SEMINAR NASIONAL 2026 Universitas Ma Chung
Prosiding Seminar Nasional Universitas Ma Chung 2026
(Sistem Informasi dan Informatika)
pp. 113-125
E-ISSN: 2808-7771

TEXT SIMILARITY PADA DOKUMEN PRODUK *CERTIPORT* DENGAN DOKUMEN PROFIL CALON KLIEN SERTIFIKASI

Hilman Zahrawa Budiarto, Dika Rizky Yunianto,Vit Zuraida

Jurusan Teknologi Informasi, Politeknik Negeri Malang, Jl. Soekarno Hatta No.9, Malang, Indonesia, 65141

***Email korespondensi:** 2241760051@student.polinema.ac.id

Received: 27 05 2026 – Revised: 29 07 2026 - Accepted: 04 08 2026 - Published: 05 08 2026

Abstrak. Proses profiling kebutuhan sertifikasi calon klien di Certiport Authorized Testing Center (CATC) JTI Polinema yang masih dilakukan secara manual rentan terhadap subjektivitas dan terkendala kesenjangan antara narasi kompetensi calon klien dengan terminologi teknis pada dokumen produk sertifikasi. Penelitian ini bertujuan merancang dan mengimplementasikan sistem rekomendasi sertifikasi otomatis berbasis *hybrid text similarity* untuk mengukur tingkat kecocokan antara dokumen *Objective Domain* sertifikasi Certiport dengan dokumen profil calon klien sertifikasi. Sistem dibangun menggunakan arsitektur komputasi hybrid yang mengintegrasikan TF-IDF pada dimensi numerik dan *Sentence Transformers* model *all-MiniLM-L6-v2* pada dimensi semantik, disertai pipeline penerjemahan otomatis untuk menjembatani perbedaan bahasa antara profil calon klien berbahasa Indonesia dengan *Objective Domain* berbahasa Inggris. Pengujian dilakukan terhadap 16 dokumen profil calon klien sertifikasi dari tiga institusi yang merepresentasikan variasi kebutuhan kompetensi lintas domain. Evaluasi dilakukan menggunakan metrik *Precision*, *Recall*, *Accuracy*, dan *F1-Score* untuk mengukur kualitas rekomendasi yang dihasilkan sistem. Hasil pengujian menunjukkan performa optimal pada batas rekomendasi Top-20 dengan rasio pembobotan 40% numerik dan 60% semantik, menghasilkan rata-rata *Precision* sebesar 29,37%, *Recall* sebesar 40,20%, *Accuracy* sebesar 74,68%, dan *F1-Score* sebesar 29,85%. *Ablation study* membuktikan bahwa pendekatan hybrid meningkatkan *F1-Score* sebesar 48,9% dibandingkan TF-IDF Only. Pengujian penerimaan pengguna menghasilkan tingkat kelayakan sebesar 88,57%, yang termasuk dalam kategori sangat layak diimplementasikan. Hasil penelitian ini menunjukkan bahwa CertiMatch mampu membantu proses rekomendasi sertifikasi menjadi lebih cepat, konsisten, dan objektif sehingga dapat mendukung pengambilan keputusan dalam proses profiling kebutuhan sertifikasi calon klien.

Kata kunci: Hybrid Text Similarity, Sentence Transformers, TF-IDF, Sistem Rekomendasi Sertifikasi, Natural Language Processing

PENDAHULUAN

Certiport Authorized Testing Center (CATC) di Jurusan Teknologi Informasi Politeknik Negeri Malang (JTI Polinema) merupakan mitra resmi layanan sertifikasi internasional yang mencakup berbagai bidang sertifikasi dengan kredibilitas global. CATC JTI Polinema menerapkan strategi pemasaran aktif dengan mendekati calon klien institusional, meliputi instansi pemerintah, sektor swasta, serta lembaga pendidikan yang membutuhkan sertifikasi secara kolektif. Certiport menyediakan lebih dari 98 variasi

produk sertifikasi, mulai dari aplikasi perkantoran, desain grafis, hingga pemrograman teknis (Certiport, 2026).

Tantangan utama dalam strategi tersebut adalah melakukan profiling kebutuhan calon klien secara akurat sebelum mengajukan penawaran sertifikasi. Pada tahap pendekatan awal, informasi yang tersedia mengenai calon klien sering kali terbatas pada dokumen profil publik seperti deskripsi bidang kompetensi atau visi-misi organisasi. Proses identifikasi kebutuhan yang saat ini masih dilakukan secara manual sangat mengandalkan perspektif manusia dalam mencocokkan dokumen klien satu per satu, sehingga memicu hambatan operasional berupa waktu yang lama dan tingkat subjektivitas yang tinggi (Pawestri & Suyanto, 2024).

Untuk mengatasi permasalahan tersebut, pendekatan otomatis berbasis text similarity diperlukan untuk mengukur tingkat kecocokan antara dokumen profil calon klien dengan dokumen Objective Domain sertifikasi Certiport. Metode Cosine Similarity dengan pembobotan TF-IDF umum digunakan karena efisiensinya, namun memiliki kelemahan saat menghadapi dokumen dengan gaya bahasa berbeda (Pawestri & Suyanto, 2024). Profil calon klien cenderung menggunakan narasi kompetensi tersirat dalam Bahasa Indonesia, sedangkan Objective Domain Certiport menggunakan istilah teknis terstruktur dalam Bahasa Inggris. Kesenjangan ini memerlukan pendekatan hibrida yang mengintegrasikan analisis semantik berbasis Sentence Transformers (Reimers & Gurevych, 2019) dengan pipeline penerjemahan otomatis lintas bahasa.

Penelitian ini bertujuan merancang dan mengimplementasikan sistem rekomendasi sertifikasi berbasis hybrid text similarity yang menggabungkan TF-IDF dan Sentence Transformers model all-MiniLM-L6-v2, serta mengevaluasi performa sistem menggunakan metrik Precision, Recall, F1-Score, dan Accuracy terhadap ground truth dari pakar.

MASALAH

Permasalahan utama yang diangkat dalam penelitian ini adalah ketidakmampuan pendekatan text similarity konvensional berbasis TF-IDF dalam menjembatani dua kesenjangan struktural sekaligus pada domain profiling sertifikasi. Pertama, kesenjangan bahasa: dokumen Objective Domain Certiport disusun dalam Bahasa Inggris dengan

format teknis terstruktur, sedangkan dokumen profil calon klien umumnya ditulis dalam Bahasa Indonesia dengan gaya narasi kompetensi yang lebih umum. Kedua, kesenjangan kosakata: profil klien sering menggunakan istilah managerial tersirat yang secara leksikal tidak bersinggungan dengan terminologi teknis sertifikasi, namun secara semantik keduanya memiliki kesamaan makna.

Belum terdapat penelitian yang secara spesifik memodelkan sistem rekomendasi profiling pemasaran B2B (Business-to-Business) untuk menjembatani perbedaan struktural antara dokumen standar kompetensi global industri dengan dokumen profil kebutuhan calon klien sertifikasi institusional (Moise & Norbert, 2025; Alsharef et al., 2023). Penelitian terdahulu umumnya berfokus pada domain monolingual dan homogen seperti pencocokan resume rekrutmen atau penilaian esai otomatis.

Dengan demikian, rumusan masalah penelitian ini adalah: (1) Bagaimana merancang sistem rekomendasi sertifikasi berbasis hybrid text similarity yang mampu menangani dokumen lintas bahasa? (2) Bagaimana mengimplementasikan integrasi TF-IDF dan Sentence Transformers dalam satu pipeline komputasi? (3) Bagaimana performa sistem yang dihasilkan berdasarkan validasi ground truth dari pakar?

METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan simulasi Ipteks dengan mengembangkan model komputasi hybrid text similarity. Sistem diimplementasikan menggunakan bahasa pemrograman Python 3.10 pada lingkungan Google Colaboratory. Penelitian dilaksanakan pada periode November 2025 hingga Mei 2026 di Laboratorium Data Jurusan Teknologi Informasi Politeknik Negeri Malang.

Data Penelitian

Data penelitian terdiri dari dua entitas utama. Pertama, korpus Objective Domain Certiport yang diperoleh dari laman resmi Certiport (certiport.com) dalam format PDF berbahasa Inggris, memuat deskripsi domain pengetahuan dan indikator kompetensi inti setiap skema sertifikasi. Kedua, dokumen profil calon klien sertifikasi berupa dokumen profil resmi institusi berbentuk PDF dalam Bahasa Indonesia, mencakup deskripsi kompetensi dan capaian bidang keahlian. Pengujian menggunakan 16 dokumen profil dari tiga institusi: Politeknik Negeri Malang, Universitas Gadjah Mada, dan Universitas Brawijaya.

Pipeline Sistem

Sistem CertiMatch dibangun melalui tujuh tahapan komputasi. (1) Ekstraksi teks dari dokumen PDF menggunakan pdfplumber. (2) Preprocessing teks mencakup tokenisasi, penghapusan stopword, dan lemmatisasi menggunakan NLTK. (3) Penerjemahan otomatis dokumen profil klien dari Bahasa Indonesia ke Bahasa Inggris menggunakan GoogleTranslator dengan algoritma sentence chunking untuk menangani teks panjang (batas 4.500 karakter per chunk). (4) Komputasi skor TF-IDF menggunakan TfidfVectorizer dari scikit-learn dan Cosine Similarity untuk menghasilkan skor numerik leksikal. (5) Pembentukan embedding semantik menggunakan model all-MiniLM-L6-v2 dari library Sentence Transformers dan Cosine Similarity untuk menghasilkan skor semantik. (6) Penggabungan skor hybrid menggunakan formula: Final Score = (α × Skor TF-IDF) + (β × Skor Semantik), dengan α = 0,40 dan β = 0,60. (7) Pemeringkatan dan output Top-N rekomendasi sertifikasi.

Pengumpulan Ground Truth

Ground truth dikumpulkan melalui instrumen kuesioner berbasis expert judgment dengan prinsip purposive sampling. Setiap program studi atau institusi diwakili oleh pemangku kepentingan yang memahami kebutuhan kompetensi, seperti Ketua Program Studi, dosen, atau praktisi industri. Instrumen berisi daftar lengkap skema sertifikasi Certiport dengan pilihan biner (memilih/tidak memilih) untuk setiap sertifikasi yang dianggap relevan.

Evaluasi Performa

Evaluasi dilakukan menggunakan Confusion Matrix dengan empat metrik: Precision, Recall, Accuracy, dan F1-Score. Pengujian mencakup lima skenario: (1) optimasi batas rekomendasi Top-N (Top-5 hingga Top-25), (2) optimasi rasio bobot hybrid (tujuh konfigurasi dari 0:100 hingga 100:0), (3) ablation study membandingkan TF-IDF Only, Semantic Only, dan Hybrid, (4) evaluasi kualitas terjemahan menggunakan metrik BLEU-1 hingga BLEU-4, dan (5) perbandingan lima model Sentence Transformer. User Acceptance Testing (UAT) dilakukan dengan melibatkan pengguna sistem untuk menilai fungsionalitas, kemudahan penggunaan, kinerja, dan kebermanfaatan.

HASIL DAN PEMBAHASAN

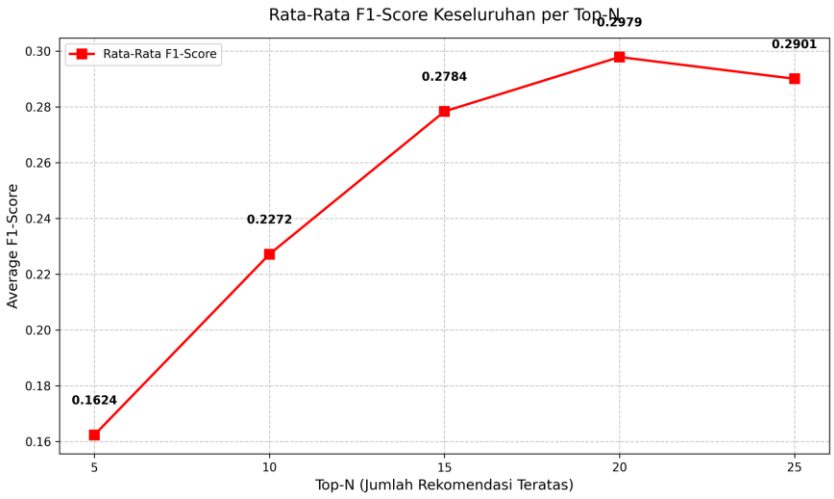
Bagian ini menyajikan hasil pengujian sistem CertiMatch secara menyeluruh, mencakup enam skenario evaluasi: optimasi parameter Top-N dan rasio bobot hybrid, analisis kinerja sistem terhadap 16 dokumen profil calon klien, ablation study, perbandingan model Sentence Transformer, evaluasi kualitas terjemahan, serta User Acceptance Testing. Seluruh pengujian menggunakan data ground truth yang dikumpulkan melalui expert judgment dari pemangku kepentingan di masing-masing institusi. Konfigurasi optimal dari skenario awal digunakan sebagai parameter tetap pada skenario-skenario berikutnya.

Optimasi Parameter Top-N dan Rasio Bobot

Hasil pengujian optimasi batas rekomendasi disajikan pada Tabel 1. Terdapat pola peningkatan F1-Score yang konsisten dari Top-5 (0,1624) hingga mencapai puncaknya pada Top-20 (0,2979), kemudian menurun pada Top-25 (0,2901). Penurunan pada Top-25 terjadi karena penambahan slot rekomendasi tidak lagi menemukan item relevan baru, melainkan memasukkan False Positive yang menekan nilai Precision. Berdasarkan hasil ini, Top-20 ditetapkan sebagai batas rekomendasi optimal.

Tabel 1. Rata-rata F1-Score per Batas Rekomendasi Top-N (* = Optimal)

Metrik	Top-5	Top-10	Top-15	Top-20*	Top-25
Rata-rata F1-Score	0,1624	0,2272	0,2784	0,2979	0,2901



Gambar 1. Kurva Rata-rata F1-Score per Batas Rekomendasi Top-N

Berdasarkan Tabel 1 dan Gambar 1, nilai F1-Score mengalami peningkatan secara bertahap dari Top-5 (0,1624) hingga mencapai nilai tertinggi pada Top-20 (0,2979), kemudian menurun pada Top-25 (0,2901). Penurunan pada Top-25 menunjukkan bahwa

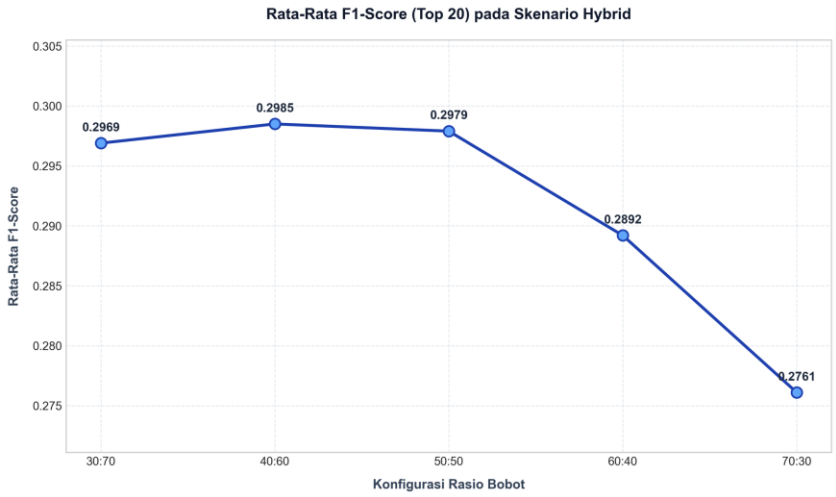
penambahan jumlah rekomendasi tidak lagi menghasilkan banyak item yang relevan, tetapi justru menambah *False Positive* sehingga nilai *Precision* menurun. Selain itu, kurva pada Gambar 1 juga menunjukkan bahwa Top-20 merupakan titik terbaik yang memberikan keseimbangan antara *Recall* dan *Precision*. Oleh karena itu, Top-20 dipilih sebagai batas rekomendasi yang optimal dan digunakan pada skenario pengujian berikutnya.

Kinerja Sistem pada 16 Program Studi

Tabel 3 menunjukkan rekapitulasi hasil evaluasi sistem CertiMatch menggunakan konfigurasi optimal Top-20 dengan rasio bobot 40:60 pada 16 program studi yang berasal dari tiga institusi. Evaluasi dilakukan menggunakan empat metrik yang berasal dari *Confusion Matrix*, yaitu *Precision*, *Recall*, *Accuracy*, dan *F1-Score*. Keempat metrik tersebut digunakan untuk menilai kualitas rekomendasi sistem dari berbagai aspek (Sathyanarayanan & Tantri, 2024).

Tabel 2. Rata-rata F1-Score per Konfigurasi Rasio Bobot (* = Optimal)

Metrik	**30:70**	**40:60***	**50:50**	**60:40**	**70:30**
Rata-rata F1-Score	0,2969	0,2990	0,2979	0,2890	0,2760



Gambar 2. Kurva Rata-rata F1-Score per Konfigurasi Rasio Bobot (Top-20)

Berdasarkan Tabel 2 dan Gambar 2, konfigurasi 40:60 (TF-IDF : Semantik) menghasilkan F1-Score tertinggi sebesar 0,2990, diikuti 50:50 (0,2979) dan 30:70 (0,2969). Kurva pada Gambar 2 menunjukkan pola puncak yang tegas di konfigurasi 40:60, dengan penurunan yang semakin curam ke arah dominasi TF-IDF (60:40 dan 70:30). Dominasi kontribusi semantik pada zona optimal mengkonfirmasi bahwa model Sentence Transformers lebih efektif menangkap kesamaan makna lintas kosakata dibandingkan pendekatan frekuensi kata murni (Pawestri & Suyanto, 2024). Penurunan signifikan pada

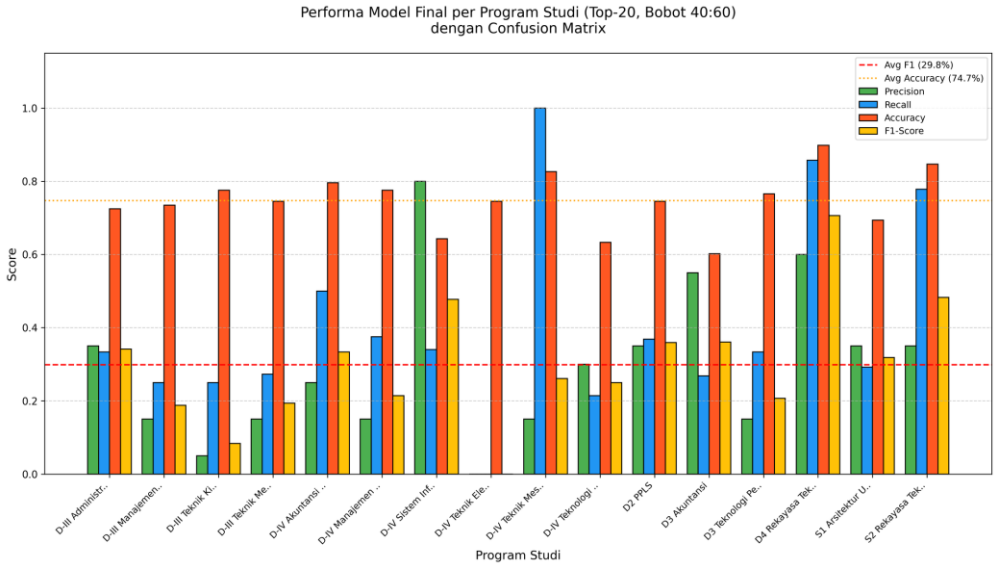
konfigurasi 60:40 (0,2890) dan 70:30 (0,2760) membuktikan bahwa peran TF-IDF tidak sebaiknya mendominasi karena keterbatasannya dalam menangani kesenjangan leksikal lintas bahasa. Konfigurasi 40:60 ditetapkan sebagai parameter optimal untuk seluruh skenario berikutnya.

Kinerja Sistem pada 16 Program Studi

Dengan konfigurasi akhir berupa Top-20 dan bobot 40:60, sistem CertiMatch diuji menggunakan 16 dokumen profil calon klien. Tabel 3 menunjukkan rekapitulasi hasil evaluasi pada setiap program studi, sedangkan Gambar 3 memperlihatkan distribusi metrik evaluasi secara keseluruhan beserta garis rata-ratanya. Evaluasi dilakukan menggunakan empat metrik yang berasal dari *Confusion Matrix*, yaitu *Precision*, *Recall*, *Accuracy*, dan *F1-Score*. Keempat metrik tersebut digunakan untuk menilai kualitas rekomendasi sistem dari berbagai aspek (Sathyanarayanan & Tantri, 2024).

Tabel 3. Rekapitulasi Metrik Evaluasi Sistem CertiMatch (Top-20, Bobot 40:60)

No	Program Studi	Precision	Recall	Accuracy	F1-Score
1	D-III Administrasi Bisnis	0,35	0,3333	0,7245	0,3158
2	D-III Manajemen Informatika	0,15	0,25	0,7347	0,1875
3	D-III Teknik Kimia	0,05	0,25	0,7755	0,0833
4	D-III Teknik Mesin	0,15	0,2727	0,7449	0,1935
5	D-IV Akuntansi Manajemen	0,25	0,50	0,7959	0,3333
6	D-IV Manajemen Rekayasa Konstruksi	0,15	0,375	0,7755	0,2143
7	D-IV Sistem Informasi Bisnis	0,80	0,3404	0,6429	0,4776
8	D-IV Teknik Elektronika	0,00	0,00	0,7449	0,0000
9	D-IV Teknik Mesin Produksi dan Perawatan	0,15	1,00	0,8265	0,2609
10	D-IV Teknologi Kimia Industri	0,30	0,2143	0,6327	0,2542
11	D2 PPLS	0,35	0,3684	0,7449	0,3585
12	D3 Akuntansi	0,55	0,2683	0,6020	0,3607
13	D3 Teknologi Pertambangan	0,15	0,3333	0,7653	0,2069
14	D4 Rekayasa Teknologi Internet UGM	0,60	0,8571	0,8980	0,7059
15	S1 Arsitektur UB	0,35	0,2917	0,6939	0,3191
16	S2 Rekayasa Teknologi Informasi	0,35	0,7778	0,8469	0,4810
	Rata-Rata	**29,37%**	**40,20%**	**74,68%**	**29,85%**



Gambar 3. Performa Sistem CertiMatch per Program Studi (Top-20, Bobot 40:60)

Variasi performa antar program studi yang cukup signifikan sebagaimana terlihat pada Gambar 3 mencerminkan ketergantungan sistem terhadap kedekatan domain keilmuan dengan spektrum sertifikasi Certiport. Program studi D4 Rekayasa Teknologi Internet UGM mencapai F1-Score tertinggi sebesar 0,7059, jauh melampaui garis rata-rata F1 sebesar 29,8%, diikuti S2 Rekayasa Teknologi Informasi (0,4810) dan D-IV Sistem Informasi Bisnis (0,4776). Ketiganya memiliki domain keilmuan yang dekat dengan terminologi teknis Objective Domain Certiport sehingga irisan kosakata antara profil calon klien dan dokumen sertifikasi lebih besar, baik secara leksikal maupun semantik. Sebaliknya, D-III Teknik Kimia, D3 Teknologi Pertambangan, dan D-IV Teknologi Kimia Industri menghasilkan performa terendah karena katalog sertifikasi Certiport secara dominan berorientasi pada teknologi informasi dan aplikasi perkantoran, sehingga sistem memiliki keterbatasan struktural dalam mencocokkan profil dari bidang keilmuan yang jauh dari domain tersebut.

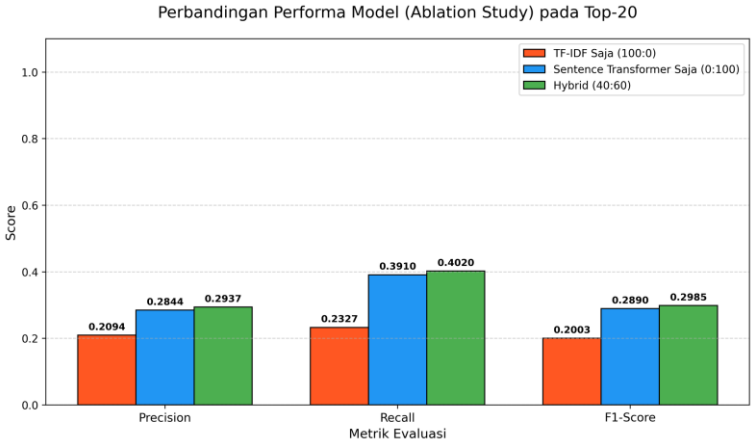
Ablation Study

Ablation study membandingkan tiga konfigurasi yang disajikan pada Tabel 4 dan Gambar 4: TF-IDF Only (100:0), Semantic Only (0:100), dan Hybrid terbaik (40:60). TF-IDF Only menghasilkan F1-Score terendah sebesar 0,2003, mengkonfirmasi keterbatasan pendekatan statistik berbasis frekuensi kata dalam menangkap kesamaan makna lintas kosakata dan lintas bahasa.

Tabel 4. Hasil Ablation Study (Top-20)

Konfigurasi Model	Precision	Recall	F1-Score
TF-IDF Only	0,2094	0,2327	0,2003

Konfigurasi Model	Precision	Recall	F1-Score
Sentence Transformer Only	0,2844	0,3910	0,2890
Hybrid (40:60)	**0,2937**	**0,4020**	**0,2990**



Gambar 4. Perbandingan Performa Ablation Study pada Top-20

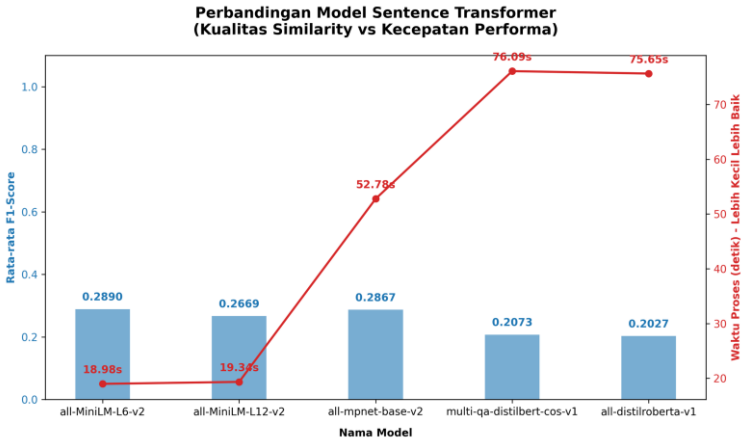
Berdasarkan Gambar 4, metode *Semantic Only* memperoleh F1-Score sebesar 0,2890, lebih tinggi dibandingkan *TF-IDF Only* pada seluruh metrik evaluasi. Hal ini menunjukkan bahwa model *all-MiniLM-L6-v2* mampu menangkap hubungan makna antar teks dengan lebih baik karena telah melalui proses *fine-tuning*. Sebaliknya, *sentence embedding* tanpa *fine-tuning* cenderung kurang optimal dalam merepresentasikan makna semantik (Li et al., 2020). Selain itu, metode *Hybrid* dengan bobot 40:60 menghasilkan F1-Score tertinggi sebesar 0,2990. Hasil ini menunjukkan bahwa penambahan TF-IDF tetap memberikan kontribusi dalam mendeteksi kesamaan istilah teknis secara lebih akurat. Peningkatan F1-Score dari *TF-IDF Only* ke *Hybrid* sebesar 48,9% juga menunjukkan bahwa penggabungan pendekatan semantik dan TF-IDF mampu meningkatkan kinerja sistem CertiMatch.

Perbandingan Model Sentence Transformer

Tabel 5 menunjukkan hasil perbandingan lima model *Sentence Transformer*, sedangkan Gambar 5 memperlihatkan perbandingan antara kualitas *similarity* yang diukur menggunakan F1-Score (batang biru) dan waktu proses komputasi (garis merah). Model *all-MiniLM-L6-v2* memperoleh F1-Score tertinggi sebesar 0,2890 dengan waktu proses paling cepat, yaitu 18,98 detik. Sementara itu, model *all-mpnet-base-v2* berada di posisi kedua dengan F1-Score sebesar 0,2867, tetapi membutuhkan waktu proses 52,78 detik atau sekitar 2,78 kali lebih lama. Meskipun demikian, selisih F1-Score antara kedua model tersebut hanya sebesar 0,0023.

Tabel 5. Hasil Perbandingan Lima Model Sentence Transformer

Model	Dim.	Waktu (s)	Precision	Recall	F1-Score
all-MiniLM-L6-v2	384	18,98	0,2844	0,3910	0,2890
all-MiniLM-L12-v2	384	19,34	0,2687	0,3582	0,2669
all-mpnet-base-v2	768	52,78	0,2812	0,3920	0,2867
multi-qa-distilbert-cos-v1	768	76,09	0,2031	0,2847	0,2073
all-distilroberta-v1	768	75,65	0,2094	0,2339	0,2027



Gambar 5. Perbandingan Kualitas dan Kecepatan Lima Model Sentence Transformer

Gambar 5 secara jelas memperlihatkan lonjakan waktu proses yang dramatis pada model ke-3 hingga ke-5, sementara F1-Score justru menurun. Temuan ini konsisten dengan studi Galli et al. (2024) yang membandingkan keempat model ST yang sama pada tugas semantic query, di mana all-MiniLM-L6-v2 menunjukkan kinerja kompetitif pada biaya komputasi jauh lebih rendah. Model all-MiniLM-L12-v2 (F1 = 0,2669) menunjukkan bahwa penambahan lapisan transformer dari 6 ke 12 tidak memberikan peningkatan berarti. Dua model terakhir, multi-qa-distilbert-cos-v1 (F1 = 0,2073) dan all-distilroberta-v1 (F1 = 0,2027), menghasilkan performa dan efisiensi terendah. Keunggulan all-MiniLM-L6-v2 juga dikonfirmasi oleh Yin & Zhang (2024) yang membuktikan ketangguhan model ini pada skenario lintas domain dengan variasi gaya bahasa tinggi, sehingga all-MiniLM-L6-v2 merupakan pilihan optimal untuk lingkungan komputasi tanpa GPU (Reimers & Gurevych, 2019).

Evaluasi Kualitas Terjemahan

Evaluasi kualitas terjemahan menggunakan metrik BLEU (Papineni et al., 2002) menghasilkan rata-rata BLEU-1 sebesar 82,37%, BLEU-2 sebesar 72,21%, BLEU-3 sebesar 64,04%, dan BLEU-4 sebesar 56,67%. Nilai BLEU-4 di atas 50% mengindikasikan bahwa struktur kalimat hasil terjemahan cukup memadai untuk pemrosesan teks lebih lanjut dalam pipeline CertiMatch. Terdapat variasi yang signifikan antar dokumen; S2 Rekayasa Teknologi Informasi menghasilkan BLEU-4 terendah sebesar

7,83% akibat kompleksitas kalimat akademik tingkat lanjut, sedangkan beberapa program studi teknik menghasilkan nilai sempurna (BLEU-4 = 1,000). Variasi ini mengindikasikan perlunya eksplorasi model terjemahan yang lebih robust pada penelitian lanjutan.

User Acceptance Testing

Tabel 6 menyajikan rekapitulasi hasil User Acceptance Testing yang melibatkan 5 responden dari tim operasional CATC JTI Polinema, mencakup empat aspek: fungsionalitas, kemudahan penggunaan, kinerja sistem, dan potensi kebermanfaatan.

Tabel 6. Rekapitulasi Hasil Pengujian User Acceptance Testing

Parameter Evaluasi	Pertanyaan	Skor Maks.	Skor Aktual	Persentase
Fungsionalitas	4	80	72	90,00%
Kemudahan Penggunaan	5	100	85	85,00%
Kinerja	1	20	18	90,00%
Potensi Kebermanfaatan	4	80	73	91,25%
Total Keseluruhan	**14**	**280**	**248**	**88,57%**

Tingkat kelayakan keseluruhan sebesar 88,57% masuk dalam kategori sangat layak. Aspek potensi kebermanfaatan memperoleh persentase tertinggi (91,25%), menunjukkan bahwa sistem dipersepsikan sangat berguna untuk mendukung keputusan profiling sertifikasi calon klien. Aspek kemudahan penggunaan menghasilkan nilai terendah (85,00%), mengindikasikan ruang perbaikan pada aspek antarmuka dan pengalaman pengguna pada pengembangan selanjutnya.

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem rekomendasi sertifikasi CertiMatch berbasis hybrid text similarity yang mengintegrasikan TF-IDF dan Sentence Transformers model all-MiniLM-L6-v2 dengan pipeline penerjemahan otomatis lintas bahasa. Konfigurasi optimal diperoleh pada batas rekomendasi Top-20 dengan rasio bobot 40:60 (TF-IDF : Semantik), menghasilkan rata-rata Precision 29,37%, Recall 40,20%, Accuracy 74,68%, dan F1-Score 29,85% terhadap 16 dokumen profil calon klien dari tiga institusi. Ablation study membuktikan kontribusi nyata dimensi semantik dengan peningkatan F1-Score sebesar 48,9% dibandingkan TF-IDF Only, sementara perbandingan model mengkonfirmasi all-MiniLM-L6-v2 sebagai pilihan paling efisien. User Acceptance Testing menghasilkan tingkat kelayakan 88,57% (sangat layak), membuktikan sistem dapat diterima sebagai alat bantu keputusan profiling kebutuhan sertifikasi calon klien secara otomatis dan objektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Negeri Malang atas dukungan fasilitas penelitian, kepada CATC JTI Polinema atas keterbukaan dalam proses wawancara dan pengumpulan data, serta kepada seluruh responden pakar dari Politeknik Negeri Malang, Universitas Gadjah Mada, dan Universitas Brawijaya yang telah bersedia mengisi instrumen ground truth.

DAFTAR PUSTAKA

ALSHAREF, A., GARG, S., & NASSOUR, H. (2023). EXPLORING THE EFFICIENCY OF TEXT SIMILARITY MEASURES IN AUTOMATED RESUME SCREENING FOR RECRUITMENT. PROCEEDINGS OF THE 10TH INTERNATIONAL CONFERENCE ON COMPUTING FOR SUSTAINABLE GLOBAL DEVELOPMENT (INDIACOM). IEEE. HTTPS://WWW.RESEARCHGATE.NET/PUBLICATION/378440387

CHRISMANTO, A. R., RAHARJO, W. S., & GILANG PURNAJATI, O. (2025). TEXT SIMILARITY ANALYSIS FOR EVALUATING ALIGNMENT BETWEEN LESSON PLANS AND TEACHING REPORTS. COGITO SMART JOURNAL, 11(2). HTTPS://DOI.ORG/10.31154/COGITO.V11I2.976.414-429

GALLI, C., DONOS, N., & CALCIOLARI, E. (2024). PERFORMANCE OF 4 PRE-TRAINED SENTENCE TRANSFORMER MODELS IN THE SEMANTIC QUERY OF A SYSTEMATIC REVIEW DATASET ON PERI-IMPLANTITIS. INFORMATION, 15(2), 68. HTTPS://DOI.ORG/10.3390/INFO15020068

LI, B., ZHOU, H., HE, J., WANG, M., YANG, Y., & LI, L. (2020). ON THE SENTENCE EMBEDDINGS FROM PRE-TRAINED LANGUAGE MODELS. PROCEEDINGS OF THE 2020 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), 9119–9130. HTTPS://DOI.ORG/10.18653/V1/2020.EMNLP-MAIN.733

MOISE, S. B., & NORBERT, M. S. (2025). ANALYSIS OF THE MATCH BETWEEN UNIVERSITY SKILLS AND LABOUR MARKET REQUIREMENTS USING SEMANTIC ANALYSIS. OPEN JOURNAL OF STATISTICS, 15(6), 492–512. HTTPS://DOI.ORG/10.4236/OJS.2025.156024

PAPINENI, K., ROUKOS, S., WARD, T., & ZHU, W.-J. (2002). BLEU: A METHOD FOR AUTOMATIC EVALUATION OF MACHINE TRANSLATION. PROCEEDINGS OF THE 40TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL), 311–318. HTTPS://DOI.ORG/10.3115/1073083.1073135

PAWESTRI, S., & SUYANTO, Y. (2024). ANALISIS PERBANDINGAN METODE SIMILARITY UNTUK KEMIRIPAN DOKUMEN BAHASA INDONESIA. JURNAL MEDIA INFORMATIKA BUDIDARMA, 8(3), 1440. HTTPS://DOI.ORG/10.30865/MIB.V8I3.7648

REIMERS, N., & GUREVYCH, I. (2019). SENTENCE-BERT: SENTENCE

EMBEDDINGS USING SIAMESE BERT-NETWORKS. PROCEEDINGS OF THE 2019 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), 3982–3992. HTTPS://ARXIV.ORG/ABS/1908.10084

SATHYANARAYANAN, S., & TANTRI, B. R. (2024). CONFUSION MATRIX-BASED PERFORMANCE EVALUATION METRICS. AFRICAN JOURNAL OF BIOMEDICAL RESEARCH, 27(4S), 4023–4031. HTTPS://DOI.ORG/10.53555/AJBR.V27I4S.4345

YIN, C., & ZHANG, Z. (2024). A STUDY OF SENTENCE SIMILARITY BASED ON THE ALL-MINILM-L6-V2 MODEL WITH 'SAME SEMANTICS, DIFFERENT STRUCTURE' AFTER FINE TUNING. PROCEEDINGS OF THE 2024 2ND INTERNATIONAL CONFERENCE ON IMAGE, ALGORITHMS AND ARTIFICIAL INTELLIGENCE (ICIAAI 2024), 677–684. HTTPS://DOI.ORG/10.2991/978-94-6463-540-9_69