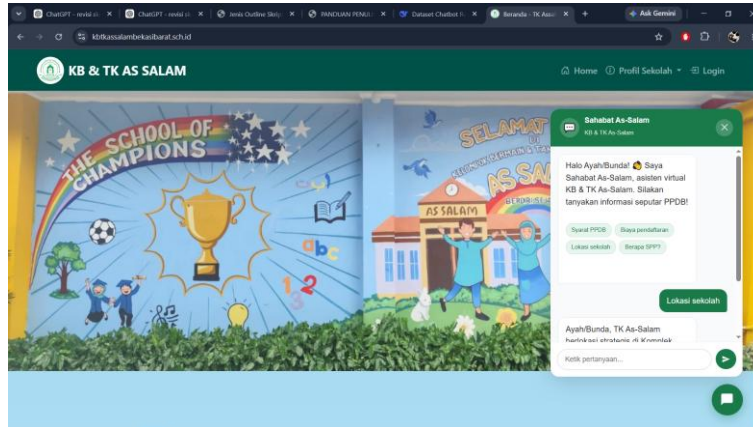
$$\text{Skor}_{maks} = 10 \times 12 \times 5 = 600$$

Hasil uji pengguna digunakan sebagai gambaran awal penerimaan sistem dan tidak dimaksudkan untuk melakukan generalisasi statistik terhadap seluruh populasi calon wali murid.

3. HASIL DAN PEMBAHASAN

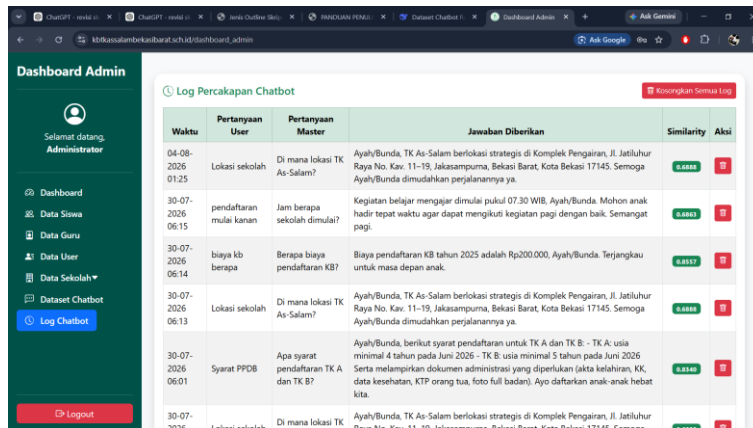
A. Implementasi Sistem

Hasil pengembangan menghasilkan widget "Sahabat As-Salam" yang terintegrasi pada website KB & TK As Salam. Pengguna dapat memilih pertanyaan cepat atau mengetik pertanyaan sendiri melalui kolom masukan. Pertanyaan yang dikirim diproses menggunakan TF-IDF dan Cosine Similarity sebelum jawaban ditampilkan. Gambar 2 memperlihatkan widget yang telah diimplementasikan pada website sekolah.



Gambar 2. Implementasi widget Sahabat As-Salam pada website KB & TK As Salam

Pada sisi administrator, sistem menyediakan pengelolaan dataset dan log percakapan. Log menyimpan waktu, pertanyaan pengguna, pertanyaan referensi terdekat, jawaban yang diberikan, dan nilai similarity. Informasi tersebut membantu administrator mengidentifikasi pertanyaan berulang, memeriksa kesesuaian jawaban, serta menentukan kebutuhan penambahan dataset. Gambar 3 menunjukkan halaman log percakapan yang dihasilkan sistem.



Waktu	Pertanyaan User	Pertanyaan Master	Jawaban Diberikan	Similarity	Aksi
04-08-2026 01:25	Lokasi sekolah	Di mana lokasi TK As-Salam?	Ayah/Bunda, TK As-Salam berlokasi strategis di Komplek Pengairan, Jl. Jatiluhur Raya No. Kav. 11-19, Jakasampurna, Bekasi Barat, Kota Bekasi 17145. Semoga Ayah/Bunda memudahkan perjalanannya ya.	0.6888	[X]
30-07-2026 06:15	pendaftaran mulai kapan	Jam berapa sekolah dimulai?	Kegiatan belajar mengajar dimulai pukul 07.30 WIB, Ayah/Bunda. Mohon anak hadir tepat waktu agar dapat mengikuti kegiatan pagi dengan baik. Semangat pagi.	0.6883	[X]
30-07-2026 06:14	biaya kb berapa	Berapa biaya pendaftaran KB?	Biaya pendaftaran KB tahun 2025 adalah Rp200.000, Ayah/Bunda. Terjangkau untuk masa depan anak.	0.6887	[X]
30-07-2026 06:13	Lokasi sekolah	Di mana lokasi TK As-Salam?	Ayah/Bunda, TK As-Salam berlokasi strategis di Komplek Pengairan, Jl. Jatiluhur Raya No. Kav. 11-19, Jakasampurna, Bekasi Barat, Kota Bekasi 17145. Semoga Ayah/Bunda memudahkan perjalanannya ya.	0.6888	[X]
30-07-2026 06:01	Syarat PPDB	Apa syarat pendaftaran TK A dan TK B?	Ayah/Bunda, berikut syarat pendaftaran untuk TK A dan TK B - TK A: usia minimal 4 tahun pada Juni 2026 - TK B: usia minimal 5 tahun pada Juni 2026. Serta melampirkan dokumen administrasi yang diperlukan (akta kelahiran, KK, data kesehatan, KTP orang tua, foto full badan). Ayo daftarkan anak-anak hebat kita.	0.6849	[X]
30-07-2026	Lokasi sekolah	Di mana lokasi TK	Ayah/Bunda, TK As-Salam berlokasi strategis di Komplek Pengairan, Jl. Jatiluhur Raya No. Kav. 11-19, Jakasampurna, Bekasi Barat, Kota Bekasi 17145. Semoga	0.6888	[X]

Gambar 3. Halaman log percakapan chatbot pada sisi administrator

B. Hasil Evaluasi Ketepatan Jawaban

Dari 32 pertanyaan uji diperoleh 26 TP, 4 TN, 2 FP, dan 0 FN. Dengan demikian, sistem memberikan 30 hasil benar dari 32 pengujian. Rekapitulasi confusion matrix dan metrik evaluasi ditunjukkan pada Tabel 6.

Tabel 6. Hasil evaluasi awal model

Indikator	Nilai
True Positive (TP)	26
True Negative (TN)	4
False Positive (FP)	2
False Negative (FN)	0
Accuracy	93,75%
Precision	92,86%
Recall	100%
F1-Score	96,30%

Dari 32 pertanyaan uji diperoleh 26 true positive (TP), 4 true negative (TN), 2 false positive (FP), dan 0 false negative (FN). Dengan demikian, sistem memberikan 30 hasil klasifikasi yang benar dari 32 pengujian. Salah satu false positive terjadi pada pertanyaan "Ada berapa jumlah ruang kelas di sekolah?" yang memperoleh nilai similarity 0,79 dan menghasilkan jawaban mengenai fasilitas dalam ruangan. Kesamaan token "ruang" dan "kelas" menyebabkan nilai kemiripan menjadi tinggi, meskipun konteks pertanyaan yang ditanyakan adalah jumlah ruang kelas. Kondisi ini menunjukkan bahwa TF-IDF dan Cosine Similarity masih bergantung pada kesamaan kata berbobot dan belum memahami konteks semantik secara mendalam.

Nilai recall sebesar 100% menunjukkan bahwa seluruh pertanyaan yang memiliki jawaban pada basis pengetahuan berhasil dikenali pada threshold yang digunakan. Precision sebesar 92,86% menunjukkan bahwa dari 28 pertanyaan yang diprediksi dapat dijawab, terdapat 2 pertanyaan yang menghasilkan jawaban tidak relevan. F1-Score sebesar 96,30% menunjukkan keseimbangan antara precision dan recall pada 32 pertanyaan uji.

C. Analisis Threshold

Pengaruh perubahan threshold Cosine Similarity dianalisis untuk menentukan nilai ambang yang mampu memberikan keseimbangan antara penerimaan pertanyaan yang relevan dan penolakan pertanyaan yang tidak memiliki jawaban pada basis pengetahuan. Pengujian dilakukan dengan membandingkan beberapa nilai threshold terhadap data uji yang sama.

Tabel 7. Perbandingan hasil berdasarkan threshold

Threshold	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
0,1	26	4	2	0	93,75%	92,86%	100%	96,30%
0,3	26	4	2	0	93,75%	92,86%	100%	96,30%
0,5	24	4	2	2	87,50%	92,31%	92,31%	92,31%
0,7	22	4	2	4	81,25%	91,67%	84,62%	88,00%
0,8	20	6	0	6	81,25%	100%	76,92%	86,96%

Berdasarkan hasil pengujian, threshold 0,1 dan 0,3 menghasilkan performa yang sama, yaitu accuracy 93,75%, precision 92,86%, recall 100%, dan F1-Score 96,30%. Kesamaan hasil tersebut terjadi karena tidak terdapat pertanyaan uji dengan nilai similarity pada rentang 0,1 hingga kurang dari 0,3. Pada threshold 0,5 mulai terdapat pertanyaan relevan dengan similarity 0,47 yang ditolak sehingga muncul false negative. Peningkatan threshold menjadi 0,7 dan 0,8 menyebabkan jumlah pertanyaan relevan yang ditolak semakin bertambah dan recall menurun. Sebaliknya, threshold 0,8 mampu menghilangkan false positive, tetapi menyebabkan enam pertanyaan relevan tidak memperoleh jawaban yang sesuai.

Berdasarkan perbandingan tersebut, threshold 0,1 dipilih sebagai threshold yang digunakan dalam sistem karena menghasilkan performa evaluasi tertinggi bersama threshold 0,3 serta mempertahankan recall sebesar 100%. Pemilihan ini juga mempertimbangkan kebutuhan sistem informasi PPDB yang lebih mengutamakan kemampuan memberikan respons terhadap pertanyaan yang memiliki informasi pada basis pengetahuan. Meskipun demikian, threshold 0,1 belum dapat dinyatakan sebagai nilai optimal secara umum karena hasil evaluasi masih bergantung pada karakteristik dan jumlah data uji yang digunakan.

D. Hasil Pengujian Fungsional

Black-Box Testing dilakukan terhadap 16 skenario. Seluruh skenario menghasilkan keluaran yang sesuai dengan hasil yang diharapkan. Ringkasan fungsi utama ditampilkan

pada Tabel 8. Hasil ini menunjukkan bahwa peningkatan algoritma pencocokan telah terintegrasi dengan fungsi operasional yang dibutuhkan pengguna dan administrator.

Tabel 8. Ringkasan Black-Box Testing

Kelompok fungsi	Skenario yang diperiksa	Hasil
Widget chatbot	Menampilkan, membuka, menutup widget, dan menerima pertanyaan	Sesuai
Pencocokan jawaban	Menjawab pertanyaan dataset dan menampilkan fallback	Sesuai
Log percakapan	Menyimpan dan menampilkan log	Sesuai
Autentikasi	Login benar dan login salah	Sesuai
Dataset chatbot	Tambah, ubah, hapus, dan menggunakan dataset baru	Sesuai
Pengelolaan log	Menghapus satu log dan seluruh log	Sesuai

Keberhasilan seluruh 16 skenario memperlihatkan bahwa sistem tidak hanya mampu menghitung kemiripan, tetapi juga mendukung penggunaan praktis melalui autentikasi administrator, pembaruan basis pengetahuan, dan pencatatan interaksi. Dataset yang baru ditambahkan dapat langsung digunakan dalam proses pencocokan, sehingga pembaruan informasi tidak lagi memerlukan perubahan pada kode program.

E. Hasil Uji Pengguna

Uji pengguna melibatkan 10 responden dan 12 pernyataan. Jumlah skor keseluruhan adalah 517 dari skor maksimum 600, sehingga persentase penerimaan mencapai 86,17%. Nilai rata-rata keseluruhan adalah 4,31 dan termasuk kategori Sangat Setuju. Rincian nilai rata-rata setiap pernyataan ditunjukkan pada Tabel 9.

Tabel 9. Hasil uji pengguna

No.	Aspek yang dinilai	Rata-rata	Kategori
1	Widget mudah ditemukan	4,30	Sangat Setuju
2	Widget mudah dibuka dan digunakan	4,60	Sangat Setuju
3	Kolom pertanyaan dan tombol mudah dipahami	4,20	Setuju
4	Dapat bertanya dengan kalimat sendiri	4,20	Setuju
5	Respons diberikan dengan cepat	4,50	Sangat Setuju
6	Jawaban sesuai dengan pertanyaan	4,30	Sangat Setuju
7	Bahasa jawaban mudah dipahami	4,20	Setuju
8	Informasi membantu memahami proses pendaftaran	4,00	Setuju
9	Cakupan informasi PPDB dinilai memadai	4,20	Setuju
10	Pesan fallback mudah dipahami	4,50	Sangat Setuju
11	Tampilan widget jelas dan nyaman	4,30	Sangat Setuju
12	Kepuasan keseluruhan	4,40	Sangat Setuju
	Rata-rata keseluruhan	4,31	Sangat Setuju

Nilai tertinggi terdapat pada kemudahan membuka dan menggunakan widget, yaitu 4,60. Nilai terendah adalah 4,00 pada pernyataan bahwa informasi chatbot membantu pengguna memahami proses pendaftaran, tetapi masih berada pada kategori Setuju. Hasil tersebut menunjukkan penerimaan awal yang positif, terutama pada kemudahan penggunaan dan kecepatan respons. Karena jumlah responden terbatas, hasil ini lebih tepat

dipahami sebagai evaluasi awal penerimaan sistem daripada ukuran representatif untuk seluruh calon wali murid.

F. Pembahasan dan Keterbatasan

Hasil penelitian mendukung penggunaan pencocokan berbasis kemiripan teks untuk layanan informasi pendidikan. Setiawan dan Adnyana melaporkan F1-score 75% pada chatbot helpdesk berbasis TF-IDF dan Cosine Similarity [6], sedangkan Nur Aziza et al. melaporkan akurasi 90% pada chatbot informasi akademik [4]. Pada penelitian ini, accuracy 93,75% dan F1-Score 96,30% menunjukkan hasil yang baik pada 32 pertanyaan uji independen. Namun, nilai tersebut tidak dapat dibandingkan secara langsung sebagai keunggulan mutlak karena setiap penelitian menggunakan objek, ukuran basis pengetahuan, komposisi pertanyaan uji, dan prosedur evaluasi yang berbeda. Temuan penting penelitian ini justru terletak pada kemampuan sistem menggantikan mekanisme pilihan menu dengan input teks bebas, sekaligus mempertahankan pengelolaan basis pengetahuan dan log percakapan melalui halaman administrator.

Kontribusi praktis penelitian ini adalah perubahan mekanisme chatbot dari pilihan menu berbasis conditional statement menjadi masukan teks bebas, penyediaan CRUD dataset oleh administrator, serta pencatatan log percakapan sebagai bahan evaluasi. Keterbatasan utama adalah ukuran basis pengetahuan 77 pasangan pertanyaan-jawaban, jumlah pertanyaan uji 32, dan uji pengguna yang masih melibatkan 10 responden. Selain itu, dua false positive menunjukkan bahwa TF-IDF dan Cosine Similarity masih sensitif terhadap kesamaan kata dan belum memahami konteks secara mendalam. Salah satu contohnya adalah pertanyaan mengenai jumlah ruang kelas yang memperoleh kecocokan berdasarkan kesamaan token. Karena itu, hasil penelitian sebaiknya dipahami sebagai evaluasi pada lingkup data dan skenario yang diuji, bukan sebagai ukuran kinerja universal untuk seluruh variasi pertanyaan PPDB.

4. KESIMPULAN

Pengembangan chatbot informasi PPDB KB & TK As Salam berhasil mengubah mekanisme layanan dari pilihan menu berbasis conditional statement menjadi pencocokan pertanyaan berbentuk teks menggunakan TF-IDF dan Cosine Similarity. Basis pengetahuan terdiri atas 77 pasangan pertanyaan-jawaban dan didukung halaman administrator untuk menambah, mengubah, serta menghapus dataset dan melihat log percakapan. Evaluasi terhadap 32 pertanyaan uji independen menghasilkan accuracy 93,75%, precision 92,86%, recall 100%, dan F1-Score 96,30%, dengan 26 true positive, 4 true negative, 2 false positive, dan 0 false negative. Dua false positive menunjukkan bahwa terdapat pertanyaan di luar basis pengetahuan yang masih dapat memperoleh kecocokan berdasarkan kesamaan kata. Seluruh 16 skenario Black-Box Testing berjalan sesuai hasil yang diharapkan. Uji pengguna terhadap 10 responden menghasilkan rata-rata 4,31 dan persentase penerimaan 86,17% atau kategori Sangat Setuju. Hasil tersebut menunjukkan bahwa tujuan penelitian telah tercapai pada lingkup data dan pengguna yang diuji: chatbot dapat menerima pertanyaan bebas, memberikan informasi PPDB, serta mendukung pembaruan basis pengetahuan dan evaluasi interaksi. Keterbatasan penelitian terletak pada ukuran dataset, jumlah pertanyaan uji dan responden yang masih terbatas, serta keterbatasan metode dalam memahami konteks semantik. Penelitian selanjutnya dapat memperluas data, mengevaluasi threshold pada sampel lebih besar, dan membandingkan pendekatan ini dengan metode representasi teks yang lebih kontekstual.

5. REFERENCES

- [1] M. Ula, R. Hardi, and I. Hipiny, "An Improved Structure for Academic Information Services through AI Chatbots," *Journal of Engineering Science and Technology Review*, vol. 16, no. 5, pp. 164–173, 2023, doi: 10.25103/jestr.165.20.
- [2] M. A. N. Islam, B. Warsito, and O. D. Nurhayati, "AI-Driven Chatbot Implementation for Enhancing Customer Service in Higher Education: A Case Study from Universitas Negeri Semarang," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 14, pp. 5690–5701, 2024.
- [3] D. Apriliani, S. F. Handayani, and I. T. Saputra, "Implementasi Natural Language Processing (NLP) Dalam Pengembangan Aplikasi Chatbot Pada SMK YPE Nusantara Slawi," *Techno.Com*, vol. 22, no. 4, pp. 1037–1047, 2023, doi: 10.33633/tc.v22i4.9155.
- [4] R. N. Aziza, E. Yosrita, F. R. Ningrum, S. T. Ardanti, and R. S. Arafazain, "Pembangunan Aplikasi dan Klasifikasi Pertanyaan Chatbot Informasi Akademik Menggunakan Metode Cosine Similarity dan Naïve Bayes," *KILAT*, vol. 12, no. 2, pp. 169–179, 2024, doi: 10.33322/kilat.v12i2.1921.
- [5] M. Al Farizi, I. N. Farida, and D. Swanjaya, "Penerapan Chatbot Berbasis Natural Language Processing untuk Layanan Informasi PPDB di SMK PGRI 1 Kediri," *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 9, no. 2, pp. 1044–1053, 2025, doi: 10.29407/c4x80271.
- [6] G. H. Setiawan and I. M. Adnyana, "Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 252–257, 2023, doi: 10.30871/jaic.v7i2.6527.
- [7] I. Amirulloh, M. W. Pertiwi, and T. Wibisono, "Rancang Bangun Chatbot Whatsapp Menggunakan Node Js Dan Model Natural Language Processing Untuk Layanan PPDB SMK YPC Tasikmalaya," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, pp. 541–550, 2024, doi: 10.23960/jitet.v12i1.3846.
- [8] B. Santoso, Jefi, B. A. Wahid, F. E. Schadu, and D. R. Zain, "Perancangan Program Chatbot Berbasis Website Dalam Sistem Informasi Dan Pelayanan PPDB Pada SMK Al-Bahri Bekasi," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11260–11268, 2024, doi: 10.36040/jati.v8i6.11247.
- [9] W. Tanuwijaya, C. E. Setiawan, H. Irsyad, and A. Rahman, "Implementasi TF-IDF dan Cosine Similarity untuk Penyaringan Dokumen Berita Program Makan Siang Gratis Pemerintah Indonesia," *DEVICE: Journal of Information System, Computer Science and Information Technology*, vol. 6, no. 2, pp. 322–334, 2025, doi: 10.46576/device.v6i2.6724.
- [10] A. Lahitani, U. S. Aesyi, N. Wulandari, and B. D. Santosa, "Cosine Similarity untuk Mengukur Tingkat Kesadaran pada Topik Software Security Berbasis Teks Komentar di Media Sosial Youtube," *Jurnal Sains dan Informatika*, vol. 8, no. 2, 2022, doi: 10.34128/jsi.v8i2.535.
- [11] T. Pricillia and Zulfachmi, "Perbandingan Metode Pengembangan Perangkat Lunak (Waterfall, Prototype, RAD)," *Jurnal Bangkit Indonesia*, vol. 10, no. 1, pp. 6–12, 2021, doi: 10.52771/bangkitindonesia.v10i1.153.
- [12] A. Maspupah, "Literature Review: Advantages and Disadvantages of Black Box and White Box Testing Methods," *Jurnal Techno Nusa Mandiri*, vol. 21, no. 2, pp. 151–162, 2024, doi: 10.33480/techno.v21i2.5776.
- [13] S. R. Widyastuti, "Pengembangan Skala Likert untuk Mengukur Sikap terhadap Penerapan Penilaian Autentik Siswa Sekolah Menengah Pertama," *Jurnal Jembatan Efektivitas Ilmu dan Akhlak Ahlussunah Wal Jama'ah*, vol. 3, no. 2, pp. 67–75, 2022, doi: 10.52188/ja.v3i02.393.

- [14] M. F. Erinsyah, G. W. Sasmito, D. S. Wibowo, and V. K. Bakti, "Sistem Evaluasi Pada Aplikasi Akademik Menggunakan Metode Skala Likert Dan Algoritma Naïve Bayes," *Komputa: Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 1, pp. 74–82, 2024, doi: 10.34010/komputa.v13i1.10940.
- [15] A. S. Atmaja, A. S. Sidabalok, M. Raihan, F. A. Putra, and N. I. L. Silalahi, "Perancangan Sistem Informasi Penerimaan Peserta Didik Baru (PPDB) Berbasis Web," *Jurnal Komputer Teknologi Informasi Sistem Komputer (JUKTISI)*, vol. 2, no. 3, pp. 515–523, 2024, doi: 10.62712/juktisi.v2i3.101.
- [16] B. Ismaya, I. Subandi, and M. Suryaman, "Mekanisme Sistem Penerimaan Peserta Didik Baru (PPDB) Dalam Rangka Meningkatkan Mutu Lulusan Di SMAN 1 Cikarang Pusat," *Jurnal Tahsinia*, vol. 6, no. 10, pp. 1590–1603, 2025, doi: 10.57171/jg464912.
- [17] D. Andreswari, D. Suranti, and D. A. Trianggara, "Application Of Text Mining In Grouping Thesis Topics Using TF-IDF Method Based On Thesis Abstract," *Jurnal Komputer Indonesia*, vol. 3, no. 2, pp. 69–78, 2024, doi: 10.37676/jki.v3i2.634.