

**PENGEMBANGAN SISTEM REKOMENDASI KURSUS
KREDENSIAL MIKRO BERBASIS KECERDASAN BUATAN
GENERATIF DENGAN ASESMEN PEMBELAJAR
OTOMATIS**



Disusun Oleh:

N a m a : Pandu Nur Afi Dewanto
NIM : 22523026

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2026

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PENGEMBANGAN SISTEM REKOMENDASI KURSUS
KREDENSIAL MIKRO BERBASIS KECERDASAN BUATAN
GENERATIF DENGAN ASESMEN PEMBELAJAR
OTOMATIS**

TUGAS AKHIR



Yogyakarta, 27 April 2026

Pembimbing,


(Dr. Nur Wijayaning R.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**PENGEMBANGAN SISTEM REKOMENDASI KURSUS
KREDENSIAL MIKRO BERBASIS KECERDASAN BUATAN
GENERATIF DENGAN ASESMEN PEMBELAJAR
OTOMATIS**

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 15 Juni 2026

Tim Penguji

Dr. Nur Wijayaning Rahayu, S.T., M.Sc.

Anggota 1

Dr. Hendrik, S.T., M.Eng.

Anggota 2

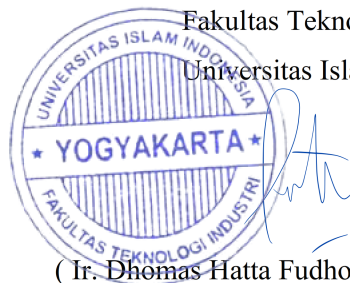
Taufiq Hidayat, S.T, M.C.S, Ph.D.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Pandu Nur Afi Dewanto

NIM : 22523026

Tugas akhir dengan judul:

PENGEMBANGAN SISTEM REKOMENDASI KURSUS KREDENSIAL MIKRO BERBASIS KECERDASAN BUATAN GENERATIF DENGAN ASESMEN PEMBELAJAR OTOMATIS

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 27 April 2026



(Pandu Nur Afi Dewanto)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, tugas akhir ini saya persembahkan sebagai bentuk terima kasih atas segala rahmat, kekuatan, dan kesempatan yang telah diberikan selama proses panjang dalam menempuh pendidikan ini. Setiap langkah dalam perjalanan ini tentu tidak selalu mudah, tetapi saya percaya bahwa setiap kesulitan, kegagalan, dan perjuangan merupakan bagian dari proses pembentukan diri.

Tugas akhir ini saya persembahkan untuk Ayah dan Bunda, yang selalu menjadi sumber doa, dukungan, dan kekuatan terbesar dalam hidup saya. Terima kasih atas kasih sayang, pengorbanan, nasihat, serta kepercayaan yang tidak pernah putus. Segala pencapaian ini tidak akan pernah lepas dari doa dan perjuangan Ayah dan Bunda yang selalu menyertai saya, bahkan dalam hal-hal yang mungkin tidak selalu saya ucapkan secara langsung.

Saya juga mempersembahkan karya ini untuk Yanguti dan Yangti, yang masih terus memberikan doa, kasih sayang, dan perhatian yang begitu berarti. Terima kasih atas kehangatan keluarga yang selalu membuat saya merasa memiliki tempat untuk pulang. Untuk Ninik dan Kakung yang telah berpulang, tugas akhir ini juga saya persembahkan dengan penuh cinta dan kerinduan. Nasihat, kebaikan, dan kenangan bersama kalian tetap hidup dalam ingatan saya, serta menjadi bagian dari kekuatan yang membantu saya terus melangkah hingga titik ini.

Untuk adik saya, Panji dan Praja, tugas akhir ini juga saya persembahkan dengan penuh kasih. Mungkin saya bukan seorang kakak yang sempurna, dan mungkin masih banyak hal yang belum bisa saya contohkan dengan baik. Namun, melalui perjalanan ini saya ingin menunjukkan bahwa proses yang sulit tetap dapat dilewati selama kita mau berusaha, belajar, dan tidak menyerah. Semoga pencapaian kecil ini dapat menjadi motivasi bagi kalian untuk terus tumbuh, berani bermimpi, dan percaya bahwa setiap perjuangan memiliki jalan dan waktunya masing-masing.

Persembahan ini juga saya tujukan kepada seluruh keluarga besar yang selalu memberikan dukungan, doa, dan semangat. Terima kasih karena telah menjadi bagian dari perjalanan saya, baik melalui bantuan, perhatian, maupun keyakinan bahwa saya mampu menyelesaikan pendidikan ini dengan baik.

Saya mengucapkan terima kasih kepada seluruh dosen Program Studi Informatika Universitas Islam Indonesia yang telah membagikan ilmu, bimbingan, dan pengalaman berharga selama masa perkuliahan. Secara khusus, saya mempersembahkan rasa terima kasih

yang mendalam kepada Ibu Nur selaku dosen pembimbing, yang telah membimbing saya dengan sabar, memberikan arahan, masukan, dan motivasi dalam proses penyusunan tugas akhir ini. Terima kasih juga kepada seluruh civitas akademik yang telah membantu, mendukung, dan menjadi bagian dari proses pendidikan saya selama menempuh studi.

Untuk orang terdekat saya, terima kasih karena telah menemani saya berproses dalam banyak keadaan, termasuk ketika berbagai rintangan datang silih berganti. Terima kasih karena tetap hadir, mendengarkan, menguatkan, dan percaya pada saya ketika saya sendiri terkadang merasa ragu. Kehadiranmu menjadi salah satu hal yang membuat perjalanan ini terasa lebih ringan dan berarti.

Untuk teman-teman seperjuangan saya, terutama teman-teman rantau dari Lombok, serta teman-teman yang saya temui dan kemudian menjadi sahabat selama merantau di Yogyakarta, terima kasih telah menjadi bagian penting dalam perjalanan ini. Kalian adalah salah satu anugerah yang membuat saya mampu bertahan di masa-masa sulit. Dari kebersamaan, tawa, bantuan, cerita, hingga perjuangan yang kita lalui bersama, saya belajar bahwa perjalanan pendidikan bukan hanya tentang menyelesaikan studi, tetapi juga tentang menemukan orang-orang baik yang membuat prosesnya menjadi lebih kuat, hangat, dan bermakna.

Terima kasih kepada ilmu pengetahuan yang telah membuka banyak jalan pemahaman, mengajarkan saya untuk berpikir lebih luas, dan membantu saya melihat bahwa belajar adalah proses yang tidak pernah benar-benar selesai. Melalui ilmu, saya belajar bahwa setiap pertanyaan, kesalahan, dan pencarian memiliki nilai yang dapat membentuk cara pandang serta arah perjalanan hidup.

Akhirnya, tugas akhir ini saya persembahkan untuk diri saya sendiri, sebagai pengingat bahwa setiap proses yang dijalani dengan sabar dan sungguh-sungguh akan membawa pelajaran berharga. Semoga ilmu, pengalaman, dan perjuangan yang tertuang dalam karya ini dapat menjadi langkah awal untuk terus bertumbuh, memberi manfaat, dan menghadapi perjalanan berikutnya dengan hati yang lebih kuat.

HALAMAN MOTO

Ilmu bukan hanya tentang mengetahui lebih banyak, tetapi tentang menggunakan pengetahuan untuk memberi manfaat yang lebih luas.

Setiap proses belajar membutuhkan keberanian untuk memulai, ketekunan untuk bertahan, dan kerendahan hati untuk terus memperbaiki diri.

Teknologi terbaik bukanlah yang paling rumit, melainkan yang mampu membantu manusia mengambil keputusan dengan lebih baik.

Kesalahan dalam proses bukan akhir dari perjalanan, melainkan bagian dari cara ilmu membentuk pemahaman.

Waktu adalah hal yang berharga dan tidak dapat dibeli menggunakan apa pun. Oleh karena itu, gunakan setiap kesempatan dengan sebaik-baiknya sebelum ia hanya menjadi bagian dari kenangan.

Yakinlah pada diri sendiri, karena keyakinan itulah yang akan selalu menemani di mana pun kamu berada.

Jangan takut pada kegagalan, tetapi takutlah ketika kamu sama sekali tidak berani untuk memulai.

Berusaha dengan sungguh-sungguh, belajar dengan hati yang terbuka, dan menyerahkan hasil terbaik kepada Allah Swt.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat, karunia, dan kekuatan yang telah diberikan, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Pengembangan Sistem Rekomendasi Kursus Kredensial Mikro Berbasis Kecerdasan Buatan Generatif dengan Asesmen Pembelajaran Otomatis”. Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Penyusunan tugas akhir ini tidak terlepas dari proses pembelajaran, tantangan, dan pengalaman berharga. Dalam prosesnya, penulis menghadapi kendala dalam memahami permasalahan penelitian, merancang sistem, melakukan implementasi, dan menyusun evaluasi. Namun, berkat bimbingan, dukungan, dan doa dari berbagai pihak, tugas akhir ini dapat diselesaikan.

Penulis menyampaikan terima kasih kepada Ibu Dr. Nur Wijyaning R., selaku dosen pembimbing, yang telah memberikan arahan, masukan, kesabaran, dan bimbingan selama penyusunan tugas akhir ini. Terima kasih juga penulis sampaikan kepada seluruh dosen Program Studi Informatika Universitas Islam Indonesia dan civitas akademik yang telah memberikan ilmu, pengalaman, serta dukungan selama masa studi.

Penulis juga berterima kasih kepada Ayah, Bunda, keluarga, orang-orang terdekat, dan teman-teman seperjuangan yang selalu memberikan doa, semangat, bantuan, dan kepercayaan selama proses perkuliahan hingga penyusunan tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini masih memiliki kekurangan dan keterbatasan. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun. Semoga tugas akhir ini dapat memberikan manfaat, khususnya dalam pengembangan sistem rekomendasi pembelajaran, asesmen otomatis, dan pemanfaatan kecerdasan buatan generatif dalam bidang pendidikan.

Yogyakarta, 27 April 2026



(Pandu Nur Afi Dewanto)

SARI

Pertumbuhan kredensial mikro pada pendidikan tinggi membuka peluang pembelajaran yang lebih fleksibel, terarah, dan berorientasi kompetensi. Namun, banyaknya pilihan kursus dan keterbatasan asesmen manual dapat menyulitkan pembelajar memperoleh rekomendasi yang sesuai dan tepat waktu. Penelitian ini bertujuan mengembangkan dan mengevaluasi awal MicroCred GenAI Backend sebagai layanan backend, yaitu bagian sistem yang menangani logika, pengolahan data, dan integrasi layanan, untuk mendukung asesmen kredensial mikro dan rekomendasi pembelajaran yang cepat, relevan, serta dapat ditelusuri.

Penelitian ini menggunakan pendekatan penelitian pengembangan sistem dengan fokus pada perancangan, implementasi, dan evaluasi awal layanan backend. Metode pengembangan perangkat lunak yang digunakan adalah Rapid Application Development (RAD), melalui tahapan pendefinisian kebutuhan, perancangan purwarupa, penyerapan umpan balik, dan finalisasi produk. Sistem dikembangkan sebagai minimum viable product (MVP) dengan arsitektur modular monolith. Proses asesmen menggabungkan penilaian berbasis aturan dan penilaian esai otomatis berbasis Generative Pre-trained Transformer (GPT). Rekomendasi pembelajaran dibangun menggunakan retrieval-augmented generation (RAG) dengan pemeringkatan hibrida ringan berbasis metadata, kata kunci, feature hashing, tag boost, dan skor kualitas kursus. Sistem juga menyimpan skor relevansi, alasan rekomendasi, dan jejak pengambilan untuk mendukung keterlacakan hasil.

Evaluasi dilakukan melalui observasi operasional, uji coba pengguna, audit manual, dan analisis *snapshot dump* lokal. Observasi terhadap 147 asesmen menunjukkan median waktu penyelesaian 13,21 detik, median latensi GPT 1696,5 ms, dan biaya per asesmen sekitar USD 0.00017–0.00020. Uji coba terhadap 10 partisipan menghasilkan rata-rata relevansi dan penerimaan 4,3 dari 5. Audit rekomendasi menghasilkan Precision@5 sebesar 0,98 dan NDCG@5 sebesar 0,9709, sedangkan audit 30 esai menghasilkan MAE 1,1998 dan korelasi Pearson 0,9990. *Snapshot dump* pada 152 asesmen menunjukkan seluruh pekerjaan GPT, RAG, dan fusion selesai 100%.

Hasil penelitian menunjukkan bahwa MicroCred GenAI Backend layak menjadi fondasi awal layanan asesmen dan rekomendasi kredensial mikro berbasis GenAI. Sistem mampu mempercepat asesmen, menghasilkan rekomendasi relevan, dan menyediakan dasar hasil yang dapat ditelusuri. Keterbatasan penelitian mencakup ukuran sampel, cakupan data kursus, personalisasi, dan belum adanya pengukuran dampak pembelajaran jangka panjang.

Kata kunci: asesmen otomatis, kredensial mikro, kecerdasan buatan generatif, rekomendasi pembelajaran, *retrieval-augmented generation*.

GLOSARIUM

AES (<i>Automated Essay Scoring</i>)	metode penilaian esai otomatis yang digunakan untuk memberi skor pada jawaban terbuka berdasarkan rubrik tertentu.
API (<i>Application Programming Interface</i>)	antarmuka pemrograman aplikasi yang memungkinkan komponen sistem saling berkomunikasi melalui aturan permintaan dan respons tertentu.
Asesmen	proses penilaian kemampuan, pemahaman, atau profil pembelajar yang digunakan sebagai dasar pemetaan kompetensi dan rekomendasi pembelajaran.
Async job (<i>asynchronous job</i>)	pekerjaan sistem yang dijalankan secara tidak langsung setelah permintaan utama selesai, sehingga pengguna tidak perlu menunggu seluruh proses berat selesai pada satu waktu.
Backend	bagian sistem yang menangani logika aplikasi, pengolahan data, autentikasi, penyimpanan, dan integrasi dengan layanan eksternal.
Baseline	tolak ukur awal yang digunakan sebagai pembanding untuk menilai kualitas atau kinerja sistem yang dikembangkan.
CBF (<i>Content-Based Filtering</i>)	pendekatan sistem rekomendasi yang menghasilkan rekomendasi berdasarkan kesamaan karakteristik konten, metadata, atau atribut objek dengan kebutuhan pengguna.
CF (<i>Collaborative Filtering</i>)	pendekatan sistem rekomendasi yang memanfaatkan pola interaksi historis pengguna untuk memperkirakan objek yang mungkin relevan bagi pengguna lain dengan pola serupa.
Cold-start	kondisi ketika sistem rekomendasi belum memiliki data interaksi pengguna yang cukup untuk menghasilkan rekomendasi berdasarkan riwayat penggunaan.
Course dataset	himpunan data kursus yang digunakan sebagai basis pengetahuan operasional dalam proses pengambilan dan penyusunan rekomendasi pembelajaran.

CSV (<i>Comma-Separated Values</i>)	format berkas tabular sederhana yang menyimpan data dalam bentuk baris dan kolom yang dipisahkan oleh tanda koma atau pemisah lain.
Degraded mode	kondisi terbatas ketika sistem tetap menghasilkan keluaran meskipun sebagian proses mengalami gangguan atau menggunakan mekanisme cadangan.
Endpoint	titik akses pada API yang digunakan untuk menerima permintaan tertentu, misalnya memulai asesmen, mengirim jawaban, memantau status, atau mengambil hasil.
Exponential backoff	strategi percobaan ulang yang memberi jeda semakin lama pada setiap percobaan berikutnya untuk mengurangi risiko kegagalan berulang.
FastAPI	kerangka kerja Python untuk membangun layanan API secara cepat dengan dukungan validasi data dan dokumentasi otomatis.
Feature hashing	teknik pembentukan representasi vektor ringan dengan menggunakan fungsi hash untuk mengubah token atau fitur menjadi indeks numerik.
Feedback	umpan balik dari pengguna terhadap hasil sistem, misalnya penilaian relevansi rekomendasi, penerimaan rekomendasi, atau komentar kualitatif.
Fusion	proses penggabungan hasil penilaian dan hasil rekomendasi menjadi ringkasan akhir yang dapat ditampilkan kepada pengguna.
GenAI (<i>Generative Artificial Intelligence</i>)	teknologi kecerdasan buatan generatif yang dapat menghasilkan teks, penilaian, atau keluaran baru berdasarkan masukan dan instruksi tertentu.
GPT (<i>Generative Pre-trained Transformer</i>)	model bahasa besar yang digunakan untuk memproses bahasa alami, termasuk menilai jawaban esai berdasarkan rubrik dalam penelitian ini.
HitRate@5	metrik evaluasi rekomendasi yang menunjukkan apakah setidaknya terdapat satu rekomendasi relevan pada lima hasil teratas.
Idempotency key	kunci unik yang digunakan untuk mencegah duplikasi pemrosesan ketika permintaan yang sama dikirim lebih dari satu kali.

JSON (<i>JavaScript Object Notation</i>)	format pertukaran data berbasis pasangan kunci dan nilai yang digunakan untuk menyusun keluaran atau metadata secara terstruktur.
JWT (<i>JSON Web Token</i>)	format token autentikasi yang digunakan untuk membuktikan identitas pengguna dan mengatur akses ke endpoint tertentu.
Kecerdasan buatan (<i>Artificial Intelligence/AI</i>)	teknologi yang memungkinkan sistem melakukan tugas yang biasanya membutuhkan kecerdasan manusia, seperti analisis, penilaian, atau rekomendasi.
Keterlacakan	kemampuan sistem untuk menunjukkan kembali dasar, sumber, skor, atau alasan yang mendukung suatu hasil penilaian atau rekomendasi.
Kredensial mikro (<i>micro- credential</i>)	bentuk pengakuan capaian belajar yang lebih kecil, terarah, dan berfokus pada keterampilan atau kompetensi tertentu.
LLM (<i>Large Language Model</i>)	model bahasa berskala besar yang mampu memahami dan menghasilkan teks berdasarkan pola yang dipelajari dari data pelatihan.
MAE (<i>Mean Absolute Error</i>)	metrik yang menghitung rata-rata selisih absolut antara skor sistem dan skor pembandingan, seperti skor penilai manusia.
Match reason	alasan kecocokan yang menjelaskan mengapa suatu kursus direkomendasikan kepada pengguna berdasarkan sinyal relevansi tertentu.
Metadata	informasi deskriptif tentang data atau objek, misalnya topik, tingkat kesulitan, tag, durasi, atau atribut kursus lain.
MicroCred GenAI Backend	layanan sisi peladen yang dikembangkan dalam penelitian untuk mendukung asesmen kredensial mikro dan rekomendasi pembelajaran berbasis GenAI.
Modular monolith	arsitektur sistem yang berjalan sebagai satu aplikasi terpadu, tetapi membagi tanggung jawab internal ke dalam modul-modul logis yang terpisah.

MOOC (<i>Massive Open Online Course</i>)	kursus daring terbuka berskala besar yang dapat diikuti banyak pembelajar melalui platform pembelajaran online.
MVP (<i>Minimum Viable Product</i>)	versi awal produk dengan fitur inti yang cukup untuk diuji, dievaluasi, dan dikembangkan melalui iterasi berikutnya.
NDCG@5 (<i>Normalized Discounted Cumulative Gain at 5</i>)	metrik evaluasi peringkat yang menilai kualitas urutan lima rekomendasi teratas dengan memberi bobot lebih besar pada posisi atas.
Pearson correlation	metrik statistik yang mengukur kekuatan hubungan linear antara dua kumpulan nilai, misalnya skor AES dan skor penilai manusia.
Precision@5	metrik evaluasi rekomendasi yang menghitung proporsi rekomendasi relevan dari lima rekomendasi teratas.
Profil kompetensi	gambaran kemampuan, kebutuhan, atau kesiapan belajar pengguna yang digunakan sebagai dasar penyusunan rekomendasi pembelajaran.
RAG (<i>Retrieval-Augmented Generation</i>)	pendekatan yang menggabungkan pengambilan informasi dari basis pengetahuan dengan proses penyusunan keluaran agar hasil tetap berlandaskan sumber yang tersedia.
Redis	penyimpanan data cepat berbasis memori yang dapat digunakan sebagai infrastruktur pendukung, misalnya untuk antrean atau pemrosesan latar.
Relevance score	nilai relevansi yang menunjukkan tingkat kecocokan suatu rekomendasi terhadap profil, kebutuhan, atau hasil asesmen pengguna.
Render Web Service	layanan hosting yang digunakan untuk menjalankan aplikasi backend pada lingkungan penerapan sistem.
Resend API	layanan pengiriman email yang digunakan untuk mengirim ringkasan hasil asesmen kepada pengguna.
Retrieval trace	jejak pengambilan yang mencatat dasar atau sumber rekomendasi agar hasil dapat diperiksa dan ditelusuri kembali.

Retry	percobaan ulang proses ketika terjadi kegagalan sementara, misalnya saat pemanggilan layanan eksternal belum berhasil.
RQ worker	pekerja latar berbasis Redis Queue yang dapat digunakan untuk menjalankan pekerjaan asinkron secara terpisah dari proses utama.
Rule-based scoring	metode penilaian berbasis aturan yang menghitung skor secara deterministik berdasarkan ketentuan yang telah ditetapkan atau biasa disebut pilihan ganda.
Scoring breakdown	rincian komponen skor yang menunjukkan bagaimana nilai akhir terbentuk dari beberapa bagian penilaian.
Skill mismatch	ketidaksesuaian antara kemampuan pengguna dengan tingkat kesulitan, topik, atau kebutuhan kompetensi dari kursus yang dipilih.
Tag boost	penambahan skor relevansi berdasarkan kecocokan tag, topik, atau label tertentu pada kursus dengan kebutuhan pengguna.
Turnaround time	waktu penyelesaian proses dari saat asesmen dikirim sampai hasil akhir tersedia bagi pengguna.
Umpan balik tepat waktu	respons atau penilaian yang diberikan cukup cepat sehingga pembelajar dapat segera memahami hasil dan menentukan langkah belajar berikutnya.
Validasi JSON	proses pemeriksaan apakah keluaran sistem atau model sudah sesuai dengan struktur JSON yang diharapkan.
Vektor token	representasi numerik ringan dari kumpulan token teks yang digunakan untuk membandingkan kedekatan antara kueri dan data kursus.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO.....	vii
KATA PENGANTAR	viii
SARI.....	ix
GLOSARIUM	xi
DAFTAR ISI	xvi
DAFTAR TABEL	xix
DAFTAR GAMBAR	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.1.1 Tren Kredensial Mikro dan Konteks Indonesia.....	2
1.1.2 Tantangan Pemilihan Jalur Belajar dan Keselarasan Kompetensi	4
1.1.3 Hambatan asesmen dan kebutuhan umpan balik tepat waktu	5
1.1.4 Peluang GenAI dan RAG untuk Asesmen serta Rekomendasi yang Dapat Ditelusuri	6
1.1.5 Ringkasan Arah Solusi Penelitian	8
1.1.6 Urgensi Penerapan Solusi pada Konteks Institusi	8
1.2 Rumusan Masalah dan Pertanyaan Penelitian.....	9
1.3 Lingkup Penelitian.....	10
1.4 Tujuan dan Manfaat Penelitian.....	10
1.4.1 Tujuan Penelitian.....	11
1.4.2 Manfaat Penelitian.....	11
1.5 Metodologi Penelitian Secara Umum.....	11
1.6 Struktur Laporan / Sistematika Penulisan	12
BAB II KAJIAN PUSTAKA.....	14
2.1 Kontekstualisasi Masalah dan Kriteria Solusi.....	14
2.2 Kategorisasi Teknik Sistem Rekomendasi	15
2.2.1 Pendekatan dengan <i>Collaborative Filtering</i> (CF)	17
2.2.2 Pendekatan dengan <i>Content-Based Filtering</i> (CBF)	18
2.2.3 Pendekatan dengan Rekomendasi Berbasis <i>Large Language Model</i> (LLM) Murni.....	19
2.2.4 Pendekatan dengan Generasi Berbantuan Pengambilan / <i>Retrieval-Augmented Generation</i> (RAG).....	20

2.2.5	Pendekatan dengan RAG Berbasis Pemeringkatan Hibrida Ringan	21
2.2.6	Simpulan Analisis Perbandingan.....	22
2.3	Pemilihan Solusi untuk Penelitian.....	23
2.4	Dasar Teori dan Konsep Solusi Terpilih	24
2.4.1	Konsep RAG untuk Rekomendasi Pembelajaran.....	24
2.4.2	Pemeringkatan Hibrida Ringan sebagai Mekanisme Pengambilan dan Pengurutan Awal	24
2.4.3	Penyelarasan Keterampilan–Kompetensi dalam Rekomendasi Pembelajaran.....	25
2.4.4	Keterjelasan dan Jejak Pengambilan	26
2.4.5	Penilaian Esai Otomatis Berbasis Rubrik dengan GPT	27
2.4.6	Risiko Halusinasi pada Penggunaan LLM	27
2.4.7	Metodologi <i>Rapid Application Development</i> (RAD)	28
2.5	Kesenjangan Penelitian dan Posisi Penelitian	28
BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM		30
3.1	Metodologi Penelitian	30
3.2	Diagram Penelitian	30
3.3	Tahapan Penelitian	31
3.3.1	Pendefinisian Kebutuhan / <i>Define Requirements</i>	31
3.3.2	Perancangan Purwarupa / <i>Prototype</i>	33
3.3.3	Penyerapan Umpan Balik / <i>Absorb Feedback</i>	50
3.4	Finalisasi Produk / <i>Finalize Product</i>	53
3.5	Ringkasan	54
BAB IV HASIL DAN PEMBAHASAN		55
4.1	Hasil Analisis Kebutuhan Awal Pengguna.....	55
4.1.1	Profil Responden	55
4.1.2	Persepsi dan Kendala Utama	56
4.1.3	Prioritas Faktor Pemilih Kursus	57
4.2	Hasil Perancangan Sistem	58
4.2.1	Gambaran Umum Sistem yang Dirancang	58
4.2.2	Hasil Perancangan Alur Asesmen	59
4.2.3	Hasil Perancangan Mekanisme RAG	60
4.2.4	Hasil Perancangan Basis Data	61
4.2.5	Hasil Perancangan Antarmuka Pengguna.....	62
4.2.6	Hasil Perancangan Transparansi dan Keterlacakan.....	63
4.2.7	Ringkasan Hasil Perancangan	63
4.3	Hasil Implementasi Sistem.....	64
4.3.1	Sistem yang Terbangun	64

4.3.2	Realisasi Kebutuhan Fungsional	65
4.3.3	Implementasi Antarmuka Pengguna.....	66
4.3.4	Implementasi Alur Asesmen dan Pemrosesan Asinkron.....	69
4.3.5	Implementasi Rekomendasi Berbasis RAG	70
4.3.6	Implementasi Penilaian Esai Berbasis GPT	71
4.3.7	Implementasi Penggabungan Hasil dan Keterlacakan.....	74
4.3.8	Implementasi Umpan Balik Pengguna	75
4.3.9	Ringkasan Hasil Implementasi	75
4.4	Hasil Pengujian dan Evaluasi Sistem	75
4.4.1	Hasil Observasi Operasional	76
4.4.2	Hasil Uji Coba Pengguna Terbatas.....	77
4.4.3	Hasil Audit Manual Kualitas Rekomendasi	79
4.4.4	Hasil Audit Penilaian Esai Otomatis	80
4.4.5	Evidensi Pendukung dari Snapshot Dump	82
4.4.6	Hasil Skenario Pengujian Sistem.....	83
4.4.7	Ringkasan Hasil Pengujian dan Evaluasi	85
4.5	Pembahasan.....	85
4.5.1	Responsivitas Alur Asesmen Asinkron	86
4.5.2	Kualitas Penilaian Esai Berbasis GPT	87
4.5.3	Relevansi Rekomendasi Berbasis RAG dengan Pemeringkatan Hibrida Ringan	88
4.5.4	Keterjelasan dan Keterlacakan Hasil.....	89
4.5.5	Implikasi Hasil Penelitian.....	90
4.5.6	Keterbatasan Interpretasi Temuan.....	91
4.6	Ringkasan	92
BAB V PENUTUP		94
5.1	Kesimpulan.....	94
5.2	Keterbatasan Penelitian	95
5.3	Saran Pengembangan Lanjutan	96
DAFTAR PUSTAKA.....		98

DAFTAR TABEL

Tabel 2. 1 Kriteria solusi sistem rekomendasi kredensial mikro berbasis GenAI	15
Tabel 2. 2 Kategorisasi teknik rekomendasi dan implikasinya.....	16
Tabel 3. 1 Struktur formulir umpan balik pemangku kepentingan	32
Tabel 3. 2 Kebutuhan fungsional sistem.....	32
Tabel 3. 3 Kebutuhan non fungsional sistem.....	33
Tabel 3. 4 Komponen utama desain RAG	39
Tabel 3. 5 Ringkasan entitas basis data utama.....	41
Tabel 3. 6 Rancangan komponen prompt penilaian esai	43
Tabel 3. 7 Parameter pemanggilan model GPT	45
Tabel 3. 8 Bobot rubrik berdasarkan tingkat kesulitan	46
Tabel 3. 9 Rancangan antarmuka pengguna MicroCred GenAI.....	48
Tabel 3. 10 Teknologi yang digunakan pada sistem.....	49
Tabel 3. 11 Indikator–parameter–sumber data evaluasi sistem.....	50
Tabel 3. 12 Tahapan uji coba pengguna	51
Tabel 3. 13 Skenario pengujian sistem	52
Tabel 3. 14 Metrik Evaluasi sistem.....	52
Tabel 4. 1 Profil 10 responden form kebutuhan awal	55
Tabel 4. 2 Ringkasan persepsi awal dan kendala utama responden	56
Tabel 4. 3 Prioritas faktor dalam memilih kursus atau sertifikasi	58
Tabel 4. 4 Realisasi kebutuhan fungsional pada implementasi	65
Tabel 4. 5 Hasil observasi operasional sistem	76
Tabel 4. 6 Kohor uji pengguna terbatas	77
Tabel 4. 7 Ringkasan indikator uji pengguna terbatas	78
Tabel 4. 8 Tema utama komentar pengguna.....	78
Tabel 4. 9 Hasil audit kualitas rekomendasi	79
Tabel 4. 10 Hasil audit keselarasan AES terhadap penilai manusia	81
Tabel 4. 11 Metrik pendukung dari snapshot dump lokal	82
Tabel 4. 12 Pemetaan skenario pengujian terhadap bukti.....	83
Tabel 4. 13 Hasil skenario representatif ujung-ke-ujung.....	84
Tabel 4. 14 Pemetaan hasil penelitian terhadap pertanyaan penelitian	86

DAFTAR GAMBAR

Gambar 1. 1 Ringkasan temuan survei pemberi kerja Indonesia terkait kredensial mikro, kredensial GenAI, dan dampaknya terhadap praktik rekrutmen	3
Gambar 1. 2 Visualisasi variasi tingkat penyelesaian pada massive open online course (MOOC) sebagai indikator tantangan keberlanjutan belajar pada pembelajaran daring terbuka	5
Gambar 1. 3 Ilustrasi kebutuhan umpan balik asesmen yang tepat waktu pada konteks pembelajaran daring	6
Gambar 1. 4 Ilustrasi konsep retrieval-augmented generation (RAG) dan jejak pengambilan untuk keterlacakan rekomendasi.....	8
Gambar 2. 1 Alur konseptual RAG dengan pemeringkatan hibrida ringan untuk rekomendasi pembelajaran.....	26
Gambar 3. 1 Tahapan Rapid Application Development (RAD) pada Pengembangan MicroCred GenAI	31
Gambar 3. 2 Arsitektur <i>modular monolith</i> MicroCred GenAI	35
Gambar 3. 3 Formula perhitungan skor akhir kandidat kursus pada proses RAG berbasis pemeringkatan hibrida ringan	36
Gambar 3. 4 Diagram alur pemrosesan asesmen MicroCred GenAI.....	37
Gambar 3. 5 Desain RAG untuk rekomendasi kursus kredensial mikro pada MicroCred GenAI Backend	38
Gambar 3. 6 ERD sistem MicroCred GenAI Backend	41
Gambar 3. 7 Diagram aliran data / data flow diagram (DFD)	42
Gambar 3. 8 Wireframe rancangan antarmuka pengguna MicroCred GenAI	47
Gambar 4. 1 Grafik distribusi kendala utama responden pada proses pencarian kursus dan sertifikasi daring	57
Gambar 4. 2 Halaman mulai asesmen.....	66
Gambar 4. 3 Halaman pengerjaan soal asesmen.....	67
Gambar 4. 4 Halaman hasil asesmen dengan rincian penilaian / <i>scoring breakdown</i>	68
Gambar 4. 5 Tampilan daftar rekomendasi kursus dengan nilai kecocokan / <i>match score</i>	69
Gambar 4. 6 Cuplikan kode pemicu alur pemrosesan asinkron setelah pengguna mengirim asesmen.....	70

Gambar 4. 7 Cuplikan kode formula perhitungan skor akhir kandidat kursus pada layanan RAG dengan pemeringkatan hibrida ringan	71
Gambar 4. 8 Cuplikan kode pembentukan prompt dan pemanggilan GPT untuk penilaian esai	72
Gambar 4. 9 Cuplikan kode validasi dan perhitungan skor akhir esai berbasis rubrik di sisi backend	73
Gambar 4. 10 Perbandingan kualitas rekomendasi antara baseline dan MicroCred GenAI....	80
Gambar 4. 11 Perbandingan rerata skor AES dan rerata skor penilai manusia	81

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kredensial mikro atau *micro-credential* merupakan bentuk pengakuan capaian belajar dalam ruang lingkup yang lebih kecil, spesifik, dan terarah dibandingkan kualifikasi pendidikan tradisional. Kredensial ini menunjukkan bahwa seseorang telah menguasai pengetahuan, keterampilan, atau kompetensi tertentu setelah mengikuti proses pembelajaran dan memenuhi standar asesmen yang ditetapkan. Secara sederhana, kredensial mikro dapat dipahami sebagai bukti atas penguasaan kemampuan tertentu pada bidang yang spesifik. Misalnya, seorang mahasiswa yang mengikuti kursus singkat tentang dasar pemrograman Python, menyelesaikan asesmen, dan memenuhi standar kelulusan tertentu dapat memperoleh kredensial mikro yang menunjukkan bahwa ia telah menguasai kompetensi dasar Python.

Dalam konteks pendidikan tinggi, kredensial mikro menjadi relevan karena perguruan tinggi menghadapi tekanan yang semakin besar untuk menyediakan pembelajaran yang lebih fleksibel, terarah, dan berorientasi kompetensi di tengah perubahan kebutuhan industri yang berlangsung cepat. Dengan karakteristiknya yang ringkas dan spesifik, kredensial mikro tidak menggantikan gelar akademik, tetapi melengkapi pembelajaran formal melalui pengakuan atas capaian belajar tertentu. Karakter tersebut membuat kredensial mikro berpotensi mendukung pembelajaran sepanjang hayat, peningkatan kesiapan kerja, dan modularisasi pembelajaran pada pendidikan tinggi. Namun, pertumbuhannya juga menuntut institusi pendidikan untuk menata kualitas, asesmen, pengakuan, dan dukungan pengambilan keputusan pembelajar secara lebih sistematis (Kato, 2023; Stein, 2025).

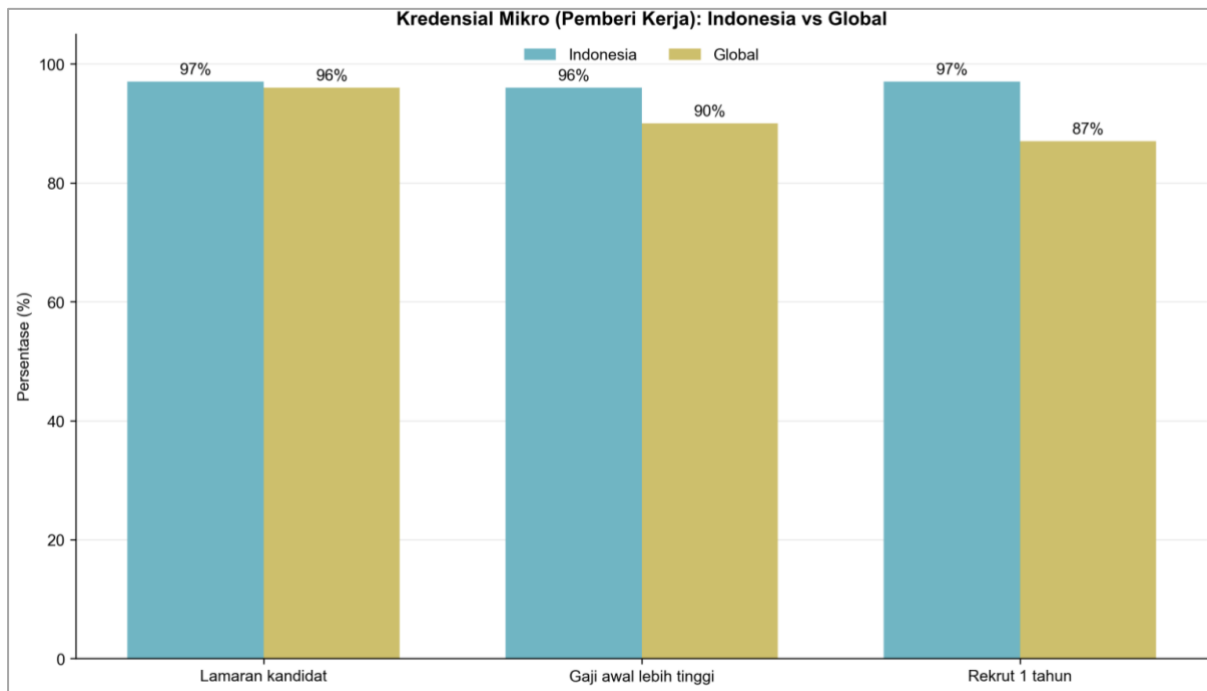
Implementasi kredensial mikro di perguruan tinggi masih menghadapi berbagai tantangan, baik dari sisi pembelajar, pengembang, organisasi, industri, maupun teknologi. Hal ini menunjukkan bahwa persoalan kredensial mikro tidak hanya berkaitan dengan bertambahnya jumlah penawaran belajar, tetapi juga dengan bagaimana capaian pembelajaran dapat dinilai secara tepat dan diarahkan menjadi rekomendasi belajar yang relevan. Pada pembelajaran daring berskala besar, kebutuhan tersebut menjadi semakin penting karena proses asesmen manual sulit dipertahankan ketika jumlah pembelajar meningkat dan umpan balik perlu diberikan secara cepat. Situasi ini membuka ruang bagi pemanfaatan kecerdasan buatan generatif atau *generative artificial intelligence* (GenAI) untuk membantu asesmen dan

personalisasi jalur belajar pada ekosistem kredensial mikro ((Halim, 2024; Harizan, 2025; Holt dkk., 2024).

1.1.1 Tren Kredensial Mikro dan Konteks Indonesia

KATO mencatat bahwa jumlah penawaran kredensial mikro pada platform pembelajaran besar meningkat dari sekitar 600 pada 2018 menjadi sekitar 1.900 pada 2022. Pertumbuhan tersebut berkaitan dengan kebutuhan peningkatan dan pembaruan keterampilan tenaga kerja, sekaligus dorongan agar sistem pendidikan menjadi lebih fleksibel dan berpusat pada pembelajar. Pada sisi pasar kerja, kredensial mikro juga semakin dipandang sebagai sinyal kesiapan kerja yang spesifik. Secara global, 96% pemberi kerja menyatakan bahwa kredensial mikro memperkuat kualitas lamaran kandidat, 87% pernah merekrut kandidat yang memiliki kredensial mikro dalam satu tahun terakhir, dan sekitar sembilan dari sepuluh pemberi kerja bersedia menawarkan gaji awal yang lebih tinggi kepada kandidat dengan kredensial mikro (Kato, 2023; Stein, 2025).

Pada konteks Indonesia, peluang tersebut terlihat dari tingginya penerimaan pemberi kerja terhadap kredensial yang relevan. Sebanyak 97% pemberi kerja menyatakan lebih cenderung merekrut kandidat yang memiliki kredensial GenAI, 91% lebih cenderung merekrut kandidat dengan kredensial yang dapat dikonversi menjadi kredit akademik, dan 93% melaporkan adanya penghematan biaya pelatihan untuk perekrutan tingkat awal yang memiliki kredensial mikro relevan. Ringkasan temuan survei pemberi kerja Indonesia tersebut ditampilkan pada Gambar 1.1.



Gambar 1. 1 Ringkasan temuan survei pemberi kerja Indonesia terkait kredensial mikro, kredensial GenAI, dan dampaknya terhadap praktik rekrutmen

Sumber: Marni Baker Stein, 2025

Berdasarkan Gambar 1.1, terlihat bahwa kredensial yang relevan memiliki penerimaan yang tinggi dari sisi pemberi kerja di Indonesia. Persentase yang tinggi pada aspek kecenderungan merekrut kandidat dengan kredensial GenAI, penerimaan terhadap kredensial yang dapat dikonversi menjadi kredit akademik, dan penghematan biaya pelatihan menunjukkan bahwa kredensial mikro tidak lagi dipandang hanya sebagai pelengkap pembelajaran. Kredensial mikro mulai berfungsi sebagai penanda kompetensi yang dapat dibaca langsung oleh dunia kerja.

Di sisi lain, tingginya minat terhadap kredensial mikro tidak otomatis membuat implementasinya sederhana. Perguruan tinggi tetap menghadapi persoalan pengembangan, pengakuan, infrastruktur, kualitas pelaksanaan, serta jaminan mutu. Jaminan mutu yang kuat diperlukan untuk membangun kepercayaan antar-institusi dan mendukung pengakuan akademik kredensial mikro secara lebih luas. Oleh karena itu, pertumbuhan kredensial mikro perlu diikuti oleh mekanisme asesmen, dokumentasi hasil, dan pengelolaan mutu yang memadai. Dalam konteks Indonesia, kondisi ini menunjukkan kombinasi antara peluang yang tinggi dari sisi permintaan pasar kerja dan kebutuhan yang kuat terhadap mekanisme institusional yang lebih matang pada sisi penyelenggaraan (Halim, 2024; Kato, 2023).

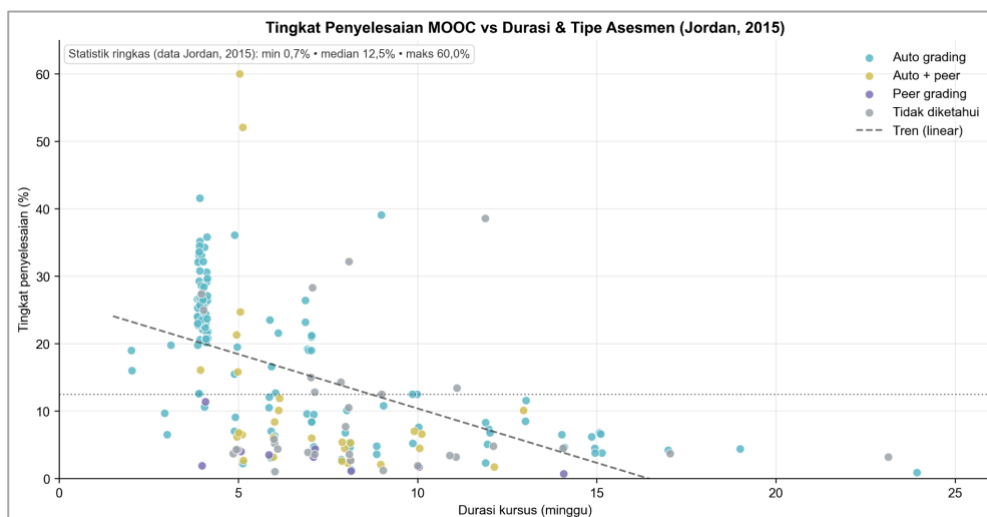
1.1.2 Tantangan Pemilihan Jalur Belajar dan Keselarasan Kompetensi

Semakin banyaknya penawaran kredensial mikro tidak otomatis memudahkan pembelajar dalam memilih jalur belajar yang paling tepat. Pembelajar tetap memerlukan informasi yang jelas, mekanisme pencarian yang terstruktur, serta dukungan pengambilan keputusan agar dapat membandingkan berbagai penawaran secara lebih terarah. Nilai kredensial jangka pendek juga bergantung pada kecocokan antara program, profil pembelajar, dan konteks dampak yang ingin dicapai. Dengan demikian, pilihan pembelajaran tidak dapat dipisahkan dari kebutuhan individual pembelajar, sehingga platform pencarian kredensial mikro perlu dilengkapi dengan bimbingan yang lebih personal (Kato, 2023).

Bertambahnya jumlah pilihan kursus justru dapat menambah beban pengambilan keputusan apabila pembelajar tidak memiliki dukungan yang memadai untuk menilai relevansi setiap penawaran. Pada konteks kursus daring terbuka masif atau *massive open online course* (MOOC), tingkat penyelesaian kursus dilaporkan bervariasi antara 0,7% hingga 52,1% dengan nilai median 12,6%. Tingkat penyelesaian tersebut dipengaruhi oleh beberapa faktor, seperti panjang kursus, waktu peluncuran, dan jenis asesmen yang digunakan. Selain itu, minggu pertama dan kedua menjadi fase penting yang menentukan keterlibatan pembelajar pada kursus daring berskala besar. Temuan ini menunjukkan bahwa keberlanjutan belajar tidak hanya bergantung pada ketersediaan materi, tetapi juga pada bagaimana jalur belajar, tingkat kesulitan, dan dukungan belajar disusun sejak awal (Jordan, 2015).

Pada saat yang sama, integrasi kecerdasan buatan atau *artificial intelligence* (AI) pada kredensial mikro dalam konteks pembelajaran jarak jauh berpotensi menyediakan jalur belajar yang dipersonalisasi, umpan balik waktu nyata, dan jalur adaptif sesuai kebutuhan pembelajar. Dengan kata lain, tantangan utama tidak berhenti pada banyaknya pilihan pembelajaran, tetapi juga pada kemampuan sistem untuk mencocokkan kebutuhan pembelajar dengan sumber belajar yang paling relevan (Harizan, 2025)

Dalam penelitian ini, keselarasan keterampilan-kompetensi didefinisikan sebagai tingkat kesesuaian rekomendasi pembelajaran dengan profil kompetensi pembelajar. Definisi ini digunakan untuk menekankan bahwa rekomendasi pembelajaran tidak seharusnya bersifat generik, tetapi perlu berangkat dari hasil asesmen dan profil kompetensi yang dapat dipertanggungjawabkan.



Gambar 1. 2 Visualisasi variasi tingkat penyelesaian pada massive open online course (MOOC) sebagai indikator tantangan keberlanjutan belajar pada pembelajaran daring terbuka

Sumber: Jordan, 2015

1.1.3 Hambatan asesmen dan kebutuhan umpan balik tepat waktu

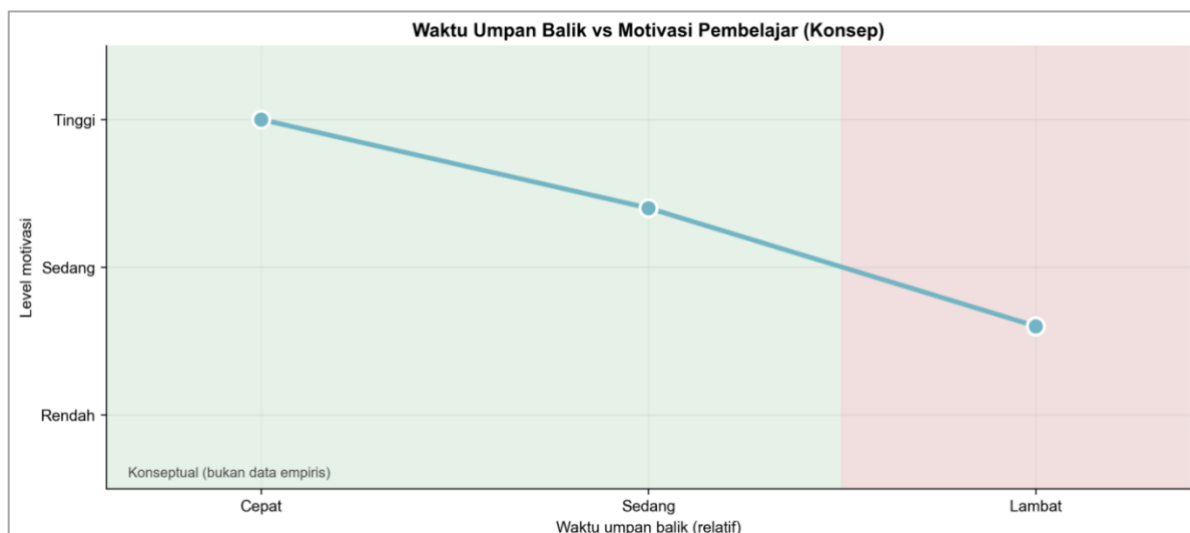
Asesmen merupakan komponen penting dalam proses belajar karena membantu pembelajar memahami kualitas kinerja, mengenali kekuatan, serta mengidentifikasi area yang perlu diperbaiki. Dalam konteks ini, umpan balik yang baik tidak hanya berfungsi sebagai penjelasan atas nilai akhir, tetapi juga perlu bersifat spesifik, konsisten, dan berorientasi pada pengembangan kemampuan pembelajar. Namun, ekspektasi terhadap umpan balik yang berkualitas sering kali berbenturan dengan keterbatasan waktu dan sumber daya institusi, terutama ketika jumlah pembelajar meningkat. Pada skala yang lebih besar, kualitas umpan balik menjadi semakin sulit dijaga apabila proses asesmen masih sepenuhnya bergantung pada penilaian manual (Holt dkk., 2024)

Pada konteks kredensial mikro, hasil asesmen tidak hanya dipakai untuk memberi nilai, tetapi juga untuk memetakan kompetensi dan menjadi dasar rekomendasi pembelajaran lanjutan. Akibatnya, keterlambatan atau rendahnya spesifisitas umpan balik bukan hanya memengaruhi pengalaman belajar, tetapi juga berpotensi menurunkan ketepatan keputusan belajar berikutnya.

Shanmughan dkk. (2024) menekankan bahwa automasi asesmen dan umpan balik berbasis AI berpotensi mengurangi beban administratif dan memberikan umpan balik yang lebih cepat kepada pembelajar.

Dengan demikian, hambatan asesmen pada ekosistem kredensial mikro dapat dipahami sebagai hambatan ganda, yaitu hambatan dalam memberi penilaian dan hambatan dalam menghasilkan dasar rekomendasi belajar yang cukup cepat serta cukup relevan.

Ilustrasi hubungan antara kecepatan umpan balik dan motivasi pembelajar ditunjukkan pada Gambar 1.3.



Gambar 1. 3 Ilustrasi kebutuhan umpan balik asesmen yang tepat waktu pada konteks pembelajaran daring

Sumber: Holt dkk., 2024

1.1.4 Peluang GenAI dan RAG untuk Asesmen serta Rekomendasi yang Dapat Ditelusuri

Kajian sistematis Harizan (2025) menunjukkan bahwa integrasi kecerdasan buatan pada kredensial mikro dan pembelajaran jarak jauh berpotensi meningkatkan personalisasi, automasi dukungan belajar, dan aksesibilitas layanan pembelajaran. Dalam penelitiannya juga menegaskan bahwa sistem berbasis AI dapat menyediakan jalur belajar adaptif dan umpan balik waktu nyata yang lebih sesuai dengan kebutuhan pembelajar.

Penelitian sebelumnya Shanmughan dkk. (2024) menunjukkan bahwa AI dapat menganalisis performa pembelajar, menyesuaikan tingkat kesulitan tugas, menyarankan sumber belajar tambahan, dan mengotomatisasi asesmen serta umpan balik sesuai kebutuhan pembelajar. Potensi tersebut menjadikan GenAI relevan untuk membantu dua persoalan yang telah diuraikan sebelumnya, yaitu hambatan asesmen dan rendahnya keselarasan jalur belajar.

Meskipun demikian, integrasi AI pada kredensial mikro juga tidak lepas dari persoalan kualitas, pengakuan, privasi data, aspek etis, dan kesenjangan. Bahwa kualitas jaminan mutu, pengakuan, serta kerangka standar yang kuat menjadi penting untuk memastikan penerapan AI pada kredensial mikro dapat diterima secara luas (Harizan, 2025).

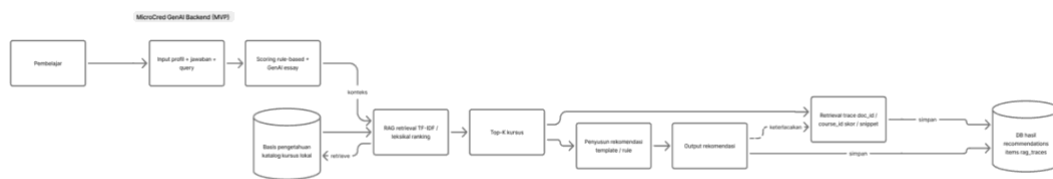
Selain itu, penelitian yang dilakukan oleh Shanmughan dkk. (2024) juga menunjukkan bahwa privasi data, bias algoritmik, dan kesenjangan digital merupakan tantangan yang harus ditangani agar pemanfaatan AI tidak menimbulkan ketidakadilan baru dalam pendidikan. Kebutuhan tersebut sejalan dengan pandangan Kato (2023) bahwa kepercayaan dan pengakuan terhadap kredensial mikro sangat bergantung pada kualitas jaminan mutu, kejelasan deskriptor, dan keterhubungan dengan kebijakan pengakuan akademik yang lebih luas. Karena itu, solusi berbasis AI pada penelitian ini tidak cukup hanya cepat dan personal, tetapi juga perlu dirancang agar dasar hasilnya dapat diperiksa dan ditelusuri.

Pada penelitian ini, GenAI dimanfaatkan terutama untuk *automated essay scoring* (AES) sebagai bagian dari proses asesmen. Untuk penyusunan rekomendasi pembelajaran, penelitian ini menggunakan *retrieval-augmented generation* (RAG), yaitu pendekatan yang menggabungkan pengambilan dokumen dari basis pengetahuan dengan proses generasi keluaran agar rekomendasi tetap berlandaskan sumber yang digunakan (Lewis dkk., 2020).

Pada tahap *minimum viable product* (MVP), pengambilan dan pemeringkatan kandidat rekomendasi dilakukan dari himpunan *course dataset* menggunakan pemeringkatan hibrida ringan. Mekanisme ini menggabungkan pencocokan metadata yang diperkaya, yaitu pencocokan berbasis data deskriptif kursus, relevansi kata kunci, kemiripan vektor token berbasis *feature hashing*, *tag boost*, dan skor kualitas kursus. Dengan pendekatan tersebut, rekomendasi tetap terikat pada himpunan data kursus yang tersedia, tetapi masih dapat mempertimbangkan beberapa sinyal relevansi yang berkaitan dengan profil kompetensi pembelajar (Weinberger dkk., 2009).

Melalui *retrieval trace*, sistem dapat menunjukkan dasar dokumen yang mendukung rekomendasi yang dihasilkan dan sekaligus mendukung keterjelasan alasan rekomendasi. Dengan pendekatan tersebut, rekomendasi pembelajaran tidak hanya diarahkan agar relevan terhadap profil kompetensi, tetapi juga memiliki dasar sumber yang lebih mudah dipahami oleh pengguna.

Ilustrasi konsep RAG dan retrieval trace sebagai dasar keterlacakan rekomendasi ditunjukkan pada Gambar 1.4.



Gambar 1. 4 Ilustrasi konsep retrieval-augmented generation (RAG) dan jejak pengambilan untuk keterlacakan rekomendasi

Sumber: Lewis dkk., 2020

1.1.5 Ringkasan Arah Solusi Penelitian

Berdasarkan latar belakang tersebut, penelitian ini mengembangkan MicroCred GenAI Backend sebagai layanan inti pada sisi *backend* untuk mendukung asesmen kredensial mikro dan rekomendasi pembelajaran berbasis GenAI. Pada tahap MVP, sistem dibangun dalam bentuk *modular monolith* agar batas modul domain tetap jelas, tetapi pengembangan awal tetap sederhana dan terkelola. Sistem menggabungkan penilaian berbasis aturan untuk pertanyaan teori dan profil dengan penilaian esai berbasis model *Generative Pre-trained Transformer* (GPT) yang diproses secara asinkron agar layanan tetap responsif.

Untuk rekomendasi pembelajaran, sistem memanfaatkan RAG berbasis himpunan data kursus dan menyimpan jejak pengambilan sebagai dasar keterlacakan hasil. Melalui pendekatan tersebut, penelitian ini berupaya menjembatani kebutuhan asesmen yang lebih cepat dengan kebutuhan rekomendasi pembelajaran yang lebih relevan dan lebih mudah diperiksa.

1.1.6 Urgensi Penerapan Solusi pada Konteks Institusi

Pertumbuhan kredensial mikro, tingginya perhatian pemberi kerja terhadap kredensial yang relevan, dan tantangan implementasinya di perguruan tinggi menunjukkan bahwa institusi membutuhkan mekanisme asesmen dan rekomendasi yang lebih adaptif serta lebih terukur (Stein, 2025).

Di perguruan tinggi, implementasi kredensial mikro juga menghadapi tantangan tata kelola dan pelaksanaan yang menuntut mekanisme asesmen serta rekomendasi yang lebih adaptif dan lebih terukur (Halim, 2024). Di saat yang sama, pembelajar memerlukan dukungan yang memadai untuk menavigasi banyaknya pilihan pembelajaran yang tersedia (Kato, 2023).

Pembelajar juga memerlukan umpan balik yang cukup cepat agar dapat melanjutkan keputusan belajar berikutnya secara lebih tepat (Holt dkk., 2024). Tanpa dukungan asesmen yang cepat dan rekomendasi yang relevan, pertumbuhan kredensial mikro justru berisiko menambah beban pengambilan keputusan bagi pembelajar daripada membantu mereka. Oleh karena itu, penelitian ini dipandang perlu untuk menyediakan dan mengevaluasi layanan asesmen serta rekomendasi pembelajaran yang lebih cepat, lebih relevan, dan tetap dapat ditelusuri pada konteks pendidikan tinggi di Indonesia.

1.2 Rumusan Masalah dan Pertanyaan Penelitian

Bagian ini merumuskan permasalahan utama yang menjadi fokus penelitian serta pertanyaan penelitian yang dijawab melalui pengembangan dan evaluasi awal sistem.

Berdasarkan latar belakang penelitian yang telah diuraikan, penelitian ini merumuskan tiga permasalahan utama sebagai berikut.

1. Proses asesmen dan penyusunan rekomendasi pada konteks kredensial mikro sulit diskalakan apabila bertumpu pada penilaian manual, sehingga umpan balik berisiko lambat ketika jumlah pembelajar meningkat.
2. Pemilihan jalur belajar yang tidak memperhitungkan profil kompetensi pembelajar berpotensi menurunkan relevansi rekomendasi dan mengganggu keberlanjutan belajar.
3. Sistem rekomendasi berbasis kecerdasan buatan memerlukan mekanisme keterjelasan dan keterlacakan agar hasilnya dapat dipahami, diperiksa, dan diverifikasi.

Berdasarkan rumusan masalah tersebut, pertanyaan penelitian utama dalam penelitian ini dirumuskan sebagai berikut.

Bagaimana mengembangkan MicroCred GenAI sebagai layanan sisi peladen / *backend* untuk mendukung asesmen kredensial mikro berbasis kecerdasan buatan generatif / *generative artificial intelligence* (GenAI) dan rekomendasi pembelajaran yang cepat, relevan, serta dapat ditelusuri pada konteks pendidikan tinggi di Indonesia?

Agar pertanyaan utama tersebut dapat dijawab secara lebih terarah, penelitian ini menggunakan tiga pertanyaan operasional sebagai berikut.

1. Bagaimana rancangan dan implementasi alur asesmen asinkron pada MicroCred GenAI Backend dapat mendukung pemrosesan asesmen yang lebih cepat pada tahap *minimum viable product* (MVP)?

2. Bagaimana mekanisme rekomendasi berbasis *retrieval-augmented generation* (RAG) dengan pemeringkatan hibrida ringan dapat menghasilkan rekomendasi pembelajaran yang relevan terhadap profil kompetensi pembelajar?
3. Bagaimana mekanisme skor, alasan rekomendasi, dan jejak pengambilan dapat mendukung keterjelasan serta keterlacakan hasil asesmen dan rekomendasi?

Pertanyaan-pertanyaan tersebut dijawab melalui analisis rancangan sistem, keluaran fungsional, observasi operasional, uji coba pengguna, dan audit manual kualitas hasil.

1.3 Lingkup Penelitian

Penelitian ini difokuskan pada pengembangan dan evaluasi awal MicroCred GenAI Backend dengan ruang lingkup sebagai berikut.

1. Penelitian berfokus pada pengembangan komponen layanan inti sisi peladen / *backend* untuk proses asesmen dan rekomendasi pembelajaran pada konteks kredensial mikro di pendidikan tinggi.
2. Proses asesmen dibatasi pada pengolahan jawaban teori, profil, dan esai sebagai dasar pemetaan kompetensi pembelajar.
3. Rekomendasi pembelajaran dibatasi pada penawaran belajar yang dihimpun dan dikurasi dalam penelitian ini, yaitu himpunan data kursus yang didapatkan dari Kaggle yang digunakan sebagai basis pengetahuan operasional sistem.
4. Evaluasi penelitian dibatasi pada observasi operasional terhadap 147 asesmen selesai dalam jendela pengamatan satu hari, uji coba pengguna pada 10 partisipan, dan audit manual kualitas hasil pada sampel uji coba yang sama.
5. Penelitian tidak berfokus pada pengembangan antarmuka pengguna secara mendalam, pengukuran dampak pembelajaran jangka panjang, atau generalisasi hasil ke seluruh konteks pendidikan tinggi di Indonesia.

1.4 Tujuan dan Manfaat Penelitian

Bagian ini menjelaskan arah yang ingin dicapai penelitian serta kontribusi yang dihasilkan setelah penelitian selesai dilakukan.

1.4.1 Tujuan Penelitian

Tujuan umum penelitian ini adalah mengembangkan dan mengevaluasi MVP MicroCred GenAI Backend untuk mendukung proses asesmen kredensial mikro dan rekomendasi pembelajaran yang lebih relevan serta dapat ditelusuri.

Tujuan khusus penelitian ini adalah sebagai berikut.

1. Merancang dan mengimplementasikan MVP MicroCred GenAI Backend sebagai layanan sisi peladen yang mendukung pemrosesan asesmen dan rekomendasi secara asinkron.
2. Mengembangkan alur asesmen kredensial mikro dengan mengintegrasikan penilaian berbasis aturan dan penilaian esai berbasis model GPT menggunakan rubrik penilaian yang terdokumentasi.
3. Membangun mekanisme rekomendasi pembelajaran berbasis retrieval-augmented generation (RAG) yang berlandaskan himpunan data kursus dan dilengkapi jejak pengambilan untuk mendukung keterlacakan hasil.

1.4.2 Manfaat Penelitian

Penelitian ini memberikan manfaat sebagai berikut.

1. Bagi institusi pendidikan, penelitian ini menyediakan acuan awal untuk pengembangan layanan asesmen dan rekomendasi kredensial mikro berbasis GenAI yang lebih terstruktur.
2. Bagi mahasiswa, penelitian ini menyediakan proses asesmen dan rekomendasi yang lebih terdokumentasi sehingga dasar rekomendasi pembelajaran lebih mudah dipahami.
3. Bagi dosen atau pendamping akademik, penelitian ini menyediakan contoh keluaran rekomendasi yang dapat diperiksa sebagai bahan pendampingan dan pengambilan keputusan belajar.
4. Bagi pengembangan keilmuan, penelitian ini menyediakan studi kasus penerapan GenAI dan RAG pada domain kredensial mikro dengan fokus pada keterlacakan proses dan keluaran rekomendasi.

1.5 Metodologi Penelitian Secara Umum

Penelitian ini menggunakan metodologi *Rapid Application Development* (RAD), yang menekankan pembangunan purwarupa secara cepat, iteratif, dan berorientasi pada umpan balik pengguna dalam setiap siklus pengembangan (Meyliana dkk., 2024).

Metodologi tersebut dipilih karena penelitian berfokus pada pembangunan *minimum viable product* (MVP) yang memerlukan iterasi cepat antara perumusan kebutuhan, pembuatan purwarupa, pengumpulan umpan balik, dan penyempurnaan hasil. Secara operasional, penelitian ini mengikuti empat tahap RAD yang juga dipakai pada BAB III, yaitu *define requirements*, *prototype*, *absorb feedback*, dan *finalize product*.

Penekanan pada iterasi dan umpan balik tersebut penting karena penelitian ini tidak berhenti pada perancangan konseptual, tetapi menilai kelayakan awal sistem melalui observasi operasional dan penggunaan terbatas pada konteks nyata penelitian.

Tahapan umum penelitian diringkas sebagai berikut.

1. Studi literatur, yaitu mengkaji literatur mutakhir 2019–2025 serta referensi dasar yang relevan mengenai kredensial mikro, pembelajaran daring, asesmen, GenAI, RAG, sistem rekomendasi, dan keterlacakan rekomendasi.
2. Pendefinisian kebutuhan, yaitu menyusun kebutuhan fungsional, kebutuhan nonfungsional, indikator evaluasi, dan ruang lingkup MVP berdasarkan masalah penelitian.
3. Pembuatan purwarupa, yaitu mengimplementasikan MicroCred GenAI Backend dan menyiapkan basis pengetahuan berupa himpunan data kursus yang dipakai dalam rekomendasi.
4. Observasi operasional dan uji coba pengguna, yaitu mengumpulkan data operasional dari hasil asesmen, menjalankan uji coba pada partisipan, dan melakukan audit manual kualitas hasil pada sampel uji coba yang sama.
5. Analisis hasil dan penyempurnaan, yaitu menganalisis nilai awal kinerja sistem, progres pemrosesan asinkron, kualitas keluaran, hasil audit manual, dan tanggapan pengguna untuk mengidentifikasi perbaikan yang diperlukan.
6. Finalisasi produk dan penyusunan laporan, yaitu merapikan artefak sistem, mendokumentasikan hasil penelitian, dan menyusun laporan penelitian.

1.6 Struktur Laporan / Sistematika Penulisan

Laporan penelitian ini disusun dalam lima bab sebagai berikut.

a. BAB I: PENDAHULUAN

BAB menyajikan latar belakang penelitian, rumusan masalah dan pertanyaan penelitian, lingkup penelitian, tujuan dan manfaat penelitian, metodologi penelitian secara umum, serta struktur laporan.

b. BAB II: KAJIAN PUSTAKA

BAB ini menyajikan kajian pustaka dan justifikasi pemilihan solusi untuk sistem rekomendasi kredensial mikro berbasis GenAI serta komponen penilaian esai otomatis yang menopang proses asesmen.

c. BAB III: METODOLOGI PENELITIAN

BAB ini menjabarkan penerapan solusi, rancangan sistem, serta metode evaluasi yang benar-benar dilakukan.

d. BAB IV: HASIL DAN PEMBAHASAN

BAB ini menyajikan hasil observasi operasional, uji coba pengguna, dan pembahasan berdasarkan indikator evaluasi.

e. BAB V: KESIMPULAN DAN SARAN

BAB ini memuat kesimpulan, keterbatasan penelitian, dan saran pengembangan lanjutan.

BAB II

KAJIAN PUSTAKA

2.1 Kontekstualisasi Masalah dan Kriteria Solusi

BAB I menempatkan penelitian ini pada tiga masalah utama, yaitu kebutuhan mempercepat asesmen kredensial mikro, kebutuhan meningkatkan relevansi rekomendasi berdasarkan profil kompetensi pembelajar, dan kebutuhan menjaga keterjelasan serta keterlacakan hasil rekomendasi. Karena itu, solusi yang dikaji pada bab ini tidak hanya dinilai dari kemampuannya menghasilkan daftar kursus, tetapi juga dari kesesuaiannya dengan kondisi penelitian dan keterbatasan implementasi MVP.

Kondisi awal sistem masih berada pada fase minim data interaksi pengguna. Dalam literatur sistem rekomendasi, kondisi tersebut dikenal sebagai *cold-start*, yaitu situasi ketika sistem belum memiliki histori interaksi yang cukup untuk mempelajari preferensi pengguna atau kualitas objek rekomendasi secara andal (Schein dkk., 2002). Pada konteks penelitian ini, sumber utama rekomendasi bukan riwayat klik atau penilaian pengguna dalam jumlah besar, melainkan hasil asesmen, profil pembelajar, dan metadata pada himpunan *course dataset*. Karena itu, tolak ukur awal rekomendasi pada penelitian ini perlu tetap dapat bekerja tanpa histori interaksi pengguna yang kaya, tetapi juga tetap dapat ditelusuri. Tolak ukur awal yang digunakan bukan TF-IDF murni, melainkan pemeringkatan hibrida ringan yang menggabungkan beberapa sinyal relevansi dari himpunan data kursus.

Literatur sistem rekomendasi menegaskan bahwa pemilihan teknik harus mempertimbangkan karakteristik data, tujuan sistem, dan konsekuensi praktis dari teknik yang dipilih, bukan hanya performa umum pada konteks lain (Adomavicius & Tuzhilin, 2005). Pada konteks pembelajaran, pertimbangan ini menjadi lebih penting karena rekomendasi perlu membantu pembelajar memilih jalur belajar yang sesuai, bukan sekadar meniru pola interaksi pengguna yang sudah ada (Verbert dkk., 2012).

Berdasarkan konteks tersebut, penelitian ini menetapkan kriteria solusi pada Tabel 2.1. Kriteria tersebut dipakai untuk membandingkan pendekatan pada Subbab 2.2 dan menjelaskan alasan pemilihan solusi pada Subbab 2.3.

Tabel 2. 1 Kriteria solusi sistem rekomendasi kredensial mikro berbasis GenAI

Kriteria	Implikasi terhadap pemilihan solusi	Justifikasi	Referensi kunci
Ketahanan terhadap <i>cold-start</i>	Solusi tetap dapat bekerja ketika data interaksi pengguna masih terbatas	Program kredensial mikro baru cenderung belum memiliki histori interaksi yang kaya, sehingga pendekatan yang terlalu bergantung pada histori berisiko kurang layak pada fase awal	(Schein dkk., 2002)
Personalisasi rekomendasi	Solusi mampu mengaitkan rekomendasi dengan kebutuhan, profil, atau hasil asesmen pembelajar	Kajian <i>artificial intelligence</i> (AI) pada kredensial mikro menunjukkan potensi pembentukan jalur belajar yang lebih adaptif dan relevan terhadap kebutuhan pembelajar	(Harizan, 2025)
Kesesuaian dengan metadata keterampilan	Solusi dapat memanfaatkan deskripsi kursus, topik, tingkat (<i>level</i>), tag (label teknis), dan metadata keterampilan secara eksplisit	Rekomendasi penelitian perlu berangkat dari profil kompetensi dan katalog kursus lokal, bukan hanya dari pola perilaku pengguna	(Ricci dkk., 2015)
Keterjelasan dan keterlacakan hasil	Solusi memungkinkan alasan rekomendasi dijelaskan melalui sumber atau bukti yang dapat diperiksa	Rekomendasi yang memengaruhi keputusan belajar perlu dapat dipahami, ditelusuri, dan dipertanggungjawabkan	(Bernardo, 2023)
Kelayakan implementasi MVP	Solusi cukup sederhana untuk dibangun, diuji, dan dievaluasi cepat pada siklus pengembangan iteratif	Penelitian mengembangkan MVP dengan metodologi RAD, sehingga tolok ukur awal perlu realistis terhadap waktu dan kompleksitas implementasi	(Meyliana dkk., 2024)

Kriteria pada Tabel 2.1 tidak dimaksudkan sebagai daftar teori terpisah, melainkan sebagai alat analisis untuk menilai apakah suatu teknik benar-benar sesuai dengan masalah penelitian ini.

2.2 Kategorisasi Teknik Sistem Rekomendasi

Sistem rekomendasi dapat dikembangkan melalui beberapa pendekatan, mulai dari collaborative filtering, content-based filtering, rekomendasi berbasis large language model (LLM), hingga retrieval-augmented generation (RAG). Setiap pendekatan memiliki karakteristik, kebutuhan data, kelebihan, dan keterbatasan yang berbeda. Oleh karena itu,

sebelum membahas literatur masing-masing teknik secara lebih rinci, bagian ini terlebih dahulu menyajikan perbandingan dasar beberapa pendekatan yang relevan terhadap kebutuhan penelitian.

Dalam konteks penelitian ini, teknik rekomendasi tidak hanya dinilai dari kemampuan menghasilkan daftar kursus, tetapi juga dari kesesuaiannya terhadap kondisi minimum viable product (MVP), keterbatasan data interaksi awal, kebutuhan keterjelasan hasil, serta kemampuan sistem untuk menyimpan jejak pengambilan rekomendasi. Perbandingan dasar teknik rekomendasi yang dipertimbangkan dalam penelitian ini ditunjukkan pada Tabel 2.2.

Tabel 2. 2 Kategorisasi teknik rekomendasi dan implikasinya

Kategori dan teknik	Level analisis dalam penelitian	Kebutuhan data interaksi historis	Ketahanan terhadap cold-start	Keterjelasan hasil rekomendasi	Potensi personalisasi	Posisi untuk penelitian ini
Penyaringan kolaboratif / Collaborative Filtering (CF)	Teknik rekomendasi	Tinggi	Rendah	Rendah	Tinggi ketika data matang	Kurang sesuai karena sistem belum memiliki matriks interaksi pengguna-kursus yang matang
Penyaringan berbasis konten / Content-Based Filtering (CBF)	Teknik rekomendasi	Rendah	Tinggi	Sedang-tinggi	Sedang	Sesuai sebagai dasar logika pencocokan karena rekomendasi perlu memetakan profil kompetensi ke atribut dan metadata kursus
Model Bahasa Besar / Large Language Model (LLM) Murni	Teknik rekomendasi	Rendah	Sedang	Rendah jika tidak dibatasi sumber	Sedang-tinggi	Kurang sesuai sebagai mesin rekomendasi utama karena sulit menjaga keterikatan hasil pada sumber
Generasi berbantuan pengambilan / Retrieval-Augmented Generation (RAG)	Kerangka solusi umum	Rendah	Tinggi	Tinggi	Sedang-tinggi	Dipilih sebagai payung arsitektural karena menjaga keterikatan hasil pada dataset kursus dan jejak pengambilan
RAG dengan pemeringkatan hibrida ringan dan	Baseline operasional	Rendah	Tinggi	Tinggi	Sedang	Dipilih sebagai bentuk implementasi

Kategori dan teknik	Level analisis dalam penelitian	Kebutuhan data interaksi historis	Ketahanan terhadap cold-start	Keterjelasan hasil rekomendasi	Potensi personalisasi	Posisi untuk penelitian ini
penyaringan metadata						awal karena mekanisme pengambilan kursusnya sederhana, dapat diaudit, dan tetap relevan terhadap profil pembelajar

Berdasarkan Tabel 2.2, pendekatan yang paling sesuai untuk penelitian ini adalah RAG dengan pemeringkatan hibrida ringan. Collaborative filtering belum sesuai karena sistem belum memiliki data interaksi pengguna-kursus yang matang. Content-based filtering sesuai sebagai dasar pencocokan, tetapi belum cukup untuk menjawab kebutuhan keterlacakan dan pengambilan kandidat yang lebih terstruktur. Rekomendasi berbasis LLM murni juga tidak dipilih sebagai mekanisme utama karena hasilnya berisiko tidak terikat pada dataset kursus yang digunakan. Oleh karena itu, RAG dipilih sebagai kerangka utama, sedangkan pemeringkatan hibrida ringan digunakan sebagai mekanisme operasional awal pada tahap MVP.

Setelah perbandingan dasar tersebut, subbagian berikutnya membahas masing-masing pendekatan secara lebih rinci berdasarkan literatur dan implikasinya terhadap rancangan MicroCred GenAI Backend.

2.2.1 Pendekatan dengan *Collaborative Filtering* (CF)

Collaborative Filtering (CF) adalah pendekatan sistem rekomendasi yang memanfaatkan pola interaksi historis antara pengguna dan objek rekomendasi untuk menghasilkan rekomendasi. Gagasan utamanya adalah pengguna dengan pola preferensi serupa cenderung tertarik pada objek rekomendasi yang serupa pula (Adomavicius & Tuzhilin, 2005).

Pada kondisi ketika data interaksi cukup besar dan relatif stabil, pendekatan ini dapat menghasilkan personalisasi yang kuat tanpa perlu menganalisis isi objek rekomendasi secara mendalam (Ricci dkk., 2015). Namun, kekuatan tersebut baru terasa ketika sistem telah memiliki matriks interaksi pengguna-kursus yang cukup kaya, misalnya riwayat klik, pengambilan kursus, penilaian pengguna, atau sinyal umpan balik dalam skala yang memadai.

Kondisi tersebut belum sesuai dengan konteks penelitian ini. Sumber utama rekomendasi pada penelitian ini bukan histori interaksi yang kaya, melainkan hasil asesmen, profil

pembelajar, jawaban esai, dan metadata pada himpunan data kursus. Data umpan balik memang dapat dikumpulkan, tetapi pada tahap awal belum cukup banyak dan belum cukup stabil untuk membentuk fondasi CF yang andal.

Selain itu, CF tidak secara eksplisit menjadikan metadata keterampilan, tingkat kesulitan, dan deskripsi kursus sebagai dasar utama rekomendasi. Akibatnya, alasan rekomendasi menjadi lebih sulit dijelaskan kepada pengguna, padahal konteks pembelajaran menuntut hasil yang dapat ditelusuri dan dipertanggungjawabkan (Verbert dkk., 2012).

Berdasarkan pertimbangan tersebut, CF dipandang kurang sesuai sebagai solusi tolok ukur awal pada penelitian ini. Dari sisi kebutuhan data historis dan ketahanan terhadap kondisi minim data awal, CF lemah; dari sisi pemanfaatan metadata dan keterjelasan hasil, CF juga kurang mendukung; dan dari sisi kelayakan tolok ukur awal MVP, pendekatan ini belum cukup realistis untuk implementasi awal. Pendekatan ini tetap relevan sebagai arah pengembangan lanjutan ketika histori interaksi dan data umpan balik sistem telah lebih matang, tetapi belum tepat untuk konteks implementasi awal yang masih bertumpu pada asesmen dan himpunan data kursus.

2.2.2 Pendekatan dengan *Content-Based Filtering* (CBF)

Content-Based Filtering (CBF) menghasilkan rekomendasi berdasarkan kesamaan karakteristik isi objek rekomendasi dengan profil atau kebutuhan pengguna. Pada konteks pembelajaran, karakteristik tersebut dapat berupa topik, keterampilan, tingkat kesulitan, deskripsi kursus, atau metadata kursus lainnya (Adomavicius & Tuzhilin, 2005).

Kelebihan utama CBF pada penelitian ini adalah ketahanannya terhadap keterbatasan data interaksi historis. Selama objek rekomendasi memiliki deskripsi dan metadata yang cukup, rekomendasi masih dapat dihasilkan melalui pencocokan konten antara kebutuhan pembelajar dan representasi kursus (Ricci dkk., 2015). Karakter ini cocok untuk konteks ketika keputusan rekomendasi perlu diturunkan dari kesesuaian antara profil kompetensi pembelajar dan atribut kursus, bukan dari pola interaksi pengguna yang sudah mapan.

CBF juga lebih mudah dijelaskan daripada CF karena alasan rekomendasi dapat ditelusuri dari kesesuaian atribut objek rekomendasi. Karakteristik ini penting bagi sistem rekomendasi pembelajaran yang perlu menunjukkan mengapa suatu kursus dianggap relevan terhadap profil pembelajar (Verbert dkk., 2012). Dalam konteks penelitian ini, logika tersebut tampak pada penggunaan metadata kursus seperti subjek, tingkat, dan tag teknologi, durasi, preferensi pembayaran, dan sinyal profil pengguna.

Namun, CBF murni belum sepenuhnya menjelaskan kebutuhan solusi pada penelitian ini. Penelitian ini tidak hanya membutuhkan pencocokan atribut, tetapi juga membutuhkan pengambilan kursus yang terstruktur, mekanisme penyaringan, penyusunan alasan rekomendasi, dan penyimpanan jejak hasil agar keputusan sistem dapat ditelusuri kembali.

Karena itu, CBF dinilai relevan sebagai dasar logika pencocokan rekomendasi. Dari sisi ketahanan terhadap kondisi minim data awal, kesesuaian dengan metadata, dan keterjelasan hasil, CBF cukup kuat; namun dari sisi personalisasi lanjut dan kelayakan sebagai solusi lengkap, CBF murni belum cukup karena belum menjelaskan pengambilan kursus, pemeringkatan, dan penyimpanan jejak rekomendasi secara terintegrasi. Meski demikian, penelitian ini tidak berhenti pada CBF murni karena solusi yang diterapkan perlu tetap mengikat pengambilan kursus, pemeringkatan, dan penyimpanan jejak rekomendasi pada himpunan data kursus melalui kerangka RAG.

2.2.3 Pendekatan dengan Rekomendasi Berbasis *Large Language Model* (LLM) Murni

Large Language Model (LLM) dapat digunakan sebagai komponen utama untuk menyusun rekomendasi dan penjelasan secara langsung dari masukan pengguna, tanpa membatasi keluaran pada hasil pengambilan dari himpunan data kursus.

Pendekatan ini memiliki daya tarik karena mampu menghasilkan narasi rekomendasi yang fleksibel, kaya konteks, dan mudah dibaca. Dalam beberapa kasus, LLM juga dapat membantu merangkum kebutuhan pengguna dan menghubungkannya dengan konsep pembelajaran yang lebih luas.

Meskipun demikian, penggunaan LLM sebagai mesin rekomendasi utama memiliki risiko penting pada konteks penelitian ini. LLM dapat menghasilkan keluaran yang tampak meyakinkan tetapi tidak sesuai dengan sumber yang valid atau tidak benar-benar tersedia dalam himpunan data kursus yang digunakan sistem (Ji dkk., 2023). Pada sistem pembelajaran, kondisi tersebut bermasalah karena alasan rekomendasi harus dapat diaudit.

Pembedaan fungsi LLM menjadi penting pada penelitian ini. LLM cocok digunakan untuk menilai jawaban esai yang bersifat terbuka karena tugas tersebut membutuhkan interpretasi bahasa alami yang tidak mudah diselesaikan secara deterministik. Namun, untuk rekomendasi utama, sistem tetap perlu terikat pada himpunan data kursus, disusun sebagai objek rekomendasi berperingkat, dan dilengkapi alasan yang dapat ditelusuri.

Literatur tentang risiko halusinasi juga menunjukkan bahwa model dapat menghadapi ketidakpastian keluaran bahkan ketika responsnya tampak lancar secara bahasa (Ji dkk., 2024).

Berdasarkan pertimbangan tersebut, LLM tidak dipilih sebagai tolok ukur awal rekomendasi utama. Dari sisi kebutuhan data historis, LLM tidak bermasalah; tetapi dari sisi keterikatan pada metadata lokal, kejelasan hasil, dan kelayakan sebagai tolok ukur awal rekomendasi yang dapat diaudit, pendekatan ini tetap lemah.

2.2.4 Pendekatan dengan Generasi Berbantuan Pengambilan / Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) menggabungkan pengambilan informasi dari basis pengetahuan dengan proses penyusunan keluaran, sehingga hasil sistem tetap terhubung dengan sumber yang digunakan (Lewis dkk., 2020).

Pada konteks sistem rekomendasi pembelajaran, RAG memungkinkan sistem mengambil kursus yang relevan dari himpunan data kursus terlebih dahulu, kemudian menyusun hasil rekomendasi berdasarkan objek rekomendasi yang memang tersedia. Dengan cara ini, rekomendasi tidak hanya bersifat relevan, tetapi juga terikat pada sumber yang dapat diperiksa kembali (Lewis dkk., 2020)

Kekuatan utama RAG pada penelitian ini adalah kemampuannya menjembatani kebutuhan personalisasi dengan kebutuhan kejelasan. Selama mekanisme pengambilan dan sumber pengetahuan dikelola dengan baik, sistem dapat menampilkan objek rekomendasi beserta alasan yang merujuk pada metadata dan dokumen yang digunakan (Bernardo, 2023). Karakter ini penting ketika hasil asesmen perlu dihubungkan ke rekomendasi yang tetap terikat pada himpunan data kursus yang tersedia.

Keterbatasan RAG terletak pada kualitas basis pengetahuan dan kualitas mekanisme pengambilan. Jika himpunan data kursus tidak dikurasi dengan baik atau jika strategi pengambilan tidak sesuai dengan karakteristik data, maka kualitas rekomendasi juga menurun (Lewis dkk., 2020). Pada konteks penelitian ini, hal tersebut berarti kualitas pengambilan kursus yang relevan tetap sangat bergantung pada mutu metadata kursus dan strategi penyaringan yang diterapkan.

Dalam konteks penelitian ini, RAG dipandang sebagai kerangka solusi yang paling sesuai karena mampu menjaga keterikatan hasil terhadap himpunan data kursus sekaligus membuka ruang bagi penyusunan rekomendasi yang lebih mudah diaudit daripada CF atau LLM murni. Dari sisi kondisi minim data awal, pemanfaatan metadata, dan kejelasan hasil, RAG kuat; namun sebagai kerangka umum, RAG belum sekaligus menjawab bentuk operasional tolok ukur awal yang paling sederhana untuk MVP.

2.2.5 Pendekatan dengan RAG Berbasis Pemeringkatan Hibrida Ringan

Dalam penelitian ini, mekanisme operasional di bawah kerangka *Retrieval-Augmented Generation* (RAG) tidak menggunakan TF-IDF murni. Sesuai rancangan implementasi pada BAB III, RAG diterapkan melalui pemeringkatan hibrida ringan yang menggabungkan beberapa sinyal relevansi, yaitu pencocokan metadata yang diperkaya, relevansi kata kunci, kemiripan vektor token berbasis fungsi *hashing*, *tag boost*, dan skor kualitas kursus. Pemilihan pendekatan hibrida ini sejalan dengan prinsip sistem rekomendasi bahwa teknik rekomendasi perlu disesuaikan dengan karakteristik data, tujuan sistem, serta ketersediaan sinyal pengguna maupun objek rekomendasi (Adomavicius & Tuzhilin, 2005; Ricci dkk., 2015).

Pencocokan *metadata* yang diperkaya digunakan untuk menilai kesesuaian kandidat kursus dengan profil kompetensi pembelajar, topik, tingkat, label, durasi, dan atribut lain yang tersedia pada himpunan data kursus. Relevansi kata kunci digunakan untuk membaca kedekatan antara kebutuhan pembelajar dan deskripsi kursus berdasarkan istilah yang muncul pada masukan pengguna, hasil asesmen, atau metadata kursus. Logika ini dekat dengan pendekatan CBF karena rekomendasi dibangun dari kesesuaian antara karakteristik objek rekomendasi dan kebutuhan pengguna, bukan dari histori interaksi pengguna yang sudah matang (Ricci dkk., 2015; Verbert dkk., 2012).

Selain itu, sistem juga menggunakan kemiripan vektor token berbasis fungsi *hashing* sebagai sinyal tambahan untuk membandingkan representasi ringan antara kueri dan kursus. Istilah *embedding_score* pada implementasi tidak merujuk pada *neural embedding* seperti representasi vektor dari model eksternal atau penyimpanan vektor pada basis data khusus. Istilah tersebut merujuk pada kemiripan vektor token berbasis fungsi *hashing* yang dibuat secara lokal dari token kursus dan kueri. Pendekatan representasi berbasis fungsi *hashing* sejalan dengan gagasan *feature hashing*, yaitu penggunaan fungsi hash untuk membentuk representasi vektor yang ringan dan hemat ruang (Weinberger dkk., 2009). Dengan demikian, pendekatan ini tetap ringan dan sesuai untuk tahap produk minimum yang layak, tetapi belum setara dengan sistem pengambilan semantik berbasis representasi vektor neural.

Keunggulan utama pendekatan ini adalah keseimbangannya antara kesederhanaan implementasi, ketahanan terhadap kondisi minim data awal, dan keterlacakan hasil. Sistem tidak memerlukan histori interaksi pengguna yang besar seperti *collaborative filtering*, sehingga lebih sesuai untuk kondisi minim data awal / *cold-start* pada fase awal sistem (Schein dkk., 2002). Selain itu, rekomendasi tetap diambil dari himpunan data kursus yang tersedia dan

disertai alasan kecocokan yang dapat diperiksa kembali, sehingga selaras dengan prinsip RAG dan kebutuhan keterjelasan sistem rekomendasi (Bernardo, 2023; Lewis dkk., 2020).

Keterbatasan pendekatan ini adalah kemampuan semantiknya masih terbatas. Karena tidak menggunakan representasi vektor neural / *neural embedding* dan bukan TF-IDF murni, kualitas rekomendasi sangat bergantung pada kualitas token, metadata, deskripsi kursus, tag, serta logika pembobotan yang ditentukan pada sistem. Selain itu, pembatasan peran LLM tetap diperlukan agar sistem tidak menghasilkan rekomendasi bebas yang sulit diverifikasi, mengingat model LLM dapat menghasilkan keluaran yang tampak meyakinkan tetapi tidak selalu memiliki dasar sumber yang valid (Ji dkk., 2023, 2024). Oleh karena itu, pendekatan ini tepat diposisikan sebagai tolak ukur awal MVP yang sederhana dan dapat diaudit, bukan sebagai teknik rekomendasi paling kuat secara teoritis.

2.2.6 Simpulan Analisis Perbandingan

Hasil kajian menunjukkan bahwa kebutuhan utama penelitian ini bukan sekadar personalisasi setinggi mungkin, melainkan personalisasi yang selaras dengan bentuk data dan kebutuhan operasional sistem. Karena itu, teknik yang terlalu bergantung pada histori interaksi matang atau terlalu bebas menghasilkan keluaran tidak menjadi pilihan utama.

CF tersisih karena penelitian ini belum bertumpu pada data interaksi pengguna-kursus yang cukup kaya untuk menopang rekomendasi kolaboratif. CBF memberikan dasar yang kuat karena rekomendasi memang dibangun dari kesesuaian konten dan metadata kursus terhadap profil pembelajar, tetapi CBF murni belum cukup mewakili kebutuhan arsitektural penelitian yang menuntut pengambilan kursus yang terstruktur, alasan rekomendasi, dan penyimpanan jejak hasil.

LLM tetap bernilai sebagai perkembangan penting dalam AI pendidikan, tetapi sebagai mekanisme rekomendasi utama pada konteks penelitian ini pendekatan tersebut tidak dipilih karena sulit menjaga keterikatan hasil pada himpunan data kursus dan alasan rekomendasi yang dapat diaudit. Pada komponen rekomendasi, kerangka RAG menjadi wadah yang paling sesuai karena menjaga keterikatan rekomendasi pada himpunan data kursus, sedangkan pemeringkatan hibrida ringan menjadi mekanisme pengambilan dan pengurutan kursus pada tolak ukur awal yang paling masuk akal untuk karakter data, kebutuhan audit, dan batasan implementasi awal.

Dengan demikian, analisis perbandingan pada bagian ini menempatkan RAG sebagai kerangka solusi yang dipilih, sedangkan RAG dengan pemeringkatan hibrida ringan

ditempatkan sebagai tolok ukur awal operasional yang paling sesuai untuk konteks penelitian ini.

2.3 Pemilihan Solusi untuk Penelitian

Berdasarkan kriteria pada Tabel 2.1 dan hasil perbandingan pada Tabel 2.2, penelitian ini memilih RAG sebagai kerangka solusi untuk rekomendasi pembelajaran, dengan pemeringkatan hibrida ringan sebagai tolok ukur awal implementasi. Tolak ukur awal ini menghitung skor kandidat kursus melalui pencocokan metadata yang diperkaya, relevansi kata kunci, kemiripan vektor token berbasis fungsi *hashing*, *tag boost*, dan skor kualitas kursus.

Keputusan ini diambil karena pendekatan tersebut paling sesuai dengan kondisi penelitian, yaitu data interaksi historis yang masih terbatas, kebutuhan keterjelasan rekomendasi, dan kebutuhan implementasi tolok ukur awal yang layak dievaluasi cepat pada siklus pengembangan MVP. Pada pendekatan ini, rekomendasi tetap dibangun dari himpunan data kursus yang tersedia melalui logika pencocokan berbasis konten, sehingga alasan hasil dapat dirujuk kembali pada sumber yang jelas (Lewis dkk., 2020; Ricci dkk., 2015).

Untuk komponen penilaian esai otomatis, penelitian ini memilih GPT berbasis rubrik sebagai pendekatan implementasi. Keputusan ini diambil karena jawaban esai terbuka membutuhkan interpretasi bahasa alami yang tidak cukup diwakili oleh aturan murni, sementara pembangunan *automated essay scoring* (AES) tradisional memerlukan korpus skor manusia yang lebih besar daripada yang tersedia pada tahap MVP. Dalam implementasinya, keluaran GPT tetap dibatasi oleh rubrik, validasi format *JavaScript Object Notation* (JSON), mekanisme *fallback* deterministik, dan penyimpanan metadata audit agar hasil penilaian tetap dapat diperiksa kembali (Xiao dkk., 2025).

Dengan demikian, penelitian ini tidak menggunakan LLM sebagai generator rekomendasi bebas. Peran LLM difokuskan pada penilaian esai otomatis berbasis rubrik, sedangkan komponen rekomendasi dibangun melalui pengambilan dan pemeringkatan hibrida ringan agar dasar keputusan sistem tetap sederhana, terukur, dan mudah diaudit (Ji dkk., 2023).

Pendekatan hibrida yang menambahkan pemeringkatan ulang generatif, *collaborative filtering*, atau personalisasi berbasis histori interaksi tidak dijadikan bagian tolok ukur awal pada tahap ini. Pendekatan tersebut ditempatkan sebagai ruang pengembangan lanjutan setelah data penggunaan, kebutuhan evaluasi, dan mekanisme audit sistem lebih matang.

2.4 Dasar Teori dan Konsep Solusi Terpilih

Bagian ini menjelaskan teori yang benar-benar menopang solusi terpilih dan yang dibutuhkan untuk membaca rancangan penerapan pada BAB III. Teori yang tidak terkait langsung dengan solusi dan evaluasi penelitian tidak dibahas lebih lanjut pada bagian ini.

2.4.1 Konsep RAG untuk Rekomendasi Pembelajaran

RAG adalah pendekatan yang menggabungkan pengambilan informasi dari basis pengetahuan dengan penyusunan keluaran, sehingga hasil sistem tidak berdiri lepas dari sumber yang digunakan (Lewis dkk., 2020).

Dalam konteks rekomendasi pembelajaran, konsep ini penting karena sistem tidak seharusnya menyarankan item yang tidak tersedia atau tidak memiliki dasar yang jelas pada dataset kursus. Dengan RAG, sistem terlebih dahulu mengambil kursus yang relevan dari dataset kursus, lalu menyusun hasil rekomendasi berdasarkan kandidat yang sudah terpilih (Lewis dkk., 2020).

Konsep tersebut sejalan dengan kebutuhan penelitian ini untuk menjaga agar hasil rekomendasi tetap terikat pada dataset kursus dan dapat dijelaskan melalui sumber yang dipakai selama proses pengambilan.

2.4.2 Pemeringkatan Hibrida Ringan sebagai Mekanisme Pengambilan dan Pengurutan Awal

Pada penelitian ini, pengambilan dan pengurutan kandidat kursus dilakukan melalui pemeringkatan hibrida ringan. Mekanisme ini tidak bergantung pada satu sinyal tunggal, tetapi menggabungkan beberapa komponen skor yang tersedia dari himpunan data kursus dan profil pembelajar.

Komponen pertama adalah pencocokan metadata yang diperkaya, yaitu penilaian kesesuaian antara profil pembelajar dan atribut kursus seperti topik, tingkat, label, durasi, serta karakteristik kursus lain yang tersedia. Komponen kedua adalah relevansi kata kunci, yaitu penilaian kedekatan antara istilah pada kueri, hasil asesmen, dan deskripsi kursus. Komponen ketiga adalah kemiripan vektor token berbasis fungsi *hashing*, yaitu representasi ringan yang mengubah token menjadi vektor melalui fungsi pencacahan dan menghitung kedekatan antara kueri dan kursus. Komponen lain yang digunakan adalah *tag boost* dan skor kualitas kursus.

Secara operasional, skor akhir kandidat kursus dihitung melalui kombinasi berbobot dari komponen-komponen tersebut. Dalam implementasi MVP, bobot terbesar diberikan pada

pencocokan metadata yang diperkaya karena penelitian ini menekankan kesesuaian rekomendasi dengan profil kompetensi pembelajar. Relevansi kata kunci dan kemiripan vektor token berbasis fungsi *hashing* digunakan sebagai sinyal tambahan agar sistem tetap dapat membaca kedekatan tekstual antara kebutuhan pembelajar dan deskripsi kursus.

Pendekatan ini dipilih karena cukup sederhana untuk tahap MVP, tidak membutuhkan histori interaksi pengguna yang besar, dan tetap memungkinkan hasil rekomendasi ditelusuri melalui skor relevansi, alasan kecocokan, serta jejak pengambilan. Namun, karena representasi vektor yang digunakan masih berbasis fungsi *hashing* lokal, pendekatan ini belum setara dengan pengambilan berbasis representasi vektor neural / *neural embedding retrieval* yang mampu menangkap relasi semantik lebih dalam.

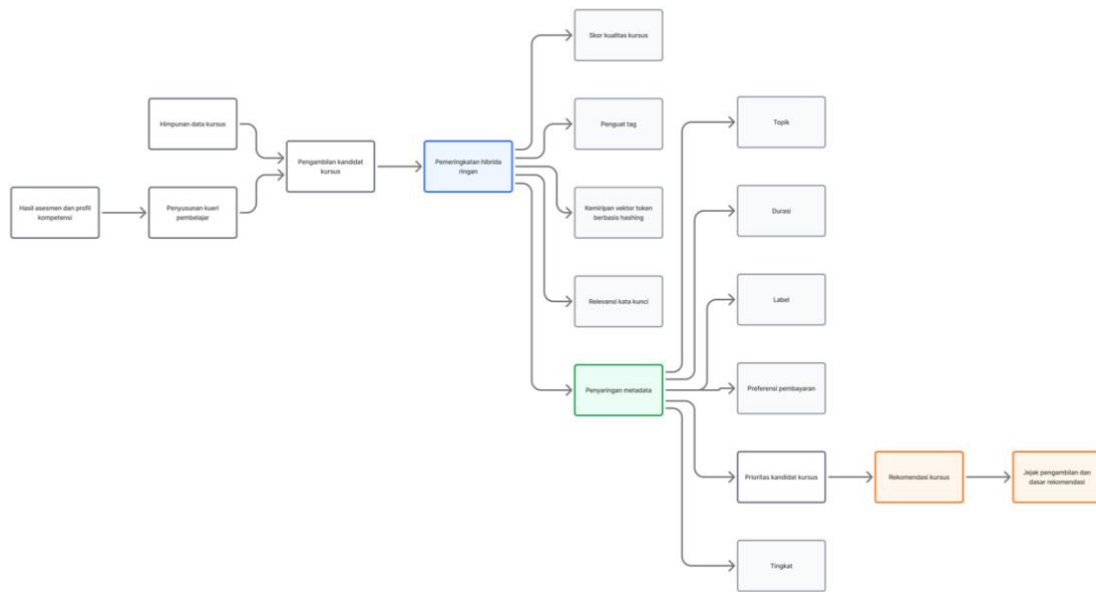
2.4.3 Penyelarasan Keterampilan–Kompetensi dalam Rekomendasi Pembelajaran

Setelah kandidat awal diperoleh, penyaringan metadata digunakan untuk memurnikan dan memprioritaskan hasil agar lebih selaras dengan kebutuhan pembelajar. Metadata yang dapat digunakan pada konteks ini meliputi topik, tingkat, label, durasi, preferensi pembayaran, atau atribut lain yang mendeskripsikan karakteristik kursus (Ricci dkk., 2015).

Peran penyaringan metadata berbeda dari relevansi kata kunci dan kemiripan vektor token berbasis fungsi *hashing*. Relevansi kata kunci dan vektor token membantu membaca kedekatan tekstual antara kebutuhan pembelajar dan deskripsi kursus, sedangkan penyaringan metadata berfokus pada kecocokan atribut terstruktur antara kandidat kursus dan profil pembelajar. Dengan demikian, sistem tidak hanya memilih kursus yang dekat secara tekstual, tetapi juga kursus yang lebih sesuai dengan konteks kebutuhan belajar yang ingin dipenuhi.

Pada penelitian ini, penyaringan metadata penting karena rekomendasi tidak cukup hanya relevan secara kata kunci. Kursus yang direkomendasikan juga perlu konsisten dengan karakteristik yang dapat dipertanggungjawabkan, sehingga hasil akhir lebih mudah dijelaskan kepada pengguna dan lebih dekat dengan logika pencocokan berbasis konten yang dipakai sebagai dasar pencocokan.

Alur konseptual solusi RAG dengan pemeringkatan hibrida ringan ditunjukkan pada Gambar 2.1.



Gambar 2. 1 Alur konseptual RAG dengan pemeringkatan hibrida ringan untuk rekomendasi pembelajaran

Pada Gambar 2.1 merangkum alur konseptual solusi yang dipilih pada penelitian ini. Hasil asesmen dan profil kompetensi dipakai untuk menyusun kueri yang kemudian digunakan untuk pengambilan kursus pada himpunan data kursus (Lewis dkk., 2020). Pada tahap awal, pemeringkatan hibrida ringan dipakai untuk mengambil dan mengurutkan kandidat kursus berdasarkan pencocokan metadata yang diperkaya, relevansi kata kunci, kemiripan vektor token berbasis fungsi *hashing*, *tag boost*, dan skor kualitas kursus (Ricci dkk., 2015; Weinberger dkk., 2009). Keluaran sistem kemudian menyertakan jejak pengambilan untuk mendukung keterjelasan rekomendasi yang dapat ditelusuri kembali ke dasar pemilihannya (Bernardo, 2023).

2.4.4 Keterjelasan dan Jejak Pengambilan

Keterjelasan merupakan kemampuan sistem untuk menunjukkan dasar keputusan yang dihasilkannya secara dapat dipahami dan dapat dipertanggungjawabkan. Pada konteks sistem rekomendasi pembelajaran, kebutuhan ini penting karena rekomendasi memengaruhi keputusan belajar pengguna (Bernardo, 2023).

Pada solusi yang dipilih dalam penelitian ini, keterjelasan diwujudkan melalui jejak pengambilan, yaitu penelusuran kembali objek rekomendasi yang direkomendasikan ke

metadata dan bukti relevansi yang mendasarinya. Dengan pendekatan tersebut, rekomendasi tidak hanya tampil sebagai daftar hasil, tetapi juga sebagai keluaran yang memiliki alasan pemilihan yang dapat diperiksa (Bernardo, 2023).

Konsep keterjelasan ini menjadi salah satu alasan utama mengapa pendekatan yang terlalu bergantung pada representasi laten atau generasi bebas tidak dipilih sebagai tolok ukur awal.

2.4.5 Penilaian Esai Otomatis Berbasis Rubrik dengan GPT

Penilaian esai otomatis berbasis GPT memanfaatkan model bahasa besar untuk menafsirkan jawaban terbuka berdasarkan instruksi dan rubrik yang telah ditentukan. Pada pendekatan ini, model tidak dibiarkan memberi skor secara bebas, tetapi diarahkan agar membaca pertanyaan, jawaban, dan rubrik sebagai satu kesatuan konteks penilaian.

Literatur terbaru menunjukkan bahwa LLM tidak selalu melampaui AES tradisional dalam akurasi mentah, tetapi memiliki kelebihan pada keterjelasan dan generalisasi lintas konteks (Xiao dkk., 2025). Karakter ini relevan bagi penelitian karena sistem memerlukan penilaian terhadap jawaban esai yang bervariasi, tetapi tetap perlu menyimpan alasan dan metadata hasil agar penilaian dapat ditelusuri kembali.

Pada penelitian ini, GPT berbasis rubrik dipilih karena sesuai untuk tugas penilaian jawaban terbuka pada tahap MVP. Agar hasilnya tetap dapat diaudit, model dibatasi oleh rubrik penilaian, validasi format keluaran, mekanisme cadangan deterministik / *fallback* ketika *parsing* gagal, dan penyimpanan metadata audit di sisi *backend*.

2.4.6 Risiko Halusinasi pada Penggunaan LLM

Halusinasi pada LLM merujuk pada keluaran yang tampak meyakinkan, tetapi tidak sesuai dengan fakta atau tidak memiliki dasar yang valid pada sumber yang seharusnya dipakai sistem (Ji dkk., 2023).

Risiko ini relevan bagi penelitian karena sistem yang mendukung proses asesmen dan rekomendasi pembelajaran tidak boleh menghasilkan penjelasan atau saran yang sulit diverifikasi. Karena itu, penggunaan LLM pada penelitian ini dibatasi agar tidak menjadi komponen utama pembangkitan rekomendasi.

Temuan mengenai risiko ketidakpastian keluaran juga menguatkan kebutuhan untuk membatasi peran LLM pada bagian yang memang dapat dikontrol lebih ketat, yaitu penilaian

esai otomatis berbasis rubrik, sementara rekomendasi utama tetap dibangun dari hasil pengambilan yang dapat ditelusuri (Ji dkk., 2024).

2.4.7 Metodologi Rapid Application Development (RAD)

Rapid Application Development (RAD) adalah metodologi pengembangan yang menekankan pembangunan purwarupa secara cepat, iteratif, dan berorientasi pada umpan balik. Pendekatan ini relevan ketika sistem perlu diuji dan disempurnakan dalam siklus pengembangan yang relatif singkat (Meyliana dkk., 2024).

Dalam penelitian ini, RAD dipilih karena penelitian berfokus pada pengembangan MVP yang perlu segera diuji pada pengguna, lalu disempurnakan dari umpan balik dan hasil evaluasi awal. Kebutuhan tersebut membuat pendekatan yang lebih linear dan menunggu spesifikasi final sejak awal menjadi kurang sesuai untuk konteks penelitian ini, karena alur asesmen, penyajian rekomendasi, dan kebutuhan keterjelasan masih perlu dipastikan melalui iterasi purwarupa (Meyliana dkk., 2024).

Dengan demikian, RAD menjadi dasar metodologis yang menghubungkan BAB II dengan BAB III. Pemilihan solusi tolok ukur awal yang sederhana, terukur, dan dapat diimplementasikan cepat pada MVP selaras dengan karakter RAD yang menekankan iterasi purwarupa, evaluasi cepat, dan perbaikan bertahap berdasarkan hasil uji coba (Meyliana dkk., 2024).

2.5 Kesenjangan Penelitian dan Posisi Penelitian

Berdasarkan kajian pada Subbab 2.1 sampai 2.4, penelitian ini mengidentifikasi tiga kesenjangan utama.

Pertama, banyak pendekatan sistem rekomendasi masih mengandalkan data interaksi historis sebagai fondasi utama, padahal konteks program kredensial mikro baru sering kali belum memiliki histori yang cukup. Kondisi ini membuat masalah *cold-start* menjadi kendala nyata bagi penerapan sistem rekomendasi pada fase awal (Schein dkk., 2002).

Kedua, kajian tentang AI pada kredensial mikro menunjukkan potensi personalisasi dan jalur belajar adaptif, tetapi belum cukup menjelaskan bagaimana solusi tersebut dioperasionalkan secara sederhana dan terukur pada konteks institusi yang membangun MVP berbasis himpunan data kursus (Harizan, 2025).

Ketiga, kebutuhan keterjelasan dan auditabilitas hasil semakin penting dalam sistem yang mendukung keputusan pembelajaran, tetapi tidak semua pendekatan menyediakan dasar

rekomendasi yang mudah diperiksa kembali oleh pengguna atau pengelola sistem (Bernardo, 2023).

Berdasarkan kesenjangan tersebut, posisi penelitian ini adalah membangun dan mengevaluasi MVP MicroCred GenAI Backend yang memadukan penilaian esai otomatis berbasis GPT dan LLM dengan rekomendasi pembelajaran berbasis RAG menggunakan pemeringkatan hibrida ringan pada himpunan data kursus. Dengan posisi tersebut, penelitian ini tidak berfokus pada pencarian teknik paling kompleks, melainkan pada solusi tolok ukur awal yang realistis, dapat dijelaskan, dan sesuai dengan kebutuhan implementasi awal pada konteks pendidikan tinggi.

BAB III

METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM

3.1 Metodologi Penelitian

Bab ini menjelaskan metodologi penelitian yang digunakan untuk merancang, membangun, dan mengevaluasi sistem MicroCred GenAI Backend. Secara umum, penelitian ini menggunakan pendekatan penelitian pengembangan perangkat lunak karena menghasilkan artefak berupa layanan sisi peladen / *backend* dan mengevaluasi kelayakan awalnya melalui observasi operasional, uji coba pengguna, audit manual, serta analisis data sistem. Dengan demikian, pembahasan pada bab ini tidak hanya mencakup proses pengembangan sistem, tetapi juga mencakup rancangan penelitian, tahapan pengumpulan data, teknik evaluasi, dan analisis hasil.

Dalam proses pengembangan perangkat lunak, penelitian ini menggunakan metode *Rapid Application Development* (RAD) karena metode ini menekankan pembangunan purwarupa secara iteratif dan kolaboratif, sehingga umpan balik pengguna dapat segera digunakan pada setiap siklus pengembangan (Meyliana dkk., 2024).

Sesuai Gambar 3.1, proses pengembangan sistem mengikuti empat tahap RAD, yaitu: (1) pendefinisian kebutuhan, (2) perancangan purwarupa, (3) penyerapan umpan balik, dan (4) finalisasi produk. Struktur bab ini disusun mengikuti tahapan tersebut, dengan tetap menempatkan RAD sebagai metode pengembangan perangkat lunak di dalam kerangka metodologi penelitian yang lebih luas.

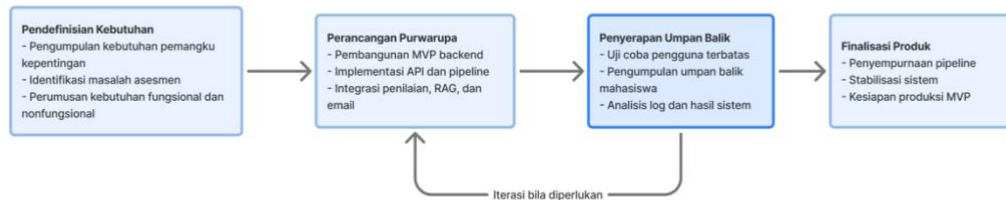
3.2 Diagram Penelitian

Sebelum tahapan penelitian dijelaskan secara rinci, alur penelitian divisualisasikan terlebih dahulu dalam bentuk diagram agar hubungan antar-tahap dapat dipahami secara runtut. Diagram penelitian ini menggambarkan penerapan metodologi *Rapid Application Development* (RAD) pada pengembangan MicroCred GenAI Backend, mulai dari pendefinisian kebutuhan, perancangan purwarupa, penyerapan umpan balik, hingga finalisasi produk.

Setiap tahap pada diagram memiliki keluaran yang menjadi masukan bagi tahap berikutnya. Selain itu, terdapat alur iteratif antara tahap perancangan purwarupa dan penyerapan umpan balik. Alur ini menunjukkan bahwa pengembangan sistem tidak dilakukan secara linear

sepenuhnya, tetapi memungkinkan adanya perbaikan berdasarkan hasil evaluasi, pengujian, dan tanggapan pengguna.

Tahapan RAD yang digunakan dalam penelitian ini divisualisasikan pada Gambar 3.1.



Gambar 3. 1 Tahapan Rapid Application Development (RAD) pada Pengembangan MicroCred GenAI

3.3 Tahapan Penelitian

Tahapan penelitian ini disusun berdasarkan empat tahap utama dalam metodologi *Rapid Application Development* (RAD). Keempat tahap tersebut digunakan untuk memastikan bahwa pengembangan MicroCred GenAI Backend dilakukan secara terarah, iteratif, dan sesuai dengan kebutuhan pengguna. Penjelasan setiap tahap adalah sebagai berikut.

3.3.1 Pendefinisian Kebutuhan / *Define Requirements*

Bagian ini menjelaskan proses perumusan kebutuhan sistem berdasarkan pengumpulan data evaluasi dari pemangku kepentingan. Kebutuhan yang dihasilkan pada tahap ini menjadi dasar bagi perancangan purwarupa pada Bagian 3.3.2.

3.3.1.1 Metode Pengumpulan Data Kebutuhan

Pengumpulan data untuk merumuskan kebutuhan sistem dilakukan melalui formulir umpan balik pemangku kepentingan berbasis Google Forms. Formulir ini disebarluaskan kepada 10 responden dengan latar belakang jurusan Informatika dan Aktuaria pada tanggal 14–16 November 2025. Pemilihan responden dilakukan dengan teknik *purposive-convenience sampling*, yaitu pemilihan responden yang relevan dengan konteks penelitian sekaligus mudah dijangkau selama periode uji agar data yang diperoleh relevan dengan konteks asesmen kredensial mikro.

Struktur formulir umpan balik disajikan pada Tabel 3.1.

Tabel 3. 1 Struktur formulir umpan balik pemangku kepentingan

Bagian	Topik	Pertanyaan Utama	Keluaran
A. Kondisi awal / <i>baseline</i>	Kondisi asesmen saat ini dan titik masalah	Durasi asesmen, kualitas rekomendasi, frekuensi ketidaksesuaian keterampilan / <i>skill mismatch</i> , tingkat keterjelasan	Data kuantitatif (Likert 1-5) dan kualitatif (jawaban terbuka)
B. Kepentingan fitur / <i>feature importance</i>	Validasi ruang lingkup MVP	Kepentingan fitur: waktu penyelesaian / <i>turnaround time</i> , personalisasi, pencocokan keterampilan / <i>skill matching</i> , keterjelasan, kurasi lima teratas / <i>Top-5</i> , putaran umpan balik / <i>feedback loop</i>	Prioritas fitur (fitur wajib / <i>must-have</i> dan fitur sebaiknya ada / <i>should-have</i>)
C. Kelayakan solusi / <i>solution viability</i>	Kelayakan solusi	Keyakinan terhadap solusi berbasis kecerdasan buatan / <i>artificial intelligence</i> (AI); kekhawatiran pengguna	Validasi arah solusi
D. Penutup / <i>closing</i>	Tindak lanjut	Kesediaan mengikuti uji coba; umpan balik tambahan	Daftar responden untuk iterasi berikutnya

Catatan: Hasil pengumpulan data (profil responden, temuan utama, distribusi jawaban) akan dilaporkan pada BAB IV.

3.3.1.2 Analisis Kebutuhan Fungsional

Berdasarkan temuan dari Bagian 3.3.1.1, dirumuskan kebutuhan fungsional yang harus dipenuhi oleh sistem seperti yang disajikan pada Tabel 3.2. Setiap kebutuhan diberi kode FR / *functional requirement* untuk memudahkan penelusuran pada tahap perancangan dan pengujian.

Tabel 3. 2 Kebutuhan fungsional sistem

Kode	Deskripsi kebutuhan
FR1	Sistem dapat memulai atau melanjutkan asesmen berdasarkan peran atau jalur / <i>role/track</i> yang dipilih pengguna.
FR2	Sistem menerima pengiriman jawaban asesmen dan mengunci respons setelah pengiriman.
FR3	Sistem melakukan penilaian otomatis untuk pertanyaan bertipe teori dan profil (berbasis aturan / <i>rule-based</i>).
FR4	Sistem melakukan penilaian otomatis untuk jawaban esai menggunakan model <i>Generative Pre-trained Transformer</i> (GPT) berbasis rubrik.
FR5	Sistem menghasilkan rekomendasi kursus menggunakan pendekatan <i>retrieval-augmented generation</i> (RAG).
FR6	Sistem menggabungkan hasil penilaian dan rekomendasi menjadi ringkasan hasil asesmen.
FR7	Sistem menyediakan titik akses layanan / <i>endpoint</i> bagi pengguna untuk memantau status pemrosesan asesmen.
FR8	Sistem mengirim ringkasan hasil melalui email setelah hasil asesmen siap.
FR9	Sistem menyediakan titik akses layanan bagi pengguna untuk mengirim umpan balik terhadap hasil rekomendasi.

3.3.1.3 Analisis Kebutuhan Non Fungsional

Kebutuhan nonfungsional menjelaskan karakteristik kualitas yang direncanakan pada sistem MicroCred GenAI Backend. Kebutuhan ini digunakan sebagai acuan dalam perancangan sistem agar layanan yang dikembangkan memiliki arah kualitas yang jelas, terutama pada aspek kinerja, keandalan, keamanan, kemudahan pemantauan, konfigurasi, dan integritas data. Rincian kebutuhan nonfungsional disajikan pada Tabel 3.3.

Tabel 3. 3 Kebutuhan non fungsional sistem

Kategori	Target/tujuan
<i>Performance</i>	Sistem dirancang agar proses komputasi berat seperti penilaian GPT, RAG, dan penggabungan hasil dapat diproses secara asinkron, sehingga proses pengiriman asesmen tetap responsif bagi pengguna.
<i>Reliability</i>	Sistem direncanakan memiliki mekanisme percobaan ulang pada proses yang berpotensi mengalami kegagalan sementara, terutama saat berinteraksi dengan layanan eksternal.
<i>Scalability</i>	Sistem dirancang agar alur pemrosesan internal dapat dikembangkan untuk mendukung pemrosesan latar atau pemrosesan terpisah apabila jumlah asesmen meningkat.
<i>Security</i>	Sistem direncanakan menggunakan autentikasi dan otorisasi untuk membatasi akses terhadap endpoint atau fungsi yang bersifat sensitif.
<i>Observability</i>	Sistem direncanakan agar parameter penting, seperti model AI, batas waktu pemrosesan, konfigurasi layanan eksternal, dan pengaturan pemrosesan pekerjaan, dapat diatur melalui konfigurasi lingkungan.
<i>Configurability</i>	Sistem direncanakan agar parameter penting, seperti model AI, batas waktu pemrosesan, konfigurasi layanan eksternal, dan pengaturan pemrosesan pekerjaan, dapat diatur melalui konfigurasi lingkungan.
<i>Data Integrity</i>	Sistem dirancang agar memiliki mekanisme pencegahan duplikasi data dan menjaga konsistensi hasil asesmen, terutama pada proses pengiriman jawaban dan penyimpanan skor.

3.3.2 Perancangan Purwarupa / *Prototype*

Bagian ini menyajikan rancangan solusi berdasarkan kebutuhan yang telah didefinisikan pada Bagian 3.3.1. Rancangan mencakup arsitektur sistem, alur proses, skema basis data, dan pilihan teknologi.

3.3.2.1 Arsitektur Sistem

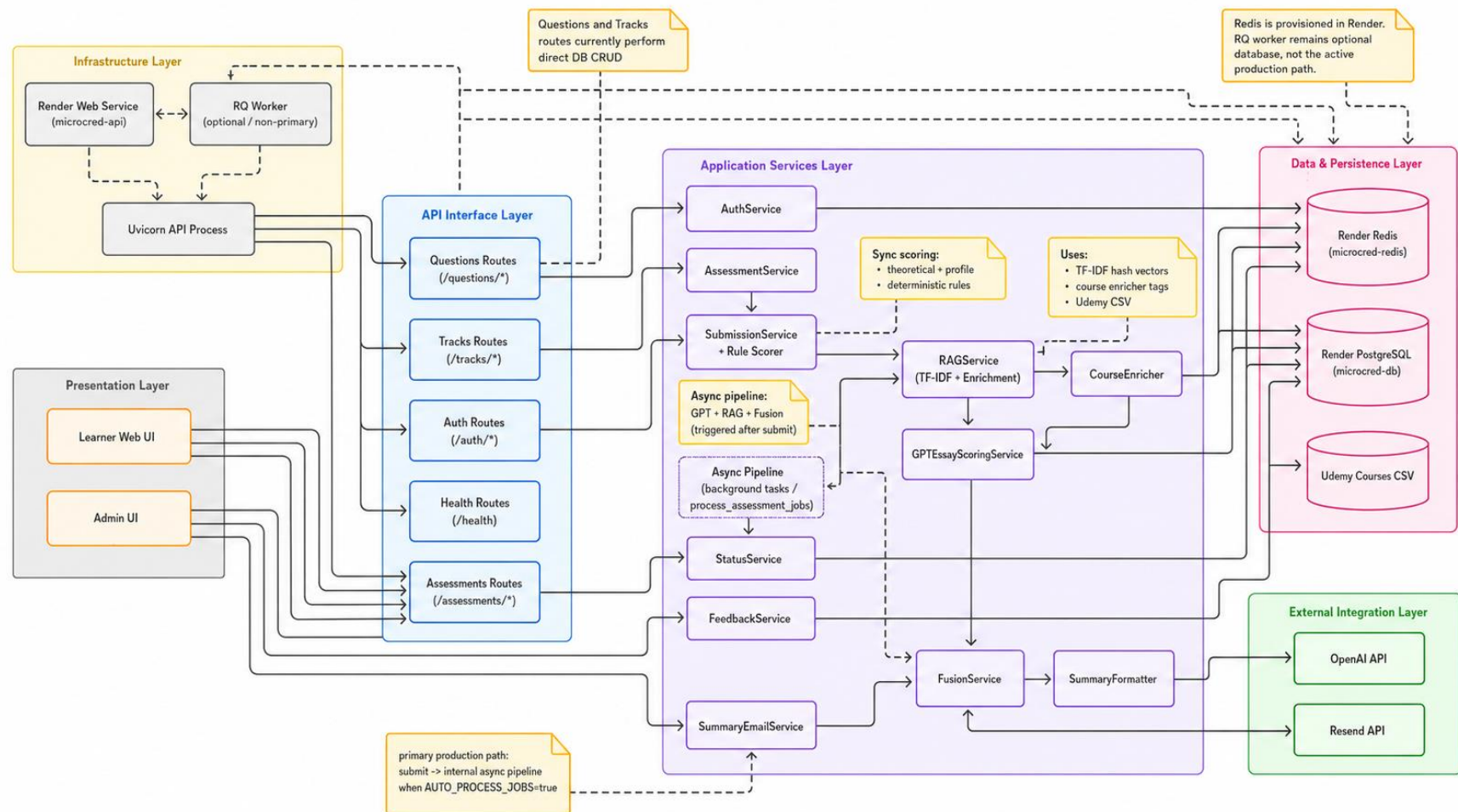
MicroCred GenAI Backend dirancang sebagai satu aplikasi backend terpadu yang tetap membagi tanggung jawab sistem ke dalam beberapa modul yang terpisah secara logis. Setiap modul menangani fungsi tertentu, seperti autentikasi, pengelolaan pengguna, pemrosesan microcredential, serta integrasi dengan layanan GenAI. Dengan pemisahan ini, struktur kode menjadi lebih terorganisasi, mudah dipahami, dan lebih mudah dikembangkan, meskipun

seluruh modul masih berjalan dalam satu basis aplikasi yang sama. Pendekatan ini dikenal sebagai arsitektur *modular monolith*. Arsitektur ini dipilih karena mampu menjaga batas antar-modul tetap jelas tanpa menambah kompleksitas koordinasi antar-layanan sejak awal, sehingga sesuai dengan ruang lingkup MVP dalam penelitian ini:

Arsitektur sistem terdiri atas enam lapisan logis, sebagaimana ditunjukkan pada Gambar 3.1, yaitu:

1. Lapisan presentasi atau presentation layer, yaitu antarmuka pengguna seperti *Learner Web UI* dan *Admin UI* yang mengakses layanan *backend* melalui HTTP.
2. Lapisan antarmuka API atau *API interface layer*, yang mencakup *route* FastAPI, validasi *request*, autentikasi, dan otorisasi.
3. Lapisan layanan aplikasi atau *application services layer*, yang berisi layanan domain seperti *AssessmentService*, *RAGService*, *FusionService*, dan layanan terkait lainnya.
4. Lapisan data dan persistensi atau data and persistence layer, yang mencakup PostgreSQL sebagai penyimpanan utama, Redis sebagai infrastruktur pendukung, serta himpunan data kursus dalam format CSV.
5. Lapisan integrasi eksternal atau *external integration layer*, yang menangani integrasi dengan API OpenAI dan layanan *email Resend*.
6. Lapisan infrastruktur atau *infrastructure layer*, yaitu lingkungan runtime aplikasi pada Render Web Service.

Gambar 3.2 menunjukkan arsitektur *modular monolith* MicroCred GenAI Backend. Perbedaan warna pada gambar digunakan untuk membedakan kelompok komponen sistem. Warna kuning menunjukkan lapisan infrastruktur, abu-abu menunjukkan lapisan presentasi, biru menunjukkan lapisan antarmuka API, ungu menunjukkan lapisan layanan aplikasi, merah muda menunjukkan lapisan data dan persistensi, sedangkan hijau menunjukkan integrasi dengan layanan eksternal. Pembedaan warna ini digunakan agar pembaca dapat memahami batas tanggung jawab setiap lapisan sistem secara lebih jelas.

Gambar 3. 2 Arsitektur *modular monolith* MicroCred GenAI

3.3.2.2 Alur Pemrosesan Asesmen

Alur pemrosesan asesmen dari awal hingga hasil tersedia digambarkan pada Gambar 3.4. Pada gambar tersebut, perbedaan warna notasi digunakan untuk membedakan jenis aktivitas pada alur sistem. Notasi berwarna biru menunjukkan interaksi pengguna dan titik akses layanan, notasi berwarna abu-abu menunjukkan proses sinkron utama pada sistem, notasi berwarna jingga menunjukkan proses asinkron yang dijalankan di belakang layar, sedangkan notasi berwarna hijau menunjukkan hasil akhir yang telah tersedia bagi pengguna.

Pada jalur sinkron, pengguna memulai asesmen dan memperoleh salinan sesaat pertanyaan / *question snapshot* sesuai peran yang dipilih. Setelah pengguna mengirim jawaban, sistem memvalidasi kepemilikan asesmen, menyimpan respons, dan menjalankan penilaian berbasis aturan / *rule-based scoring* untuk pertanyaan teori dan profil. Selanjutnya, sistem membentuk tiga pekerjaan asinkron / *job* (GPT, RAG, dan penggabungan hasil / *fusion*) dan segera mengembalikan respons awal kepada pengguna.

Pada jalur asinkron, ketiga pekerjaan tersebut dieksekusi secara berurutan. Tahap GPT digunakan untuk menilai jawaban esai berdasarkan rubrik. Tahap RAG digunakan untuk mengambil dan mengurutkan kandidat kursus dari himpunan data menggunakan pemeringkatan hibrida ringan. Setelah itu, tahap *fusion* menggabungkan hasil penilaian dan rekomendasi menjadi ringkasan akhir yang dapat ditampilkan kepada pengguna.

Secara operasional, skor akhir kandidat kursus pada tahap RAG dihitung melalui kombinasi pencocokan metadata yang diperkaya, relevansi kata kunci, kemiripan vektor token berbasis fungsi *hashing*, *tag boost*, dan skor kualitas kursus. Formula implementasinya ditunjukkan pada Gambar 3.3.

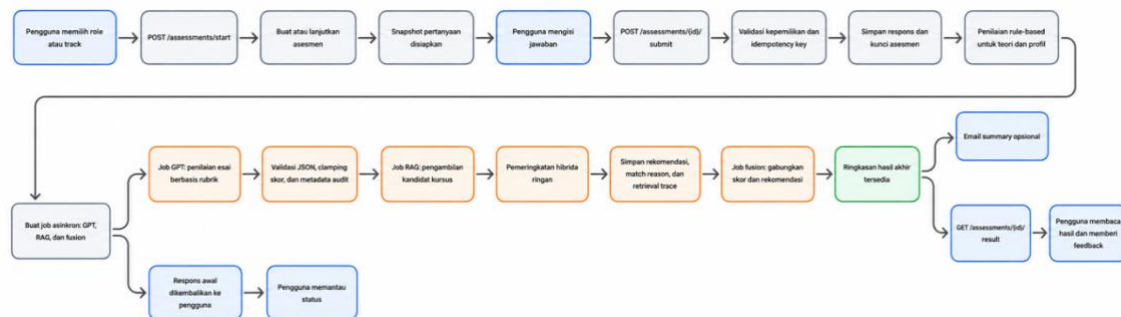
```
final_score = (
    match_score * 0.5
    + keyword_score * 0.2
    + embedding_score * 0.1
    + tag_boost
    + quality_score * 0.2
)
```

Gambar 3. 3 Formula perhitungan skor akhir kandidat kursus pada proses RAG berbasis pemeringkatan hibrida ringan

Pada formula tersebut, *match_score* menunjukkan pencocokan metadata yang

diperkaya, *keyword_score* menunjukkan relevansi kata kunci, *embedding_score* menunjukkan kemiripan vektor token berbasis fungsi *hashing*, *tag_boost* menunjukkan penguatan berdasarkan tag atau topik, dan *quality_score* menunjukkan skor kualitas kursus. Istilah *embedding_score* tidak merujuk pada representasi vektor neural / *neural embedding* eksternal, tetapi pada representasi vektor token berbasis fungsi *hashing* yang dibuat secara lokal oleh sistem.

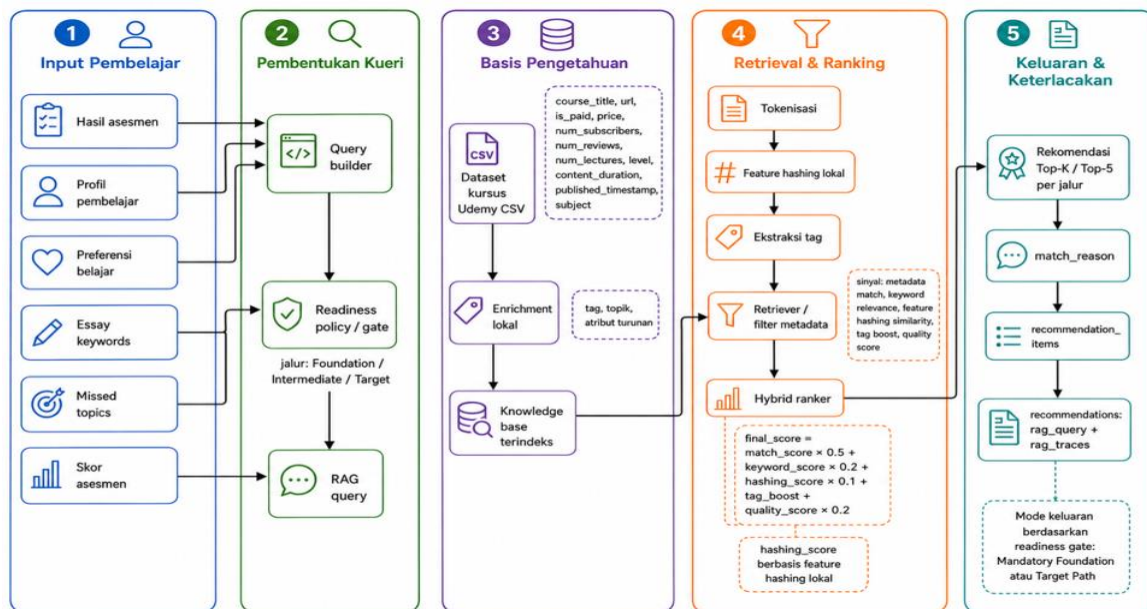
Diagram alur pemrosesan asesmen MicroCred GenAI ditunjukkan pada Gambar 3.4.



Gambar 3. 4 Diagram alur pemrosesan asesmen MicroCred GenAI

3.3.2.3 Desain RAG dan Basis Pengetahuan

Selain alur pemrosesan asesmen secara umum, penelitian ini juga merancang mekanisme *retrieval-augmented generation* (RAG) sebagai dasar penyusunan rekomendasi kursus. Pada penelitian ini, RAG dipahami sebagai alur pengambilan dan penyusunan rekomendasi yang berlandaskan pada dataset kursus lokal, bukan sebagai proses generasi bebas tanpa sumber. Mekanisme ini mencakup pembentukan kueri, penentuan kesiapan pembelajar, pemanfaatan basis pengetahuan kursus, pengambilan kandidat, pemeringkatan hibrida, penyusunan alasan rekomendasi, dan penyimpanan jejak pengambilan. Rancangan umum mekanisme RAG ditunjukkan pada Gambar 3.5.



Gambar 3. 5 Desain RAG untuk rekomendasi kursus kredensial mikro pada MicroCred GenAI Backend

Berdasarkan Gambar 3.5, proses RAG dimulai dari hasil asesmen, profil pembelajar, preferensi belajar, kata kunci esai, topik yang belum dikuasai, dan skor asesmen. Informasi tersebut digunakan oleh *query builder* untuk membentuk kueri rekomendasi. Selain itu, sistem menggunakan *readiness policy* atau *readiness gate* untuk menentukan arah rekomendasi berdasarkan tingkat kesiapan pembelajar. Dengan mekanisme tersebut, pengguna dapat diarahkan ke jalur fondasi, jalur menengah, atau jalur target sesuai hasil asesmen yang diperoleh.

Basis pengetahuan yang digunakan dalam penelitian ini berasal dari dataset kursus UdeMy yang tersedia pada Kaggle. Dataset tersebut digunakan sebagai sumber data lokal berbentuk CSV dan kemudian disiapkan sebagai basis pengetahuan operasional sistem. Field asli pada dataset meliputi *course_title*, *url*, *is_paid*, *price*, *num_subscribers*, *num_reviews*, *num_lectures*, *level*, *content_duration*, *published_timestamp*, dan *subject*. Field tersebut digunakan sebagai dasar pencocokan awal karena merepresentasikan informasi utama kursus, seperti judul, tautan kursus, status berbayar atau gratis, harga, jumlah pelanggan, jumlah ulasan, jumlah materi, tingkat kesulitan, durasi, waktu publikasi, dan kategori/subjek kursus.

Selain field asli dari dataset, sistem juga melakukan pengayaan data / *enrichment* untuk membentuk informasi tambahan yang tidak tersedia secara langsung pada CSV, seperti tag, topik, dan atribut turunan lain yang digunakan untuk mendukung proses pencocokan rekomendasi. Dengan demikian, basis pengetahuan pada penelitian ini terdiri atas dua bagian,

yaitu data asli dari dataset kursus dan hasil pengayaan yang dibentuk oleh sistem untuk kebutuhan rekomendasi.

Pada tahap pemrosesan, data dari kueri pengguna dan data kursus diproses melalui tokenisasi, ekstraksi tag, dan *feature hashing*. Sistem tidak menggunakan model *embedding* eksternal. Oleh karena itu, istilah *embedding_score* pada penelitian ini tidak merujuk pada *neural embedding* dari model eksternal, melainkan pada skor kemiripan representasi token berbasis *feature hashing* lokal.

Setelah kueri dan basis pengetahuan siap, sistem melakukan proses pengambilan kandidat melalui *retriever* dan penyaringan metadata. Kandidat kursus kemudian diurutkan menggunakan *hybrid ranker* yang menggabungkan beberapa komponen skor, yaitu *match_score*, *keyword_score*, *embedding_score*, *tag_boost*, dan *quality_score*. Hasil akhir proses ini berupa rekomendasi Top-K atau Top-5 per jalur rekomendasi, disertai *match_reason* agar alasan kecocokan kursus dapat dipahami oleh pengguna. Pada mode tertentu, rekomendasi dapat disusun dalam dua jalur, yaitu *Mandatory Foundation* dan *Target Path*, sehingga penyebutan Top-K digunakan agar tetap sesuai dengan variasi keluaran sistem.

Hasil rekomendasi disimpan pada tabel *recommendation_items*. Sementara itu, informasi *rag_query* dan *rag_traces* disimpan pada tabel *recommendations*. Penyimpanan ini digunakan untuk mendukung keterlacakan hasil rekomendasi, sehingga dasar pengambilan dan penyusunan rekomendasi dapat diperiksa kembali.

Ringkasan komponen utama rancangan RAG disajikan pada Tabel 3.4.

Tabel 3. 4 Komponen utama desain RAG

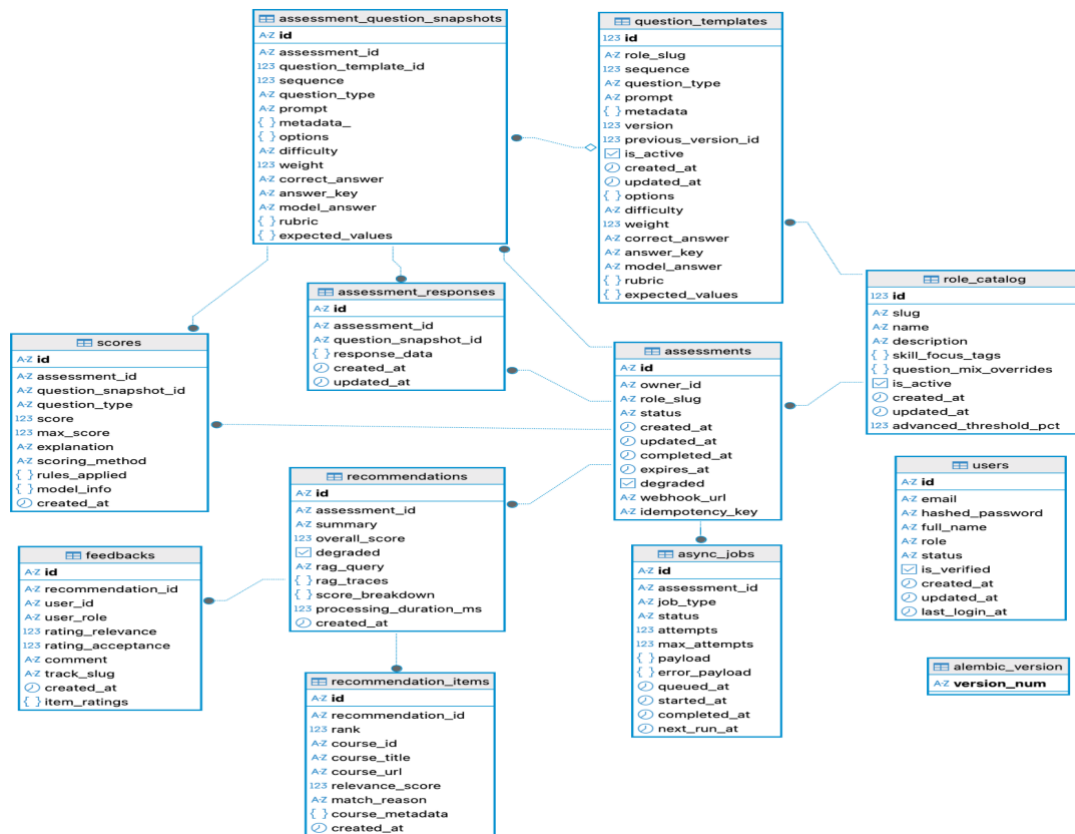
Kategori	Target/tujuan
Hasil asesmen dan profil pembelajar	Menjadi masukan awal untuk memahami kemampuan pengguna, preferensi belajar, kata kunci dari jawaban esai, topik yang belum dikuasai, serta skor hasil asesmen.
Query builder	Membentuk kueri rekomendasi dengan memanfaatkan hasil asesmen, profil pembelajar, preferensi belajar, <i>essay keywords</i> , dan <i>missed topics</i> .
Readiness policy / readiness gate	Menentukan arah rekomendasi berdasarkan tingkat kesiapan pembelajar, misalnya ke jalur fondasi, jalur menengah, atau jalur target sesuai kebutuhan pengguna.
Dataset kursus Udemy CSV	Menjadi sumber data utama dalam proses rekomendasi. Dataset yang digunakan berasal dari Udemy Courses Dataset di Kaggle, dengan field utama seperti <i>course_title</i> , <i>url</i> , <i>is_paid</i> , <i>price</i> , <i>num_subscribers</i> , <i>num_reviews</i> , <i>num_lectures</i> , <i>level</i> , <i>content_duration</i> , <i>published_timestamp</i> , dan <i>subject</i> .
Enriched knowledge base	Menyimpan hasil pengayaan dari dataset kursus, seperti tag, topik, dan atribut turunan lain yang digunakan untuk membantu proses pencocokan rekomendasi.
Tokenisasi dan feature hashing	Mengubah teks dari kueri dan informasi kursus menjadi representasi token lokal. Pada rancangan ini, proses representasi teks tidak menggunakan model

Kategori	Target/tujuan
	embedding eksternal, melainkan pendekatan tokenisasi dan <i>feature hashing</i> secara lokal.
Ekstraksi tag dan topik	Menghasilkan informasi tambahan dari data kursus maupun masukan pengguna untuk membantu mencocokkan kebutuhan pembelajar dengan kandidat kursus yang tersedia.
Retriever / filter metadata	Mengambil kandidat kursus awal berdasarkan kecocokan kueri, metadata kursus, tag, topik, level, durasi, harga atau status berbayar, jumlah pelanggan, jumlah ulasan, jumlah materi, serta waktu publikasi.
Hybrid ranker	Mengurutkan kandidat kursus menggunakan kombinasi beberapa skor, yaitu <i>match_score</i> , <i>keyword_score</i> , <i>embedding_score</i> , <i>tag_boost</i> , dan <i>quality_score</i> . Dalam konteks ini, <i>embedding_score</i> merujuk pada skor kemiripan berbasis representasi fitur lokal, bukan embedding dari model eksternal.
Recommendation output	Menghasilkan rekomendasi Top-K atau Top-5 sesuai jalur rekomendasi, disertai alasan kecocokan atau <i>match_reason</i> untuk menjelaskan mengapa kursus tersebut direkomendasikan.
Penyimpanan hasil dan jejak	Menyimpan hasil rekomendasi pada tabel <i>recommendation_items</i> , sedangkan <i>rag_query</i> dan <i>rag_traces</i> disimpan pada tabel <i>recommendations</i> sebagai jejak proses rekomendasi.

Berdasarkan Tabel 3.4, desain RAG pada penelitian ini dirancang agar tetap ringan dan sesuai dengan ruang lingkup *minimum viable product* (MVP). Sistem tidak bergantung pada model *embedding* eksternal, tetapi menggunakan tokenisasi, ekstraksi tag, dan *feature hashing* lokal untuk mendukung pencocokan kandidat kursus. Selain itu, keberadaan *readiness policy*, *match_reason*, dan *rag_traces* membantu sistem tidak hanya menghasilkan rekomendasi, tetapi juga menyediakan dasar yang dapat ditelusuri.

3.3.2.4 Skema Basis Data

Struktur basis data dirancang untuk mendukung proses asesmen, penilaian, rekomendasi, pelacakan pekerjaan asinkron / *job*, dan pengelolaan umpan balik. Diagram hubungan entitas / *entity relationship diagram* (ERD) sistem ditunjukkan pada Gambar 3.6, sedangkan ringkasan entitas utama disajikan pada Tabel 3.5.



Gambar 3. 6 ERD sistem MicroCred GenAI Backend

Gambar 3.6 menampilkan ERD yang merepresentasikan struktur basis data sistem, yang dirancang untuk mendukung autentikasi pengguna, pelaksanaan assessment, penilaian, rekomendasi, pelacakan proses asinkron, serta pengelolaan feedback pengguna.

Untuk memperjelas peran masing-masing tabel utama, ringkasan tabel basis data disajikan pada Tabel 3.5.

Tabel 3. 5 Ringkasan entitas basis data utama

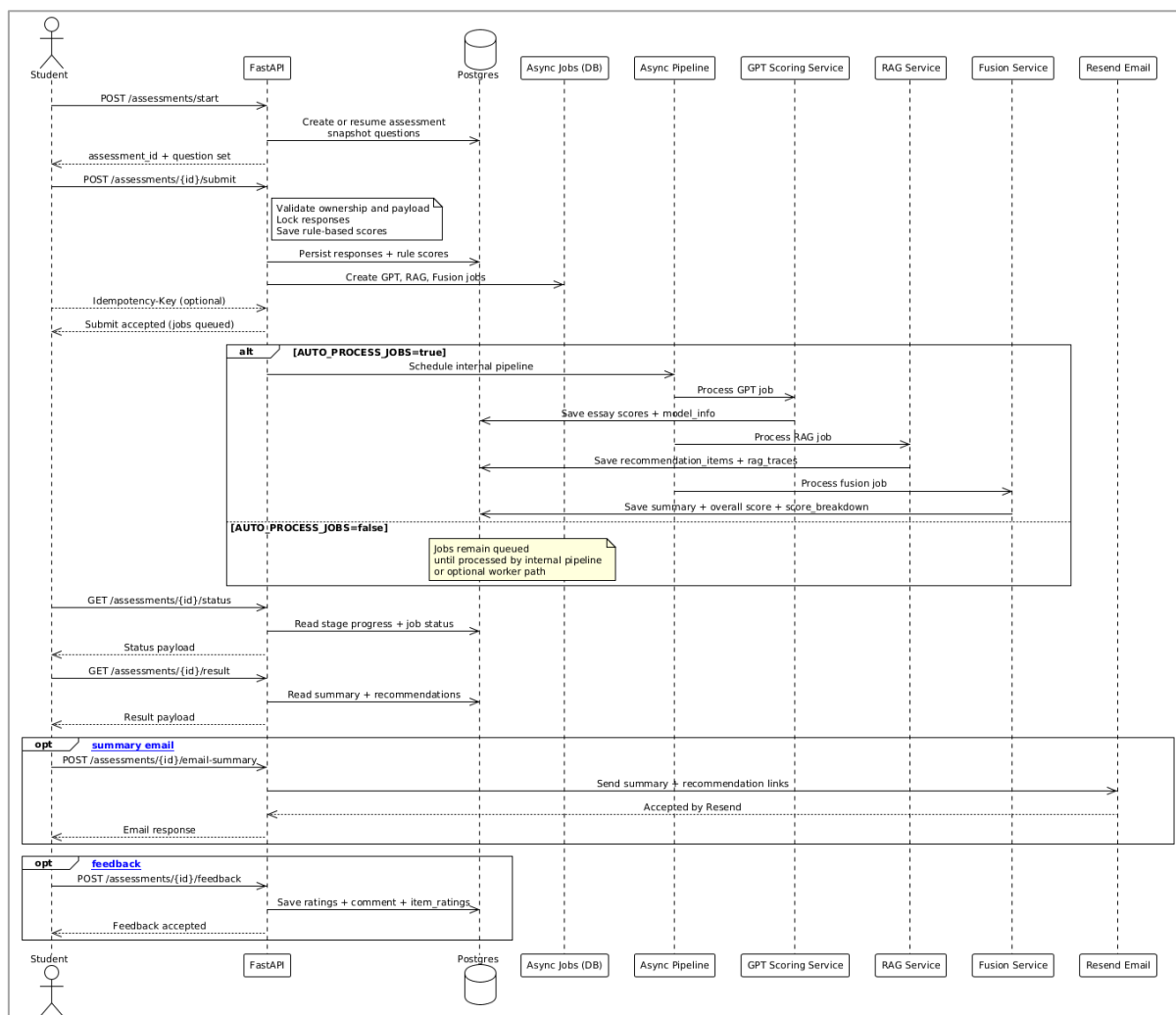
Entitas	Fungsi
<i>users</i>	Menyimpan akun pengguna untuk autentikasi dan otorisasi.
<i>role_catalog</i>	Menyimpan katalog peran atau jalur asesmen dan konfigurasi fokus <i>skill</i> .
<i>question_templates</i>	Menyimpan bank pertanyaan per peran atau jalur beserta rubrik dan versi.
<i>assessments</i>	Menyimpan entitas asesmen dan status siklus hidupnya.
<i>assessment question snapshots</i>	Menjaga konsistensi pertanyaan sejak asesmen dimulai.
<i>assessment responses</i>	Menyimpan jawaban pengguna.
<i>scores</i>	Menyimpan hasil penilaian per pertanyaan (berbasis aturan / <i>rule-based</i> dan GPT).
<i>async_jobs</i>	Melacak pemrosesan pekerjaan asinkron / <i>job</i> .
<i>recommendations</i>	Menyimpan hasil penggabungan / <i>fusion</i> , ringkasan, dan metadata rekomendasi.

Entitas	Fungsi
<i>recommendation_items</i>	Menyimpan butir rekomendasi beserta alasan pemilihan.
<i>feedbacks</i>	Menyimpan umpan balik pengguna terhadap rekomendasi.

3.3.2.5 Diagram Aliran Data

Diagram aliran data / *data flow diagram* (DFD) pada Gambar 3.7 memperlihatkan perpindahan data antar-komponen sistem selama siklus *request-response*, mencakup alur penyimpanan, pembacaan, dan pertukaran data dengan layanan eksternal.

Untuk memperjelas perpindahan data antarproses, diagram aliran data sistem ditunjukkan pada Gambar 3.7.



Gambar 3. 7 Diagram aliran data / data flow diagram (DFD)

Berdasarkan Gambar 3.7, data pengguna mengalir dari antarmuka menuju backend, kemudian diproses melalui layanan asesmen, GPT, RAG, dan penyimpanan hasil.

3.3.2.6 Rancangan Penilaian Esai Berbasis GPT

Penilaian esai pada penelitian ini dirancang menggunakan model *Generative Pre-trained Transformer* (GPT) yang dipanggil melalui antarmuka pemrograman aplikasi / *Application Programming Interface* (API) Chat Completions. Pemilihan GPT didasarkan pada karakteristik jawaban esai yang berbentuk bahasa alami, sehingga penilaian tidak cukup dilakukan hanya melalui pencocokan kata kunci sederhana. Dalam rancangan ini, GPT digunakan untuk membaca jawaban pengguna, membandingkannya dengan konteks pertanyaan dan rubrik, lalu menghasilkan skor per dimensi beserta penjelasan singkat.

Penggunaan model GPT dipilih untuk tahap *minimum viable product* (MVP) karena mendukung instruksi berbasis *prompt*, dapat dipanggil melalui API, dan memungkinkan integrasi dengan proses *backend* secara asinkron. Pemilihan ini tidak dimaksudkan untuk menyatakan bahwa model tertentu merupakan satu-satunya atau model terbaik untuk penilaian esai, tetapi sebagai pilihan rancangan yang sesuai dengan ruang lingkup penelitian, kebutuhan integrasi cepat, dan keterbatasan waktu pengembangan. Agar rancangan tetap fleksibel, pemilihan model direncanakan dapat dikonfigurasi melalui variabel lingkungan, sehingga model dapat diganti tanpa mengubah logika utama sistem.

Prompt penilaian esai dirancang terdiri atas dua bagian utama, yaitu *system prompt* dan *user prompt*. *System prompt* berisi instruksi tetap untuk mengatur peran dan batasan model, sedangkan *user prompt* berisi data penilaian yang berubah sesuai pertanyaan dan jawaban pengguna. Rincian rancangan komponen *prompt* penilaian esai ditunjukkan pada Tabel 3.6.

Tabel 3. 6 Rancangan komponen prompt penilaian esai

Komponen Prompt	Posisi dalam Rancangan	Isi Rancangan	Tujuan
<i>System prompt</i>	<i>System message</i>	Instruksi berbahasa Inggris yang menetapkan model sebagai evaluator esai yang ketat / <i>strict essay evaluator</i> pada platform asesmen kredensial mikro.	Membatasi peran model sebagai evaluator, bukan sebagai tutor atau pemberi saran belajar.
Batasan evaluasi	<i>System message</i>	Model hanya diperbolehkan menggunakan informasi yang tersedia dalam prompt, yaitu pertanyaan, jawaban pengguna, dan jawaban acuan apabila tersedia. Model tidak diperbolehkan menggunakan pengetahuan eksternal, memberikan saran, menggunakan format <i>markdown</i> , atau menghasilkan teks bebas di luar struktur JSON.	Mengurangi risiko penilaian bebas, menjaga konsistensi evaluasi, dan memastikan keluaran tetap terstruktur.

Dimensi rubrik	<i>System message</i>	Dimensi penilaian terdiri atas <i>relevance</i> , <i>depth</i> , <i>clarity</i> , <i>completeness</i> , dan <i>technical</i> , dengan masing-masing dimensi berada pada rentang skor 0–100.	Menyeragamkan dasar penilaian esai berdasarkan dimensi rubrik yang terdokumentasi.
Kontrak keluaran JSON	<i>System message</i>	Keluaran model wajib berupa satu objek JSON yang memuat <i>scores</i> , <i>total_score</i> , dan <i>explanation</i> .	Memudahkan proses <i>parsing</i> , validasi struktur keluaran, dan penghitungan ulang skor oleh <i>backend</i> .
Konteks pertanyaan	<i>User prompt</i>	Berisi teks pertanyaan atau tugas esai yang harus dijawab oleh pengguna.	Memberikan konteks penilaian agar model mengevaluasi jawaban sesuai maksud pertanyaan.
Jawaban pengguna	<i>User prompt</i>	Berisi jawaban esai yang dikirimkan oleh pengguna.	Menjadi objek utama yang dinilai oleh model berdasarkan rubrik.
Jawaban acuan	<i>User prompt</i>	Berisi jawaban rujukan apabila tersedia pada data pertanyaan.	Membantu model menyelaraskan penilaian dengan konsep, kata kunci, atau poin penting yang diharapkan.
Rubrik penilaian	<i>User prompt</i>	Berisi daftar dimensi rubrik dan bobot penilaian yang digunakan dalam proses evaluasi.	Menjaga penilaian tetap mengacu pada kriteria yang jelas dan terdokumentasi.
Tingkat kesulitan	Diproses oleh <i>backend</i> dan memengaruhi rubrik	Tingkat kesulitan, seperti <i>easy</i> , <i>medium</i> , atau <i>hard</i> , digunakan untuk menentukan bobot rubrik, batas bawah / <i>floor</i> , dan batas atas / <i>ceiling</i> penilaian.	Menyesuaikan fokus dan rentang penilaian dengan karakteristik serta tingkat kompleksitas pertanyaan.

Berdasarkan Tabel 3.6, *system prompt* dan *user prompt* memiliki fungsi yang berbeda. *System prompt* mengatur model agar bertindak sebagai evaluator yang ketat, sedangkan *user prompt* membawa data penilaian spesifik, seperti pertanyaan, jawaban pengguna, jawaban acuan, dan bobot rubrik. Dengan pemisahan ini, instruksi evaluator dapat dijaga tetap konsisten, sementara data yang dinilai dapat berubah sesuai setiap asesmen.

Selain rancangan *prompt*, penelitian ini juga menetapkan parameter pemanggilan model agar proses penilaian lebih terkendali. Parameter tersebut digunakan sebagai acuan agar keluaran model lebih konsisten, proses tidak berjalan terlalu lama, dan respons yang dihasilkan dapat divalidasi oleh *backend*. Rancangan parameter pemanggilan model ditunjukkan pada Tabel 3.7.

Tabel 3. 7 Parameter pemanggilan model GPT

Parameter	Penggunaan aktual	Catatan
<i>model</i>	Model GPT dirancang dapat diatur melalui variabel lingkungan, sehingga dapat diganti tanpa mengubah kode utama sistem.	Menjaga fleksibilitas pemilihan model pada tahap pengembangan dan evaluasi.
<i>base_url</i>	Endpoint API dirancang dapat dikonfigurasi agar kompatibel dengan layanan Chat Completions.	Memudahkan integrasi dengan layanan pemrosesan bahasa alami berbasis API.
<i>messages</i>	Berisi <i>system prompt</i> dan <i>user prompt</i> .	Memisahkan instruksi tetap evaluator dan data penilaian yang berubah pada setiap asesmen.
<i>temperature</i>	Nilai temperatur dirancang rendah untuk menjaga konsistensi penilaian.	Mengurangi variasi keluaran model pada tugas penilaian berbasis rubrik.
<i>max_tokens</i>	Jumlah token keluaran dibatasi agar respons tetap ringkas dan sesuai kebutuhan penilaian.	Mencegah keluaran terlalu panjang dan membantu menjaga efisiensi pemrosesan.
<i>response_format</i>	Keluaran diarahkan dalam format JSON dan divalidasi kembali oleh <i>backend</i> .	Memastikan keluaran dapat diproses secara terstruktur oleh sistem.
<i>timeout</i>	Proses pemanggilan model dirancang memiliki batas waktu..	Mencegah proses penilaian menggantung terlalu lama apabila terjadi gangguan layanan.
<i>Retry API / jaringan</i>	Sistem dirancang memiliki mekanisme percobaan ulang pada kegagalan sementara.	Meningkatkan ketahanan proses terhadap <i>timeout</i> , <i>rate limit</i> , atau gangguan jaringan.
<i>Retry parsing</i>	Sistem dirancang dapat meminta ulang keluaran apabila respons awal tidak	Mengurangi risiko kegagalan penilaian akibat

	dapat diproses sebagai JSON yang valid.	format keluaran yang tidak sesuai.
--	---	------------------------------------

Keluaran model tidak langsung digunakan sebagai nilai akhir. Backend terlebih dahulu memvalidasi struktur JSON, memeriksa kelengkapan field, dan membatasi skor setiap dimensi ke rentang 0–100 melalui mekanisme *clamping*. Setelah itu, backend menghitung skor akhir secara deterministik berdasarkan bobot rubrik yang disesuaikan dengan tingkat kesulitan pertanyaan. Bobot rubrik tersebut ditunjukkan pada Tabel 3.8.

Tabel 3. 8 Bobot rubrik berdasarkan tingkat kesulitan

Tingkat kesulitan	Relevansi	Kejelasan	Kelengkapan	Kedalaman	Teknis
Default	0,20	0,20	0,20	0,20	0,20
<i>Easy</i>	0,30	0,25	0,25	0,10	0,10
<i>Medium</i>	0,25	0,20	0,20	0,20	0,15
<i>Hard</i>	0,20	0,15	0,15	0,25	0,25

Nilai bobot pada Tabel 3.8 tidak diambil sebagai angka baku dari satu sumber eksternal, tetapi ditetapkan sebagai bagian dari rancangan rubrik pada implementasi MicroCred GenAI Backend. Penetapan bobot tersebut didasarkan pada prinsip penilaian berbasis rubrik, yaitu setiap jawaban dinilai melalui beberapa dimensi yang jelas, lalu bobot tiap dimensi dapat disesuaikan dengan karakteristik tugas. Pada penelitian ini, tingkat kesulitan digunakan untuk membedakan fokus penilaian. Pertanyaan *easy* lebih menekankan relevansi, kejelasan, dan kelengkapan jawaban, sedangkan pertanyaan *hard* lebih menekankan kedalaman dan aspek teknis. Dengan demikian, bobot pada Tabel 3.8 merupakan keputusan desain sistem yang digunakan agar penilaian esai lebih sesuai dengan tingkat kesulitan pertanyaan.

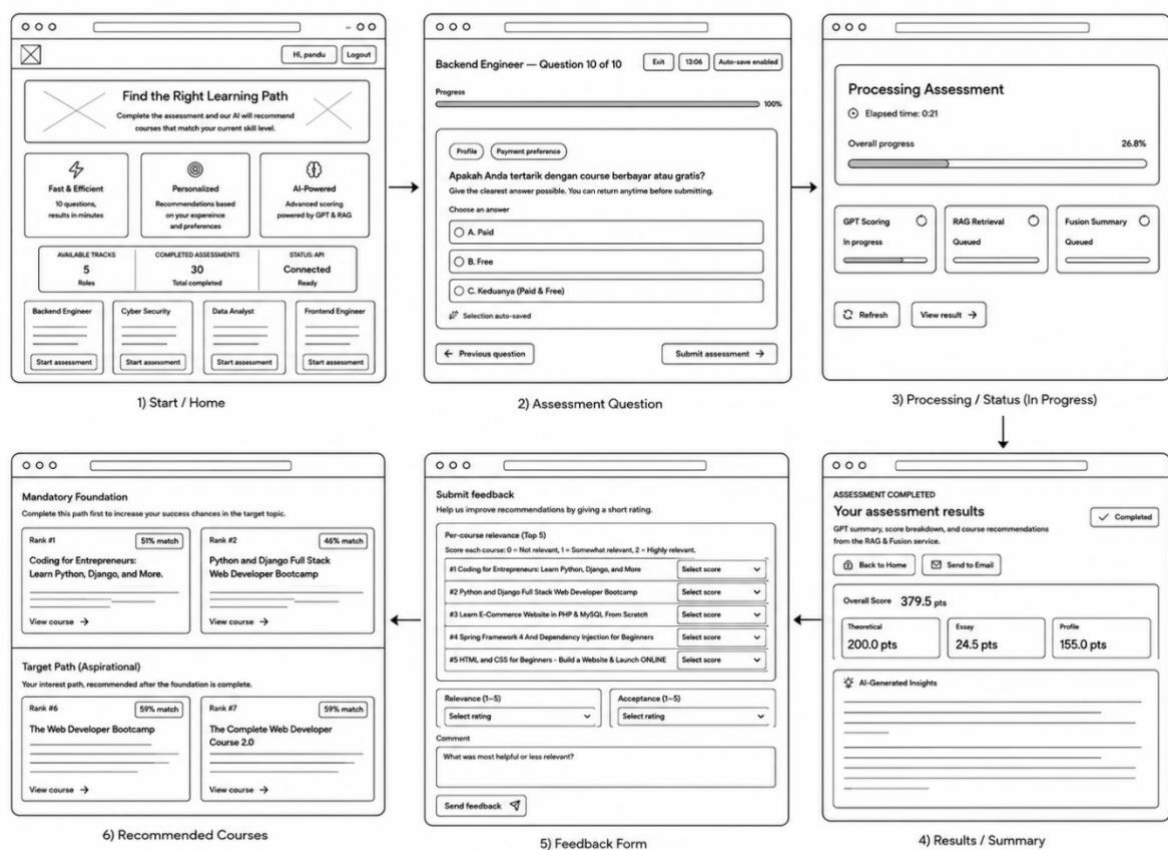
Setelah skor berbobot dihitung, sistem menerapkan batas bawah / *floor* dan batas atas / *ceiling*. Pada konfigurasi aktual, *ceiling* default dan berbasis tingkat kesulitan adalah 95. Batas bawah untuk default dan *hard* adalah 0, untuk *medium* adalah 5, dan untuk *easy* adalah 10. Penerapan *floor* dan *ceiling* digunakan untuk menjaga agar hasil akhir tetap berada dalam rentang yang sesuai dengan rancangan penilaian. Dengan demikian, GPT berperan sebagai evaluator awal berbasis rubrik, sedangkan nilai akhir tetap ditentukan oleh logika backend yang lebih terstruktur dan dapat diaudit.

3.3.2.7 Rancangan Antarmuka Pengguna / *User Interface*

Meskipun penelitian ini berfokus pada pengembangan layanan sisi peladen / *backend*, rancangan antarmuka pengguna tetap disusun untuk menggambarkan bagaimana pengguna

berinteraksi dengan alur asesmen dan rekomendasi. Rancangan UI pada penelitian ini tidak diarahkan sebagai fokus utama evaluasi, tetapi digunakan sebagai gambaran alur penggunaan sistem, mulai dari memulai asesmen, menjawab pertanyaan, menunggu pemrosesan, hingga membaca hasil rekomendasi.

Rancangan antarmuka pengguna dibuat sederhana agar dapat mendukung kebutuhan uji coba MVP. Antarmuka dirancang untuk menampilkan informasi utama yang dibutuhkan pengguna, yaitu instruksi asesmen, pertanyaan, status pemrosesan, skor hasil asesmen, rekomendasi kursus, alasan rekomendasi, dan akses umpan balik. Untuk memperjelas alur interaksi pengguna dengan sistem, rancangan antarmuka pengguna divisualisasikan dalam bentuk wireframe seperti ditunjukkan pada Gambar 3.8.



Gambar 3. 8 Wireframe rancangan antarmuka pengguna MicroCred GenAI

Berdasarkan Gambar 3.8, alur antarmuka pengguna dimulai dari halaman mulai asesmen yang berfungsi memberikan pengantar dan menyiapkan sesi asesmen. Pengguna kemudian mengerjakan pertanyaan teori, esai, dan profil pembelajar pada halaman pengerjaan soal. Setelah jawaban dikirim, sistem menampilkan status pemrosesan karena proses GPT, RAG,

dan fusion berjalan secara asinkron. Hasil akhir ditampilkan melalui halaman hasil asesmen dan halaman rekomendasi kursus yang memuat skor, tingkat kompetensi, daftar rekomendasi, nilai kecocokan, alasan rekomendasi, dan akses terhadap informasi keterlacakan. Pada tahap akhir, pengguna dapat mengirimkan umpan balik terhadap relevansi dan penerimaan rekomendasi.

Wireframe tersebut tidak dimaksudkan sebagai rancangan akhir antarmuka, tetapi sebagai sketsa konseptual untuk menjelaskan kebutuhan tampilan pada tahap MVP. Dengan demikian, rancangan UI berperan sebagai media interaksi antara pengguna dan layanan backend, sedangkan fokus utama penelitian tetap berada pada rancangan asesmen, pemrosesan asinkron, mekanisme RAG, dan keterlacakan rekomendasi. Rincian fungsi setiap halaman antarmuka dan keterkaitannya dengan backend dirangkum pada Tabel 3.9.

Tabel 3. 9 Rancangan antarmuka pengguna MicroCred GenAI

Halaman / Komponen UI	Fungsi Utama	Informasi yang Ditampilkan	Keterkaitan dengan Backend
Halaman mulai asesmen	Memberikan pengantar sebelum pengguna mengerjakan asesmen.	Tujuan asesmen, jumlah pertanyaan, jenis penilaian, dan ringkasan komponen skor.	Mengambil konfigurasi asesmen dan menyiapkan sesi asesmen pengguna.
Halaman pengerjaan asesmen	Menyediakan tempat bagi pengguna untuk menjawab pertanyaan.	Pertanyaan teori, pertanyaan esai, pilihan profil, dan tombol pengiriman jawaban.	Mengirim jawaban pengguna ke endpoint asesmen untuk divalidasi dan diproses.
Status pemrosesan	Menunjukkan bahwa jawaban pengguna sedang diproses.	Status pekerjaan, progres pemrosesan, atau informasi bahwa hasil sedang disiapkan.	Berhubungan dengan proses asinkron GPT, RAG, dan <i>fusion</i> .
Halaman hasil asesmen	Menampilkan ringkasan hasil penilaian pengguna.	Skor akhir, tingkat kompetensi, dan rincian skor apabila tersedia.	Mengambil hasil asesmen yang telah dihitung oleh backend.
Halaman rekomendasi kursus	Menampilkan daftar kursus yang direkomendasikan.	Daftar rekomendasi Top-K atau Top-5 per jalur, skor kecocokan, dan <i>match reason</i> .	Mengambil data dari hasil RAG dan tabel rekomendasi.

Komponen keterlacakan rekomendasi	Membantu pengguna memahami alasan suatu kursus direkomendasikan.	Alasan rekomendasi, sinyal kecocokan, dan ringkasan dasar pengambilan.	Berhubungan dengan <i>recommendation_items</i> , <i>rag_query</i> , dan <i>rag_traces</i> .
Form umpan balik	Mengumpulkan penilaian pengguna terhadap hasil rekomendasi.	Nilai relevansi, penerimaan rekomendasi, dan komentar pengguna.	Mengirim data umpan balik ke endpoint <i>feedbacks</i> untuk kebutuhan evaluasi terbatas.

Berdasarkan Tabel 3.9, rancangan UI disusun untuk mendukung alur utama sistem tanpa memperluas fokus penelitian ke evaluasi desain antarmuka secara mendalam. Oleh karena itu, UI pada penelitian ini berperan sebagai media interaksi pengguna dengan layanan backend, sedangkan fokus utama penelitian tetap berada pada rancangan asesmen, pemrosesan asinkron, mekanisme RAG, dan keterlacakan rekomendasi.

Untuk mendukung transparansi kepada pengguna, halaman hasil dan rekomendasi dirancang tidak hanya menampilkan daftar kursus, tetapi juga menampilkan alasan rekomendasi / *match reason*. Informasi tersebut membantu pengguna memahami mengapa suatu kursus muncul sebagai rekomendasi. Dengan demikian, rancangan UI ikut mendukung tujuan sistem, yaitu menghasilkan rekomendasi yang tidak hanya relevan, tetapi juga lebih mudah dipahami dan ditelusuri.

3.3.2.8 Teknologi yang Digunakan

Komponen teknologi yang digunakan pada setiap lapisan sistem dirangkum pada Tabel 3.10.

Tabel 3. 10 Teknologi yang digunakan pada sistem

Lapisan	Komponen	Teknologi	Fungsi
HTTP	Layanan API / <i>API service</i>	FastAPI, Pydantic, JWT	Menyediakan endpoint HTTP, validasi request, autentikasi, dan otorisasi
Penilaian / <i>scoring</i>	Berbasis aturan / <i>rule-based</i>	Python	Menilai pertanyaan theoretical dan profile secara deterministik
Penilaian / <i>scoring</i>	Penilaian esai GPT / <i>GPT essay scoring</i>	API <i>Chat Completions</i> yang kompatibel dengan OpenAI	Menilai jawaban esai berdasarkan rubric scoring
Orkestrasi / <i>orchestration</i>	Pekerjaan asinkron / <i>async jobs</i>	PostgreSQL (async_jobs), FastAPI	Menyimpan status job dan mengorkestrasi pipeline internal

Lapisan	Komponen	Teknologi	Fungsi
		<i>BackgroundTasks</i>	
Pekerja latar / <i>worker</i> (opsional)	RQ <i>worker</i>	Redis, python-rq	Mendukung eksekusi job terpisah bila diaktifkan
Rekomendasi / <i>recommendation</i>	RAG dengan pemeringkatan hibrida ringan	Python, himpunan data CSV, pencocokan metadata, relevansi kata kunci, vektor token berbasis fungsi <i>hashing</i>	melakukan <i>Retrieval</i> lalu menghasilkan rekomendasi dan pencatatan <i>recommendation_items</i> serta <i>rag_traces</i> .
Basis data / <i>database</i>	Penyimpanan utama / <i>primary store</i>	PostgreSQL, SQLAlchemy ORM	Penyimpanan data transaksi dan metadata sistem.
Notifikasi / <i>notification</i>	Layanan email	Resend API	Pengiriman email ringkasan hasil asesmen.
Penerapan sistem / <i>deployment</i>	Hos / <i>hosting</i>	Render Web Service, Render PostgreSQL, Render Redis	Menjalankan aplikasi backend pada environment produksi yang aktif
Keteramatan / <i>observability</i>	Log terstruktur / <i>structured logs</i>	Structlog (JSON)	Pencatatan aktivitas untuk observabilitas dan debugging.

3.3.3 Penyerapan Umpan Balik / *Absorb Feedback*

Bagian ini menjelaskan rancangan metode evaluasi yang akan digunakan untuk menyerap umpan balik dari pengguna dan data operasional sistem. Pelaksanaan dan hasil evaluasi dilaporkan pada Bab 4.

3.3.3.1 Indikator Evaluasi

Indikator evaluasi disusun untuk membaca tiga aspek keberhasilan sistem: (1) kecepatan pemrosesan asesmen, (2) relevansi rekomendasi pembelajaran, dan (3) kualitas penilaian otomatis. Rincian indikator, parameter pengukuran, dan sumber data disajikan pada Tabel 3.11.

Tabel 3. 11 Indikator–parameter–sumber data evaluasi sistem

No.	Indikator	Parameter Pengukuran	Sumber Data
1	Kecepatan pemrosesan asesmen	Waktu penyelesaian / <i>turnaround time</i> dari <i>submit</i> hingga hasil tersedia	<i>assessments</i> , <i>recommendations</i>
2	Reliabilitas pemrosesan asinkron	Tingkat keberhasilan pekerjaan asinkron / <i>job</i> , jumlah percobaan ulang / <i>retry</i>	<i>async_jobs</i>
3	Keterlacakan penilaian	Metadata metode penilaian / <i>scoring method</i> dan aturan yang diterapkan /	<i>scores</i>

No.	Indikator	Parameter Pengukuran	Sumber Data
		<i>rules applied</i>	
4	Keterjelasan rekomendasi	Keberadaan skor relevansi / <i>relevance score</i> dan <i>match reason</i>	<i>recommendation_items</i>
5	Kesesuaian rekomendasi	<i>Precision@5, HitRate@5, NDCG@5</i>	Audit manual
6	Keselaran AES dengan penilai manusia	MAE, korelasi Pearson	Audit manual
7	Penerimaan pengguna	Penilaian relevansi / <i>rating</i> dan penerimaan dari pengguna	<i>feedbacks</i>

Beberapa istilah metrik pada Tabel 3.11 digunakan sebagai nama teknis. Waktu penyelesaian / *turnaround time* adalah selisih waktu dari pengiriman asesmen sampai hasil tersedia. Presisi lima rekomendasi teratas / *Precision@5* mengukur proporsi butir rekomendasi relevan pada lima rekomendasi teratas. Rasio temuan lima teratas / *HitRate@5* menunjukkan apakah setidaknya ada satu rekomendasi relevan dalam lima hasil teratas.

Ukuran kualitas peringkat lima teratas / *Normalized Discounted Cumulative Gain at 5* (NDCG@5) menilai kualitas urutan rekomendasi dengan memberi bobot lebih tinggi pada posisi atas. Galat absolut rata-rata / *mean absolute error* (MAE) mengukur rata-rata selisih absolut antara skor sistem dan skor manusia.

3.3.3.2 Rancangan Pelaksanaan Evaluasi

Evaluasi dirancang dalam tiga lapis yang saling melengkapi. Lapis pertama adalah observasi operasional, yaitu pengamatan kinerja sistem pada beban penggunaan nyata dengan data berupa waktu penyelesaian atau *turnaround time*, latensi GPT, tingkat keberhasilan job asinkron, dan kondisi terbatas atau *degraded*. Lapis kedua adalah uji coba pengguna, yaitu pengujian bertahap terhadap sepuluh partisipan untuk menilai fungsionalitas inti, integrasi *pipeline* GPT, RAG, dan *fusion*, serta stabilitas sistem pada kondisi beban awal. Lapis ketiga adalah audit manual kualitas hasil, yaitu pemeriksaan oleh peneliti terhadap sampel rekomendasi dan skor esai untuk mengukur kualitas relatif terhadap penilaian manusia. Rancangan tahapan uji coba pengguna disajikan pada Tabel 3.12.

Tabel 3. 12 Tahapan uji coba pengguna

Tahap	Ukuran Sampel	Fokus Evaluasi
Tahap 1	10 percobaan asesmen / <i>attempt</i>	Fungsi inti: mulai asesmen, pengiriman, pemantauan status, pengambilan hasil
Tahap 2	10 percobaan asesmen / <i>attempt</i>	Integrasi alur pemrosesan /

Tahap	Ukuran Sampel	Fokus Evaluasi
		<i>pipeline</i> GPT, RAG, dan penggabungan hasil / <i>fusion</i>
Tahap 3	10 percobaan asesmen / <i>attempt</i>	Perilaku sistem pada beban awal dan kondisi terbatas / <i>degraded</i>

3.3.3.3 Skenario Pengujian

Skenario pengujian dirancang agar setiap proses bisnis inti teruji pada kondisi normal maupun kondisi kegagalan. Ringkasan skenario disajikan pada Tabel 3.13.

Tabel 3. 13 Skenario pengujian sistem

No.	Proses Bisnis	Skenario	Keluaran yang Diharapkan
1	Mulai asesmen	Jalur normal	Salinan pertanyaan / <i>snapshot</i> sesuai peran
2	Pengiriman asesmen	Jalur normal	Asesmen terkunci, pekerjaan asinkron dibuat
3	Pengiriman asesmen	Idempotensi / <i>idempotency</i>	Tidak ada duplikasi
4	Pemrosesan GPT	Sukses	Skor dan metadata tersimpan
5	Pemrosesan GPT	batas waktu dan percobaan ulang / <i>timeout/retry</i>	Pekerjaan asinkron melakukan percobaan ulang / <i>retry</i> sesuai batas
6	Pengambilan RAG	Sukses	Item rekomendasi tersimpan
7	Pengambilan RAG	kondisi terbatas dan mekanisme cadangan / <i>degraded/fallback</i>	Keluaran tetap dihasilkan dengan penanda kondisi terbatas / <i>degraded flag</i>
8	Penggabungan hasil / <i>fusion</i>	Sukses	Ringkasan dan skor keseluruhan tersimpan
9	Pemantauan status	Konsistensi status / <i>state</i>	Status sesuai pekerjaan asinkron
10	Pengambilan hasil	Asesmen selesai	Hasil lengkap dapat diambil

3.3.3.4 Metrik Evaluasi

Metrik evaluasi dikelompokkan menurut lapis evaluasi pada Bagian 3.3.3.2. Rincian disajikan pada Tabel 3.14.

Tabel 3. 14 Metrik Evaluasi sistem

Lapis Evaluasi	Metrik	Metode Pengukuran
Observasi operasional	Waktu penyelesaian / <i>turnaround time</i>	Selisih waktu <i>created_at</i> dan <i>completed_at</i>
Observasi operasional	Latensi GPT	Nilai <i>latency</i> ms pada metadata <i>scoring</i>
Observasi operasional	Tingkat keberhasilan pekerjaan asinkron / <i>job</i>	Proporsi status = completed per jenis <i>job</i>
Observasi operasional	Tingkat kondisi terbatas / <i>degraded rate</i>	Proporsi asesmen dengan penanda kondisi terbatas / penanda kondisi terbatas / <i>degraded flag</i>
Uji coba pengguna	Relevansi	Rata-rata <i>rating relevance</i>

Lapis Evaluasi	Metrik	Metode Pengukuran
Uji coba pengguna	Penerimaan	Rata-rata <i>rating acceptance</i>
Audit manual	<i>Precision@5</i> , <i>HitRate@5</i> , <i>NDCG@5</i>	Perbandingan hasil sistem dengan label relevansi manusia
Audit manual	MAE, korelasi Pearson	Perbandingan skor AES dengan skor penilai manusia

3.3.3.5 Kriteria Interpretasi

Kriteria interpretasi digunakan sebagai pedoman pembacaan hasil pada BAB IV dan disusun dengan memperhatikan keterbatasan ukuran sampel penelitian:

1. Hasil dari lapis observasi operasional dibaca sebagai tolak ukur awal / *baseline* kelayakan operasional pada konteks penggunaan penelitian, bukan sebagai klaim umum untuk semua situasi.
2. Hasil dari lapis uji coba pengguna dibaca sebagai arah penerimaan awal, bukan sebagai simpulan kepuasan pengguna secara umum.
3. Hasil dari lapis audit manual dibaca sebagai indikasi awal kualitas rekomendasi dan keselarasan AES. Karena sampel audit terbatas (10 percobaan asesmen / *attempt*), hasil ini diposisikan sebagai bukti pendukung yang perlu dibaca secara hati-hati.

3.3.3.6 Instrumen Umpan Balik

Umpan balik pengguna akan dikumpulkan melalui dua jalur:

1. Formulir umpan balik pemangku kepentingan (pada tahap pendefinisian kebutuhan / *define requirements*, lihat Bagian 3.3.1.1).
2. Titik akses layanan / *endpoint* umpan balik pada sistem (selama uji coba pengguna), yang menyimpan *rating_relevance*, *rating_acceptance*, penilaian per butir / *item rating* opsional, dan komentar pada tabel feedbacks.

3.4 Finalisasi Produk / *Finalize Product*

Tahap finalisasi dalam penelitian ini didefinisikan sebagai kondisi ketika sistem telah memenuhi seluruh kebutuhan fungsional pada Tabel 3.2 dan lolos seluruh skenario pengujian pada Tabel 3.13. Kriteria finalisasi meliputi:

1. Seluruh titik akses layanan / *endpoint* pada Tabel 3.2 dapat diakses dan berfungsi sesuai spesifikasi.
2. Seluruh skenario pada Tabel 3.8 menghasilkan keluaran yang diharapkan atau, dalam kondisi kegagalan yang direncanakan, perilaku kondisi terbatas / *degraded* yang tercatat.

3. Indikator dan metrik evaluasi pada Tabel 3.11 dan Tabel 3.14 terekam secara konsisten.
4. Instrumen umpan balik pengguna (Bagian 3.3.3.6) dapat dioperasikan tanpa kendala.

Kondisi finalisasi yang dicapai dan perbandingannya dengan kriteria di atas dilaporkan pada Bab 4.

3.5 Ringkasan

Bab ini menjabarkan metode penelitian pengembangan MicroCred GenAI Backend menggunakan metodologi RAD dengan empat tahap, yaitu pendefinisian kebutuhan, perancangan purwarupa, penyerapan umpan balik melalui evaluasi berlapis, dan kriteria finalisasi produk. Kebutuhan sistem diturunkan dari pengumpulan data pemangku kepentingan, lalu diwujudkan dalam rancangan arsitektur monolit modular / *modular monolith*, skema basis data, dan mekanisme penilaian otomatis. Evaluasi dirancang dalam tiga lapis, yaitu observasi operasional, uji coba pengguna, dan audit manual, dengan kriteria interpretasi yang memperhatikan keterbatasan ukuran sampel. BAB IV akan menyajikan pelaksanaan dan hasil evaluasi berdasarkan kerangka yang dirumuskan pada BAB II

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Analisis Kebutuhan Awal Pengguna

Bagian ini menyajikan hasil analisis kebutuhan awal berdasarkan formulir umpan balik pemangku kepentingan yang telah dirancang pada BAB III. Analisis ini digunakan untuk memahami karakteristik responden, persepsi terhadap kredensial mikro, kendala dalam memilih kursus, serta prioritas pengguna dalam menentukan rekomendasi pembelajaran.

4.1.1 Profil Responden

Berdasarkan Tabel 4.1, delapan dari sepuluh responden berada pada rentang usia 21–23 tahun, enam responden baru menggunakan platform pembelajaran daring kurang dari enam bulan, dan enam responden sudah pernah menggunakan sistem rekomendasi berbasis AI. Komposisi ini menunjukkan bahwa sampel didominasi pengguna muda yang cukup akrab dengan ekosistem digital, namun belum seluruhnya memiliki pengalaman panjang dalam memilih kursus secara terarah.

Distribusi kendala utama responden divisualisasikan pada Tabel 4.1.

Tabel 4. 1 Profil 10 responden form kebutuhan awal

Kode	Usia (tahun)	Lama menggunakan platform pembelajaran online	Familiaritas terhadap sistem rekomendasi AI
R1	18-20	6-12 bulan	Sudah pernah menggunakan
R2	21-23	Lebih dari 2 tahun	Sudah pernah menggunakan
R3	18-20	Kurang dari 6 bulan	Tahu konsepnya, tetapi belum pernah menggunakan
R4	21-23	Kurang dari 6 bulan	Sudah pernah menggunakan
R5	21-23	Kurang dari 6 bulan	Tahu konsepnya, tetapi belum pernah menggunakan
R6	21-23	Kurang dari 6 bulan	Tahu konsepnya, tetapi belum pernah menggunakan
R7	21-23	Kurang dari 6 bulan	Sudah pernah menggunakan
R8	21-23	6-12 bulan	Tahu konsepnya, tetapi belum pernah menggunakan
R9	21-23	Kurang dari 6 bulan	Sudah pernah menggunakan
R10	21-23	1-2 tahun	Sudah pernah menggunakan

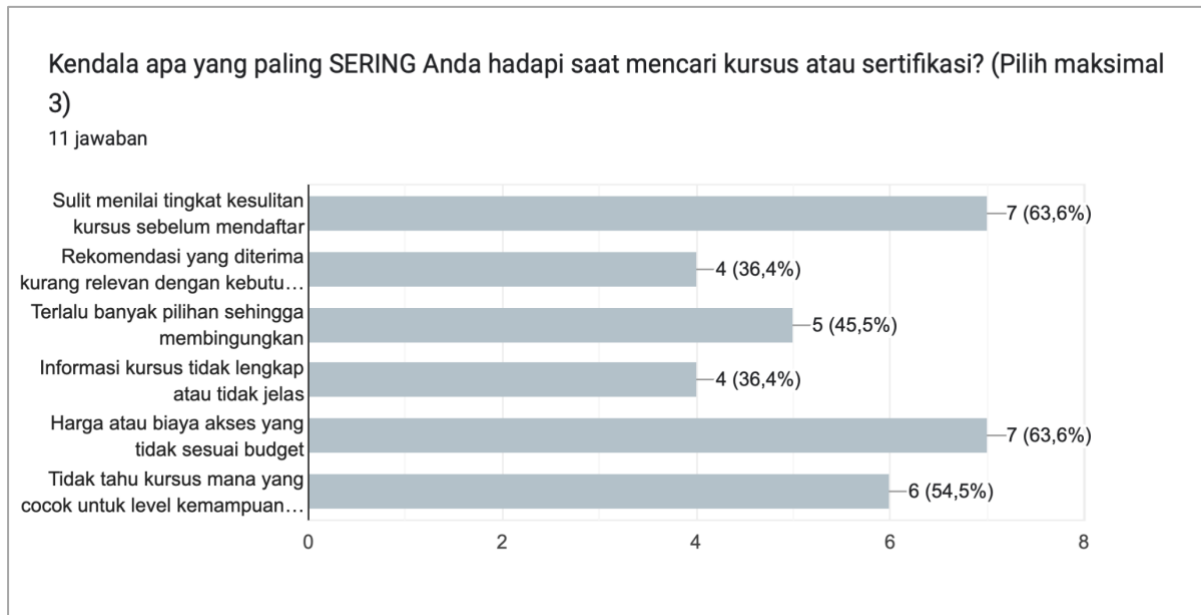
4.1.2 Persepsi dan Kendala Utama

Selain melihat profil responden, penelitian ini juga mengidentifikasi persepsi awal responden terhadap proses pencarian dan pemilihan kursus atau sertifikasi daring. Bagian ini bertujuan untuk mengetahui sejauh mana responden merasa mudah dalam memilih kursus yang sesuai, serta kendala apa saja yang paling sering mereka alami sebelum adanya sistem rekomendasi yang dikembangkan dalam penelitian ini.

Data pada bagian ini diperoleh dari jawaban responden terhadap pertanyaan terkait kemudahan memilih kursus, pengalaman ketidaksesuaian tingkat kesulitan, serta hambatan utama dalam menentukan kursus yang sesuai dengan kebutuhan dan kemampuan mereka. Ringkasan hasil persepsi awal dan kendala utama responden disajikan pada Tabel 4.2.

Tabel 4. 2 Ringkasan persepsi awal dan kendala utama responden

Indikator	Hasil
Rata-rata persepsi kemudahan mencari dan memilih kursus	2,3 dari 5
Distribusi persepsi	skor 1: 1 responden, skor 2: 5 responden, skor 3: 4 responden
Frekuensi mengalami mismatch tingkat kesulitan	jarang: 6 responden, sering: 3 responden, tidak pernah: 1 responden
Kendala paling sering: harga atau biaya tidak sesuai budget	7 responden
Kendala paling sering: sulit menilai tingkat kesulitan sebelum mendaftar	6 responden
Kendala paling sering: tidak tahu kursus mana yang cocok untuk level kemampuan	6 responden
Kendala cukup sering: terlalu banyak pilihan sehingga membingungkan	4 responden
Kendala cukup sering: informasi kursus tidak lengkap atau tidak jelas	4 responden
Kendala cukup sering: rekomendasi yang diterima kurang relevan	3 responden



Gambar 4. 1 Grafik distribusi kendala utama responden pada proses pencarian kursus dan sertifikasi daring

Hasil pada Tabel 4.2 menunjukkan bahwa masalah utama responden bukan sekadar banyaknya pilihan kursus, tetapi ketidakjelasan kecocokan antara kursus dan kondisi pengguna. Nilai rata-rata 2,3 dari 5 menunjukkan bahwa proses pencarian dan pemilihan kursus masih dipersepsi cenderung sulit. Kombinasi kendala berupa budget, tingkat kesulitan, dan ketidakjelasan level kemampuan menguatkan bahwa pengguna memerlukan sistem yang bukan hanya menampilkan daftar kursus, tetapi juga membantu menilai kesesuaian kursus terhadap kesiapan belajar, kemampuan saat ini, dan keterbatasan pengguna.

4.1.3 Prioritas Faktor Pemilih Kursus

Tabel 4.3 menunjukkan bahwa responden tetap menempatkan mutu isi kursus dan reputasi penyelenggara sebagai faktor utama. Akan tetapi, harga, review, dan tingkat kesulitan juga memperoleh skor tinggi. Artinya, masalah pemilihan kursus bersifat multi-faktor. Pada konteks ini, kebutuhan terhadap MicroCred GenAI menjadi lebih jelas karena pengguna memerlukan dukungan yang dapat membaca banyak sinyal sekaligus, bukan hanya satu kata kunci minat. Dengan kata lain, hasil form memberikan dasar yang kuat bahwa sistem rekomendasi perlu mempertimbangkan keterampilan, kesiapan, dan metadata kursus secara bersamaan.

Prioritas faktor pemilihan kursus disajikan pada Tabel 4.3.

Tabel 4. 3 Prioritas faktor dalam memilih kursus atau sertifikasi

Faktor	Rata-rata skor kepentingan
Materi atau isi kursus	4,1
Reputasi platform atau instruktur	4,1
Harga atau biaya kursus	4,0
Review atau rating pengguna lain	4,0
Tingkat kesulitan sesuai kemampuan	3,9
Sertifikat yang diberikan	3,8
Durasi atau waktu penyelesaian	3,4

4.2 Hasil Perancangan Sistem

Bagian ini menyajikan hasil perancangan MicroCred GenAI Backend berdasarkan kebutuhan sistem dan rancangan teknis yang telah dijelaskan pada BAB III. Hasil perancangan ini berfungsi sebagai jembatan antara hasil analisis kebutuhan awal pengguna pada Bagian 4.1 dan hasil implementasi sistem pada Bagian 4.3. Dengan demikian, bagian ini tidak membahas sistem yang telah berjalan, tampilan antarmuka yang telah diimplementasikan, maupun cuplikan kode, tetapi berfokus pada bagaimana kebutuhan penelitian diterjemahkan menjadi rancangan sistem.

Hasil perancangan sistem mencakup gambaran umum sistem, alur asesmen, mekanisme rekomendasi berbasis RAG, rancangan basis data, rancangan antarmuka pengguna, serta mekanisme transparansi dan keterlacakan. Setiap rancangan tersebut disusun untuk mendukung tujuan utama penelitian, yaitu menyediakan layanan asesmen kredensial mikro dan rekomendasi pembelajaran yang cepat, relevan, serta dapat ditelusuri.

4.2.1 Gambaran Umum Sistem yang Dirancang

MicroCred GenAI Backend dirancang sebagai layanan sisi peladen / *backend* yang mendukung dua proses utama, yaitu asesmen pembelajar dan rekomendasi kursus kredensial mikro. Proses asesmen digunakan untuk mengolah jawaban teori, jawaban esai, dan profil pembelajar. Sementara itu, proses rekomendasi digunakan untuk menghasilkan daftar kursus yang sesuai dengan hasil asesmen, profil pembelajar, preferensi belajar, serta tingkat kesiapan pengguna.

Secara arsitektural, sistem dirancang menggunakan pendekatan *modular monolith*. Pendekatan ini dipilih karena sistem tetap dijalankan sebagai satu aplikasi terpadu, tetapi tanggung jawab internalnya dipisahkan ke dalam beberapa modul logis. Pembagian tersebut membuat sistem tetap sederhana untuk tahap *minimum viable product* (MVP), namun tetap

memiliki batas fungsi yang jelas antara komponen asesmen, penilaian esai, rekomendasi, penyimpanan data, dan integrasi layanan eksternal.

Rancangan arsitektur sistem telah ditunjukkan pada Gambar 3.1. Berdasarkan rancangan tersebut, MicroCred GenAI Backend terdiri atas beberapa lapisan utama, yaitu lapisan antarmuka API, lapisan layanan aplikasi, lapisan data dan persistensi, lapisan integrasi eksternal, serta lapisan infrastruktur. Lapisan antarmuka API berfungsi menerima permintaan pengguna dan mengembalikan respons sistem. Lapisan layanan aplikasi menangani proses utama seperti asesmen, penilaian esai berbasis GPT, rekomendasi RAG, dan penggabungan hasil / *fusion*. Lapisan data dan persistensi digunakan untuk menyimpan data asesmen, hasil penilaian, rekomendasi, dan jejak pengambilan. Lapisan integrasi eksternal digunakan untuk menghubungkan sistem dengan layanan GPT dan layanan pendukung lainnya.

Hasil perancangan ini menunjukkan bahwa sistem tidak hanya dirancang sebagai mesin rekomendasi kursus, tetapi sebagai layanan asesmen dan rekomendasi yang saling terhubung. Hasil asesmen menjadi dasar pembentukan profil kompetensi, sedangkan profil tersebut digunakan untuk menyusun rekomendasi kursus yang lebih sesuai dengan kebutuhan pembelajar.

4.2.2 Hasil Perancangan Alur Asesmen

Alur asesmen dirancang untuk memisahkan proses yang perlu dijalankan secara langsung dan proses yang membutuhkan komputasi lebih berat. Pemisahan ini diperlukan agar pengguna tetap memperoleh respons awal secara cepat setelah mengirim jawaban, sementara proses penilaian esai, rekomendasi RAG, dan penggabungan hasil dapat dijalankan melalui pemrosesan asinkron.

Rancangan alur pemrosesan asesmen telah ditunjukkan pada Gambar 3.3. Berdasarkan rancangan tersebut, proses asesmen dimulai ketika pengguna memilih konteks atau peran asesmen, kemudian sistem menyiapkan salinan pertanyaan / *question snapshot*. Salinan ini digunakan agar pertanyaan yang dikerjakan pengguna tetap terdokumentasi secara konsisten meskipun data master pertanyaan dapat berubah pada waktu berikutnya.

Setelah pengguna mengirim jawaban, sistem dirancang untuk melakukan validasi kepemilikan asesmen, menyimpan respons pengguna, mengunci status asesmen, dan menjalankan penilaian awal berbasis aturan / *rule-based scoring*. Penilaian berbasis aturan digunakan untuk jenis pertanyaan yang dapat dihitung secara deterministik, seperti pertanyaan teori dan profil pembelajar.

Proses yang lebih berat dirancang untuk dijalankan melalui pekerjaan asinkron / *asynchronous job*. Pekerjaan ini mencakup penilaian esai berbasis GPT, pengambilan dan pemeringkatan rekomendasi melalui RAG, serta penggabungan hasil / *fusion*. Dengan rancangan ini, proses pengiriman jawaban tidak harus menunggu seluruh proses GPT dan RAG selesai dalam satu permintaan yang sama.

Tahap *fusion* dirancang untuk menyatukan hasil penilaian berbasis aturan, hasil penilaian esai, skor akhir, tingkat kompetensi, dan rekomendasi kursus ke dalam satu keluaran akhir. Dengan demikian, hasil yang diterima pengguna tidak hanya berupa nilai, tetapi juga memuat arah pembelajaran lanjutan yang sesuai dengan profil kompetensinya.

4.2.3 Hasil Perancangan Mekanisme RAG

Mekanisme *retrieval-augmented generation* (RAG) dirancang sebagai dasar penyusunan rekomendasi kursus kredensial mikro. Dalam penelitian ini, RAG tidak dipahami sebagai proses generasi bebas tanpa sumber, tetapi sebagai alur pengambilan kandidat kursus dari basis pengetahuan lokal, kemudian mengurutkannya menggunakan beberapa sinyal relevansi. Rancangan mekanisme RAG telah ditunjukkan pada Gambar 3.4 dan komponen utamanya telah dirangkum pada Tabel 3.4.

Masukan utama pada mekanisme RAG berasal dari hasil asesmen, profil pembelajar, preferensi belajar, kata kunci esai / *essay keywords*, topik yang belum dikuasai / *missed topics*, dan skor asesmen. Informasi tersebut digunakan oleh *query builder* untuk membentuk kueri rekomendasi. Selain itu, sistem dirancang memiliki *readiness policy* atau *readiness gate* untuk menentukan arah rekomendasi berdasarkan tingkat kesiapan pengguna.

Melalui *readiness policy*, pengguna dapat diarahkan ke jalur rekomendasi yang berbeda. Pengguna yang masih membutuhkan penguatan dasar dapat diarahkan ke jalur fondasi, sedangkan pengguna dengan kesiapan lebih baik dapat diarahkan ke jalur menengah atau jalur target. Dengan demikian, rekomendasi tidak hanya mempertimbangkan kecocokan topik, tetapi juga kesiapan belajar pengguna.

Basis pengetahuan yang digunakan dalam rancangan RAG berasal dari dataset kursus Udemy berbentuk CSV. Field asli pada dataset mencakup *course_title*, *url*, *is_paid*, *price*, *num_subscribers*, *num_reviews*, *num_lectures*, *level*, *content_duration*, *published_timestamp*, dan *subject*. Field tersebut digunakan sebagai sumber informasi utama dalam proses pencocokan awal karena merepresentasikan informasi

kursus, seperti judul, tautan, status berbayar atau gratis, harga, jumlah pelanggan, jumlah ulasan, jumlah materi, tingkat kesulitan, durasi, waktu publikasi, dan subjek kursus.

Selain menggunakan field asli, sistem juga dirancang melakukan pengayaan data / *enrichment*. Pengayaan ini mencakup pembentukan tag, topik, dan atribut turunan lain yang membantu proses pencocokan rekomendasi. Dengan demikian, basis pengetahuan sistem terdiri atas data asli dari dataset kursus dan hasil pengayaan lokal yang dibuat untuk kebutuhan rekomendasi.

Pada tahap pemrosesan, kueri dan data kursus dirancang melalui tokenisasi, ekstraksi tag, dan *feature hashing*. Sistem tidak dirancang menggunakan model *embedding* eksternal. Oleh karena itu, istilah *embedding_score* pada rancangan ini merujuk pada skor kemiripan representasi token berbasis *feature hashing* lokal, bukan pada *neural embedding* dari model eksternal.

Kandidat kursus yang telah diperoleh kemudian dirancang untuk diurutkan menggunakan *hybrid ranker*. Pemeringkatan ini menggabungkan beberapa komponen skor, yaitu *match_score*, *keyword_score*, *embedding_score*, *tag_boost*, dan *quality_score*. Hasil akhir proses ini berupa rekomendasi Top-K atau Top-5 per jalur rekomendasi, disertai *match_reason* agar alasan kecocokan kursus dapat dipahami oleh pengguna.

4.2.4 Hasil Perancangan Basis Data

Basis data dirancang untuk mendukung penyimpanan data pengguna, sesi asesmen, jawaban, skor, pekerjaan asinkron, rekomendasi, item rekomendasi, umpan balik, serta jejak pemrosesan. Rancangan basis data diperlukan agar seluruh proses dari awal pengguna memulai asesmen hingga hasil rekomendasi tersedia dapat disimpan secara terstruktur dan dapat ditelusuri kembali.

Rancangan entitas utama basis data telah diringkas pada Tabel 3.5, sedangkan hubungan antarentitas telah divisualisasikan melalui ERD pada Gambar 3.5. Berdasarkan rancangan tersebut, entitas pengguna digunakan untuk menyimpan identitas dan peran pengguna. Entitas asesmen digunakan untuk menyimpan sesi asesmen yang dikerjakan pengguna. Entitas pertanyaan dan jawaban digunakan untuk mencatat pertanyaan yang diberikan serta respons yang dikirimkan pengguna. Entitas skor digunakan untuk menyimpan hasil penilaian dari komponen berbasis aturan dan penilaian esai.

Selain itu, sistem juga dirancang memiliki entitas pekerjaan asinkron / *async jobs*. Entitas ini digunakan untuk mencatat status proses yang berjalan di belakang layar, seperti GPT, RAG,

dan *fusion*. Dengan adanya entitas ini, sistem dapat memantau apakah suatu pekerjaan masih berjalan, sudah selesai, atau mengalami kegagalan.

Pada komponen rekomendasi, entitas *recommendations* digunakan untuk menyimpan hasil rekomendasi secara umum, sedangkan *recommendation_items* digunakan untuk menyimpan detail setiap kursus yang direkomendasikan. Informasi seperti *relevance score*, *match_reason*, *rag_query*, dan *rag_traces* dirancang untuk mendukung keterlacakan hasil rekomendasi. Dengan demikian, basis data tidak hanya menyimpan hasil akhir, tetapi juga menyimpan dasar pengambilan dan bukti pendukung proses rekomendasi.

Rancangan aliran data sistem telah ditunjukkan pada Gambar 3.6. Berdasarkan rancangan tersebut, data pengguna mengalir dari antarmuka menuju layanan backend, kemudian diproses melalui layanan asesmen, layanan GPT, layanan RAG, dan layanan penyimpanan. Aliran data ini menunjukkan hubungan antara proses interaksi pengguna, pemrosesan sistem, dan penyimpanan hasil.

4.2.5 Hasil Perancangan Antarmuka Pengguna

Meskipun penelitian ini berfokus pada pengembangan layanan sisi peladen / *backend*, rancangan antarmuka pengguna tetap disusun untuk menggambarkan bagaimana pengguna berinteraksi dengan sistem. Rancangan antarmuka tidak menjadi fokus utama evaluasi, tetapi digunakan untuk mendukung uji coba MVP dan memperjelas alur penggunaan sistem.

Rancangan antarmuka pengguna telah dirangkum pada Tabel 3.9. Berdasarkan rancangan tersebut, antarmuka pengguna mencakup beberapa halaman utama, yaitu halaman mulai asesmen, halaman pengerjaan asesmen, halaman status pemrosesan, halaman hasil asesmen, halaman rekomendasi kursus, komponen keterlacakan rekomendasi, dan form umpan balik.

Halaman mulai asesmen dirancang untuk memberikan pengantar kepada pengguna sebelum mengerjakan asesmen. Halaman pengerjaan asesmen dirancang untuk menampilkan pertanyaan dan menerima jawaban pengguna. Halaman status pemrosesan dirancang untuk menunjukkan bahwa sistem sedang memproses hasil asesmen, terutama ketika proses GPT, RAG, dan *fusion* masih berjalan. Halaman hasil asesmen dirancang untuk menampilkan skor, tingkat kompetensi, dan ringkasan hasil penilaian. Halaman rekomendasi kursus dirancang untuk menampilkan daftar kursus yang sesuai dengan profil dan hasil asesmen pengguna.

Rancangan antarmuka juga memuat komponen keterjelasan rekomendasi. Pada halaman rekomendasi, setiap kursus tidak hanya dirancang untuk menampilkan nama kursus, tetapi juga skor kecocokan dan alasan rekomendasi / *match_reason*. Informasi ini penting agar pengguna

dapat memahami mengapa suatu kursus direkomendasikan. Selain itu, sistem juga dirancang menyediakan form umpan balik agar pengguna dapat menilai relevansi dan penerimaan terhadap rekomendasi yang diberikan.

Dengan rancangan tersebut, antarmuka pengguna berperan sebagai media interaksi antara pengguna dan layanan backend. UI membantu pengguna menjalani alur asesmen dan memahami hasil rekomendasi, sedangkan proses inti seperti penilaian, RAG, *fusion*, dan penyimpanan jejak tetap berada pada sisi backend.

4.2.6 Hasil Perancangan Transparansi dan Keterlacakan

Transparansi dan keterlacakan dirancang sebagai bagian penting dari MicroCred GenAI Backend karena sistem tidak hanya menghasilkan skor dan rekomendasi, tetapi juga perlu menyediakan dasar yang dapat diperiksa kembali. Keterlacakan ini diperlukan agar hasil asesmen dan rekomendasi tidak dipahami sebagai keluaran kotak hitam.

Pada proses asesmen, keterlacakan dirancang melalui penyimpanan rincian skor, rubrik penilaian, dan hasil penilaian esai. Keluaran GPT tidak langsung dijadikan nilai akhir secara bebas, tetapi dirancang untuk divalidasi dan dihitung kembali oleh backend menggunakan struktur skor dan bobot rubrik yang telah ditentukan. Dengan mekanisme ini, proses penilaian tetap memiliki batasan dan struktur yang dapat diperiksa.

Pada proses rekomendasi, keterlacakan dirancang melalui penyimpanan *relevance score*, *match_reason*, *rag_query*, dan *rag_traces*. *Relevance score* digunakan untuk menunjukkan tingkat kecocokan rekomendasi, *match_reason* digunakan untuk menjelaskan alasan rekomendasi, *rag_query* digunakan untuk menyimpan dasar kueri yang dibentuk dari profil dan hasil asesmen pengguna, sedangkan *rag_traces* digunakan untuk menyimpan jejak pengambilan kandidat kursus.

Rancangan ini membuat hasil rekomendasi tidak hanya berupa daftar kursus, tetapi juga memiliki dasar pengambilan yang dapat ditelusuri. Dengan demikian, pengguna dan pengelola sistem dapat memahami mengapa suatu kursus direkomendasikan dan sinyal apa saja yang digunakan dalam proses pemeringkatan.

4.2.7 Ringkasan Hasil Perancangan

Berdasarkan hasil perancangan yang telah dijelaskan, MicroCred GenAI Backend dirancang sebagai sistem asesmen dan rekomendasi yang saling terhubung. Hasil asesmen digunakan sebagai dasar pembentukan profil kompetensi dan kueri rekomendasi. Mekanisme

RAG digunakan untuk mengambil dan mengurutkan kandidat kursus dari basis pengetahuan lokal. Basis data dirancang untuk menyimpan hasil, proses, dan jejak pengambilan. Antarmuka pengguna dirancang untuk mendukung interaksi pengguna dengan sistem dan membantu pengguna memahami hasil yang diberikan.

Hasil perancangan ini menjadi dasar bagi tahap implementasi sistem yang dijelaskan pada Bagian 4.3. Dengan demikian, pembaca dapat memahami hubungan antara kebutuhan awal pengguna, rancangan sistem, dan komponen yang kemudian direalisasikan dalam bentuk sistem MVP.

4.3 Hasil Implementasi Sistem

Bagian ini menyajikan hasil implementasi MicroCred GenAI Backend berdasarkan hasil perancangan sistem yang telah dijelaskan pada Bagian 4.2. Fokus bagian ini adalah menunjukkan komponen sistem yang berhasil dibangun, realisasi kebutuhan fungsional, tampilan sistem yang digunakan dalam uji coba, serta cuplikan kode yang merepresentasikan implementasi komponen utama. Dengan demikian, bagian ini tidak membahas hasil evaluasi sistem, karena hasil pengujian dan evaluasi disajikan secara khusus pada Bagian 4.4.

Implementasi sistem dilakukan untuk memastikan bahwa rancangan asesmen, pemrosesan asinkron, penilaian esai berbasis GPT, rekomendasi berbasis RAG, penggabungan hasil / *fusion*, penyimpanan jejak rekomendasi, serta antarmuka pengguna dapat berjalan dalam satu alur MVP. Hasil implementasi ini menjadi dasar bagi pengujian sistem pada bagian berikutnya.

4.3.1 Sistem yang Terbangun

MicroCred GenAI Backend berhasil diimplementasikan sebagai layanan sisi peladen / *backend* yang mendukung proses asesmen kredensial mikro dan rekomendasi pembelajaran. Sistem dibangun untuk menangani alur utama mulai dari pengguna memulai asesmen, mengirim jawaban, memperoleh hasil penilaian, menerima rekomendasi kursus, hingga memberikan umpan balik terhadap rekomendasi yang diterima.

Implementasi sistem mengikuti rancangan *modular monolith* yang telah dijelaskan pada Bagian 4.2. Komponen sistem dipisahkan ke dalam beberapa lapisan, yaitu lapisan antarmuka API, lapisan layanan aplikasi, lapisan data dan persistensi, serta lapisan integrasi eksternal. Lapisan antarmuka API digunakan untuk menerima permintaan dari pengguna dan mengembalikan respons sistem. Lapisan layanan aplikasi menangani logika utama, seperti asesmen, penilaian esai, RAG, dan *fusion*. Lapisan data dan persistensi digunakan untuk

menyimpan data asesmen, rekomendasi, umpan balik, dan jejak pemrosesan. Lapisan integrasi eksternal digunakan untuk menghubungkan sistem dengan layanan GPT dan layanan pendukung lain.

Pada tahap MVP, sistem difokuskan untuk membuktikan bahwa alur asesmen, penilaian esai, rekomendasi, dan keterlacakan hasil dapat berjalan dalam satu rangkaian proses yang terintegrasi. Dengan demikian, hasil implementasi tidak hanya menunjukkan bahwa masing-masing komponen dapat dibangun, tetapi juga menunjukkan bahwa komponen tersebut dapat saling terhubung dalam satu alur penggunaan.

4.3.2 Realisasi Kebutuhan Fungsional

Realisasi kebutuhan fungsional digunakan untuk menunjukkan bahwa kebutuhan yang telah dirancang pada BAB III telah diterjemahkan ke dalam komponen sistem. Setiap kebutuhan fungsional dipetakan ke titik akses layanan / *endpoint* atau komponen implementasi yang menjalankan fungsi tersebut. Ringkasan realisasi kebutuhan fungsional ditunjukkan pada Tabel 4.4.

Tabel 4. 4 Realisasi kebutuhan fungsional pada implementasi

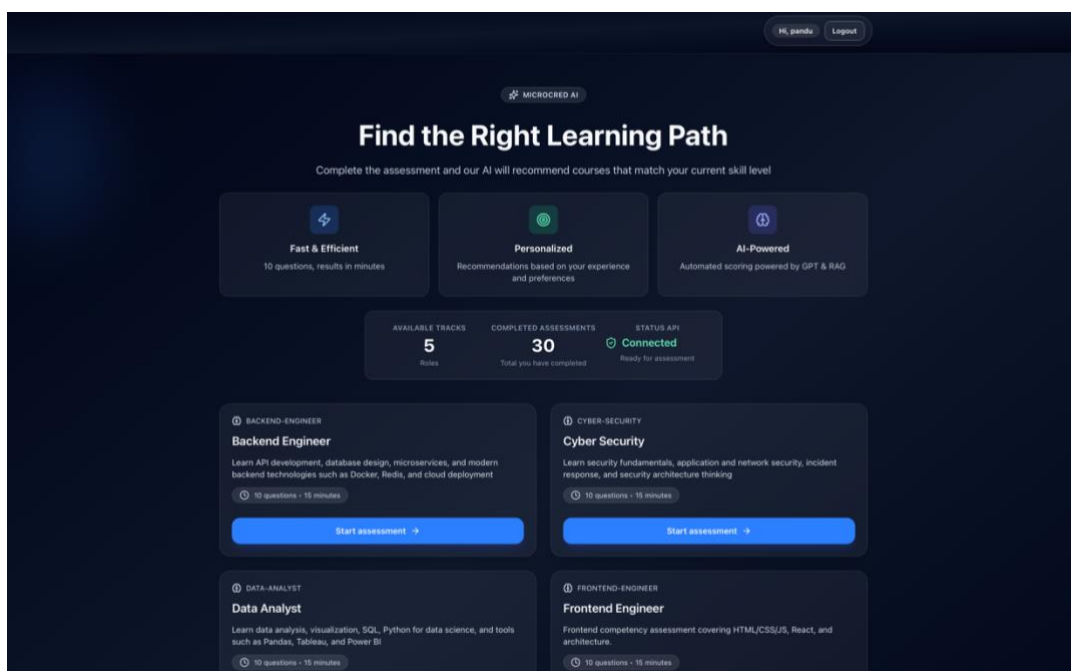
Kode	Kebutuhan Fungsional	Titik Akses Layanan / Komponen	Status
FR1	Mulai atau lanjut asesmen	POST /assessments/start	Aktif
FR2	Pengiriman asesmen	POST /assessments/{id}/submit	Aktif
FR3	Penilaian berbasis aturan / <i>rule-based scoring</i>	SubmissionService	Aktif
FR4	Penilaian esai berbasis GPT	GPTEssayScoringService	Aktif
FR5	Rekomendasi berbasis RAG	RAGService	Aktif
FR6	Penggabungan hasil / <i>fusion</i>	FusionService	Aktif
FR7	Pemantauan status asesmen	GET /assessments/{id}/status	Aktif
FR8	Pengiriman ringkasan hasil melalui email	POST /assessments/{id}/email-summary	Aktif, opsional
FR9	Pengiriman umpan balik pengguna	POST /assessments/{id}/feedback	Aktif

Berdasarkan Tabel 4.4, seluruh kebutuhan fungsional utama telah direalisasikan dalam bentuk *endpoint* atau layanan aplikasi. Fitur pengiriman ringkasan hasil melalui email diposisikan sebagai fitur opsional karena tidak menjadi jalur utama dalam evaluasi sistem. Sementara itu, fitur inti seperti memulai asesmen, mengirim jawaban, memproses penilaian, menghasilkan rekomendasi, memantau status, dan mengirim umpan balik menjadi bagian utama dari alur MVP.

Tabel 4.4 juga menunjukkan bahwa implementasi sistem tidak hanya berfokus pada satu fitur tunggal, tetapi mencakup keseluruhan alur asesmen dan rekomendasi. Kebutuhan FR1 dan FR2 berkaitan dengan interaksi awal pengguna, FR3 sampai FR6 berkaitan dengan pemrosesan inti sistem, sedangkan FR7 sampai FR9 mendukung pemantauan hasil, pengiriman ringkasan, dan pengumpulan umpan balik.

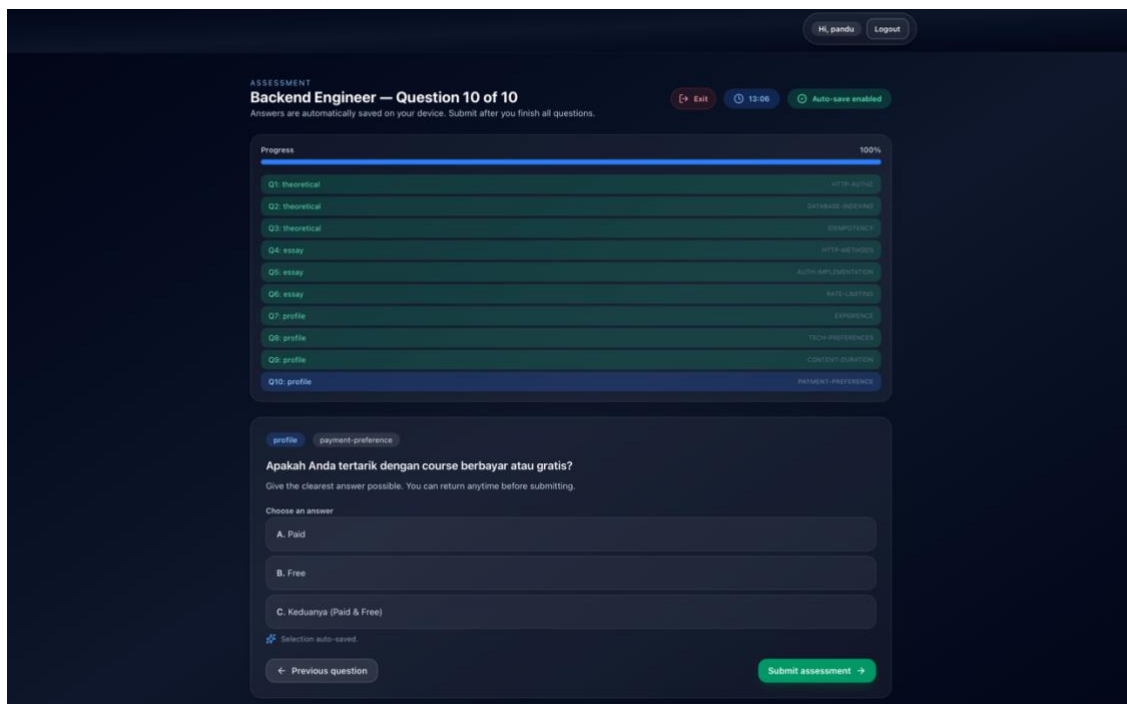
4.3.3 Implementasi Antarmuka Pengguna

Walaupun fokus utama penelitian berada pada layanan sisi peladen / *backend*, antarmuka pengguna tetap diimplementasikan untuk mendukung uji coba MVP. Antarmuka ini digunakan sebagai media agar pengguna dapat menjalani alur asesmen, melihat hasil penilaian, membaca rekomendasi kursus, dan memberikan umpan balik terhadap hasil yang diterima. Tampilan awal asesmen ditunjukkan pada Gambar 4.2.



Gambar 4. 2 Halaman mulai asesmen

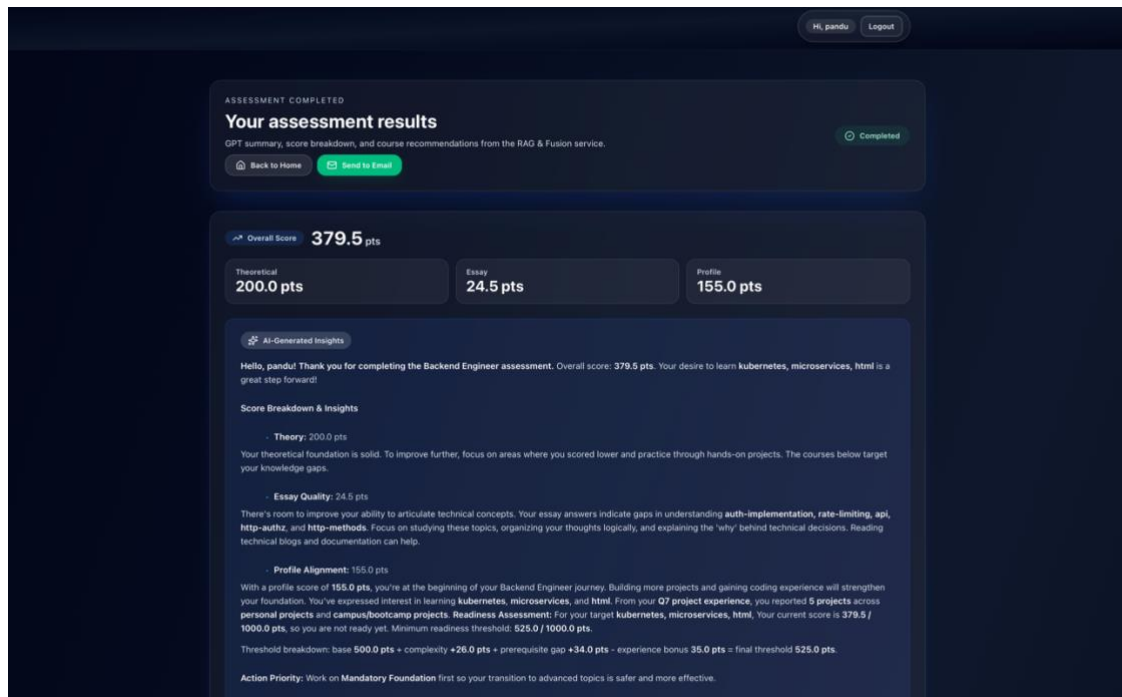
Berdasarkan Gambar 4.2, halaman mulai asesmen berfungsi sebagai titik awal interaksi pengguna dengan sistem. Pada halaman ini, pengguna memperoleh informasi awal mengenai asesmen, tujuan penggunaan sistem, serta pilihan konteks atau peran pembelajaran yang ingin diuji. Tampilan ini membantu pengguna memahami bahwa asesmen yang dilakukan tidak hanya menghasilkan skor, tetapi juga menjadi dasar penyusunan rekomendasi pembelajaran. Setelah pengguna memulai asesmen, sistem menampilkan halaman pengerjaan soal. Tampilan halaman pengerjaan soal ditunjukkan pada Gambar 4.3.



Gambar 4. 3 Halaman pengerjaan soal asesmen

Berdasarkan Gambar 4.3, halaman pengerjaan soal asesmen menampilkan daftar progres pertanyaan, jenis pertanyaan, dan area pengisian jawaban. Sistem menampilkan beberapa jenis soal, seperti pertanyaan teori, esai, dan profil pembelajar. Pada jenis pertanyaan tertentu, pengguna dapat memilih opsi jawaban yang tersedia atau mengisi jawaban secara manual. Navigasi antarpertanyaan juga disediakan agar pengguna dapat berpindah dari satu pertanyaan ke pertanyaan berikutnya secara lebih terarah.

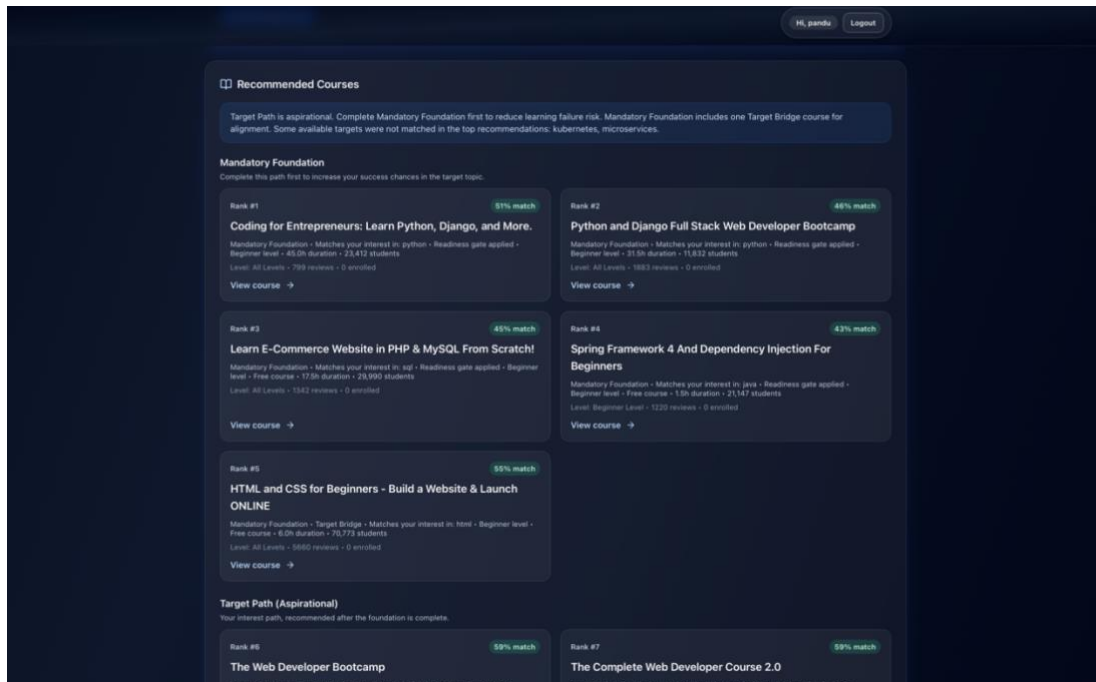
Halaman hasil asesmen ditampilkan setelah proses penilaian dan penggabungan hasil selesai. Tampilan halaman hasil asesmen ditunjukkan pada Gambar 4.4.



Gambar 4. 4 Halaman hasil asesmen dengan rincian penilaian / *scoring breakdown*

Berdasarkan Gambar 4.4, halaman hasil asesmen menampilkan status asesmen, skor keseluruhan, rincian skor teori, rincian skor esai, serta wawasan yang dihasilkan sistem. Rincian penilaian / *scoring breakdown* membantu pengguna memahami bagaimana nilai akhir terbentuk dari beberapa komponen asesmen. Selain itu, halaman hasil juga menampilkan kekuatan pengguna, area yang perlu ditingkatkan, serta arahan jalur pembelajaran yang sesuai dengan hasil asesmen.

Bagian lanjutan dari hasil asesmen berupa rekomendasi kursus ditunjukkan pada Gambar 4.5.



Gambar 4. 5 Tampilan daftar rekomendasi kursus dengan nilai kecocokan / *match score*

Berdasarkan Gambar 4.5, sistem menampilkan daftar kursus yang direkomendasikan berdasarkan hasil asesmen dan profil pengguna. Setiap kartu rekomendasi memuat informasi kursus, nilai kecocokan / *match score*, alasan rekomendasi / *match reason*, tingkat kesulitan, durasi, dan tautan menuju kursus. Tampilan ini menunjukkan bahwa rekomendasi tidak hanya disajikan sebagai daftar kursus, tetapi juga dilengkapi dengan alasan yang membantu pengguna memahami dasar kecocokan kursus tersebut.

Selain menampilkan rekomendasi, halaman ini juga menyediakan form umpan balik. Melalui form tersebut, pengguna dapat memberikan penilaian terhadap relevansi dan penerimaan rekomendasi. Data umpan balik ini kemudian digunakan sebagai salah satu sumber evaluasi pada Bagian 4.4.

4.3.4 Implementasi Alur Asesmen dan Pemrosesan Asinkron

Alur asesmen diimplementasikan dengan memisahkan proses sinkron dan proses asinkron. Proses sinkron digunakan untuk menangani permintaan yang perlu segera diproses ketika pengguna berinteraksi dengan sistem, seperti memulai asesmen, mengirim jawaban, menyimpan respons, dan menjalankan penilaian berbasis aturan. Proses asinkron digunakan

untuk menjalankan pekerjaan yang lebih berat, seperti penilaian esai berbasis GPT, rekomendasi RAG, dan penggabungan hasil / *fusion*.

Setelah pengguna mengirim asesmen, sistem menjalankan penilaian awal yang bersifat deterministik, kemudian memicu alur pemrosesan asinkron. Cuplikan kode pemicu alur pemrosesan asinkron ditunjukkan pada Gambar 4.6.

Pemicu *pipeline* asinkron setelah *submit*:

```
settings = get_settings()
if settings.auto_process_jobs:
    background_tasks.add_task(_run_async_pipeline,
result.assessment_id)
```

Gambar 4. 6 Cuplikan kode pemicu alur pemrosesan asinkron setelah pengguna mengirim asesmen

Berdasarkan Gambar 4.6, sistem tidak menyelesaikan seluruh proses berat dalam satu permintaan pengguna. Setelah proses awal selesai, sistem menjadwalkan pekerjaan latar untuk menjalankan alur GPT, RAG, dan *fusion*. Implementasi ini membantu menjaga respons awal tetap cepat karena pengguna tidak perlu menunggu seluruh proses penilaian esai dan rekomendasi selesai secara langsung pada saat mengirim jawaban.

Mekanisme pemrosesan asinkron ini penting karena sistem melibatkan layanan eksternal dan proses komputasi yang tidak selalu memiliki waktu respons seragam. Dengan memisahkan proses sinkron dan asinkron, sistem dapat mengurangi risiko permintaan pengguna mengalami waktu tunggu terlalu panjang. Pengguna juga dapat memantau status asesmen melalui endpoint pemantauan status sampai hasil akhir tersedia.

4.3.5 Implementasi Rekomendasi Berbasis RAG

Rekomendasi pembelajaran diimplementasikan melalui layanan RAG yang mengambil dan mengurutkan kandidat kursus dari basis pengetahuan lokal. Masukan untuk proses rekomendasi berasal dari hasil asesmen, profil pembelajar, preferensi belajar, kata kunci esai, topik yang belum dikuasai, dan skor asesmen. Informasi tersebut digunakan untuk membentuk kueri rekomendasi dan menentukan kandidat kursus yang paling sesuai.

Pada implementasi MVP, sistem menggunakan pemeringkatan hibrida ringan untuk menghitung skor akhir kandidat kursus. Cuplikan kode formula perhitungan skor akhir kandidat kursus ditunjukkan pada Gambar 4.7.

Formula skor pada layanan RAG:

```
# Combine all scores
# Priority: enriched match score > keyword/embedding > quality
final_score = (
    match_score * 0.5 # 50% from enriched metadata matching
    + keyword_score * 0.2 # 20% from keyword relevance
    + embedding_score * 0.1 # 10% from embedding similarity
    + tag_boost # Boost for missed topics
    + enriched.quality_score * 0.2 # 20% from quality
)

scored_courses.append((final_score, course, match_reason))
```

Gambar 4. 7 Cuplikan kode formula perhitungan skor akhir kandidat kursus pada layanan

RAG dengan pemeringkatan hibrida ringan

Berdasarkan Gambar 4.7, skor akhir kandidat kursus dihitung dengan menggabungkan beberapa komponen, yaitu *match_score*, *keyword_score*, *embedding_score*, *tag_boost*, dan *quality_score*. *Match_score* merepresentasikan kecocokan metadata yang diperkaya, *keyword_score* menunjukkan relevansi kata kunci, *embedding_score* menunjukkan kemiripan representasi token berbasis *feature hashing* lokal, *tag_boost* memberikan penguatan terhadap tag atau topik tertentu, sedangkan *quality_score* merepresentasikan sinyal kualitas kursus.

Implementasi ini menunjukkan bahwa rekomendasi tidak dibentuk secara bebas oleh model generatif, tetapi diambil dari dataset kursus yang tersedia dan diurutkan berdasarkan sinyal yang dapat diperiksa. Dengan demikian, sistem tetap mempertahankan keterikatan rekomendasi terhadap basis pengetahuan lokal serta mendukung keterlacakan hasil melalui penyimpanan skor, alasan rekomendasi, dan jejak pengambilan.

Hasil rekomendasi kemudian disimpan dalam struktur data rekomendasi. Item rekomendasi menyimpan informasi kursus, skor kecocokan, dan alasan rekomendasi, sedangkan data seperti *rag_query* dan *rag_traces* digunakan untuk menelusuri dasar pengambilan rekomendasi. Penyimpanan ini mendukung tujuan sistem agar rekomendasi tidak hanya relevan, tetapi juga dapat ditelusuri kembali.

4.3.6 Implementasi Penilaian Esai Berbasis GPT

Penilaian esai berbasis GPT diimplementasikan untuk menilai jawaban terbuka pengguna yang tidak dapat dihitung secara langsung menggunakan aturan sederhana. Pada proses ini, model GPT menerima konteks pertanyaan, jawaban pengguna, jawaban acuan apabila tersedia, rubrik penilaian, dan tingkat kesulitan. Keluaran model diarahkan agar berbentuk JSON sehingga dapat diproses kembali oleh backend.

Implementasi penilaian esai tidak hanya dilakukan melalui instruksi konseptual pada laporan, tetapi juga diwujudkan dalam pembentukan prompt pada kode program. System prompt digunakan untuk menetapkan peran dan batasan model sebagai evaluator esai, sedangkan user prompt membawa data penilaian spesifik untuk setiap asesmen. Data tersebut meliputi pertanyaan, jawaban esai pengguna, jawaban acuan apabila tersedia, serta bobot rubrik penilaian. Cuplikan kode pembentukan prompt dan pemanggilan GPT untuk penilaian esai ditunjukkan pada Gambar 4.8.

Pembentukan prompt:

```

ESSAY_SCORING_SYSTEM_PROMPT = """You are a strict essay evaluator for a
micro-credential \
assessment platform.

Evaluation constraints:
- You are an evaluator, not a tutor.
- Use only the provided question, student answer, and reference answer in
the prompt.
- Do not use external knowledge.
- Do not provide suggestions, feedback format, markdown, or prose outside
JSON.

Rubric dimensions (0-100):
- relevance: How well the answer addresses the question
- depth: Depth of analysis and understanding
- clarity: Clarity and organization of expression
- completeness: Coverage of the asked scope
- technical: Technical correctness and terminology usage

Output contract (single JSON object only):
{
  "scores": {
    "relevance": <number>,
    "depth": <number>,
    "clarity": <number>,
    "completeness": <number>,
    "technical": <number>
  },
  "total_score": <number>,
  "explanation": "<short rationale linked to rubric>"
}
"""

```

Gambar 4. 8 Cuplikan kode pembentukan prompt dan pemanggilan GPT untuk penilaian esai

Berdasarkan Gambar 4.8, rubrik penilaian tidak hanya dijelaskan secara konseptual, tetapi dikirimkan ke model melalui user prompt bersama pertanyaan, jawaban pengguna, dan jawaban acuan apabila tersedia. Pemisahan antara system prompt dan user prompt digunakan agar instruksi evaluator tetap konsisten, sementara data penilaian dapat menyesuaikan setiap jawaban pengguna. Dengan demikian, model GPT tidak diminta menilai secara bebas, tetapi

diarahkan untuk melakukan evaluasi berdasarkan konteks dan rubrik yang diberikan oleh sistem.

Meskipun GPT digunakan sebagai evaluator awal, nilai akhir tidak langsung diambil secara bebas dari model. Backend tetap melakukan validasi struktur keluaran, pemeriksaan kelengkapan field, pembatasan nilai pada rentang 0–100, dan perhitungan skor akhir berdasarkan bobot rubrik. Cuplikan kode validasi dan perhitungan skor akhir esai ditunjukkan pada Gambar 4.9.

Rubrik Penilaian pada Prompt:

```

rubric = self._resolve_rubric(snapshot)
answer_key = snapshot.model_answer or snapshot.answer_key or ""

rubric_weights = rubric["dimensions"]
rubric_lines = [
    f"- {dim}: weight {weight:.2f}"
    for dim, weight in rubric_weights.items()
]

reference_section = (
    f"\nReference answer (for rubric alignment):\n{answer_key}\n"
    if answer_key
    else ""
)

rubric_section = (
    "\nRubric weights:\n" + "\n".join(rubric_lines) + "\n"
    if rubric_weights
    else ""
)

user_prompt = f"""Question: {snapshot.prompt}

Student's Essay Answer:
{essay_text}

Please score this essay according to the rubric dimensions.
{reference_section}{rubric_section}"""

messages = [
    {"role": "system", "content": ESSAY_SCORING_SYSTEM_PROMPT},
    {"role": "user", "content": user_prompt},
]

gpt_response = await self.gpt_client.chat_completion(
    messages=messages,
    temperature=0.0,
    max_tokens=500,
)

```

Gambar 4. 9 Cuplikan kode validasi dan perhitungan skor akhir esai berbasis rubrik di sisi backend

Berdasarkan Gambar 4.9, implementasi penilaian esai menunjukkan bahwa backend berperan sebagai pengendali akhir dalam perhitungan skor. Model GPT memberikan penilaian

per dimensi, tetapi skor akhir tetap dihitung kembali melalui mekanisme yang lebih terstruktur berdasarkan bobot rubrik. Selain itu, keluaran model diperiksa agar sesuai dengan struktur JSON yang diharapkan sebelum digunakan dalam proses penilaian. Dengan mekanisme ini, proses penilaian esai menjadi lebih konsisten, lebih mudah diaudit, dan tidak sepenuhnya bergantung pada keluaran mentah dari model.

Pada implementasi ini, keluaran JSON diarahkan melalui kontrak keluaran pada prompt dan divalidasi kembali oleh parser backend. Nilai akhir tidak langsung diambil dari GPT, tetapi dihitung kembali berdasarkan bobot rubrik, batas skor, dan validasi struktur keluaran. Implementasi ini juga mendukung keterlacakan hasil karena sistem menyimpan rincian penilaian dan dapat menunjukkan bagaimana skor akhir terbentuk dari beberapa dimensi. Dengan demikian, penilaian esai tidak hanya menghasilkan angka akhir, tetapi juga memiliki dasar evaluasi yang dapat diperiksa kembali.

4.3.7 Implementasi Penggabungan Hasil dan Keterlacakan

Setelah proses penilaian berbasis aturan, penilaian esai, dan rekomendasi RAG selesai, sistem menjalankan proses penggabungan hasil / *fusion*. Proses ini menyatukan skor teori, skor esai, skor akhir, tingkat kompetensi, wawasan hasil asesmen, serta rekomendasi kursus ke dalam satu keluaran yang dapat ditampilkan kepada pengguna.

Implementasi *fusion* diperlukan agar hasil asesmen dan rekomendasi tidak berdiri sebagai data terpisah. Melalui proses ini, pengguna dapat melihat hubungan antara skor asesmen, profil kompetensi, dan rekomendasi kursus yang diberikan. Dengan kata lain, sistem tidak hanya memberi daftar kursus, tetapi juga mengaitkan rekomendasi dengan hasil asesmen pengguna.

Keterlacakan hasil diimplementasikan melalui penyimpanan beberapa informasi penting. Pada komponen rekomendasi, sistem menyimpan *relevance score*, *match_reason*, *rag_query*, dan *rag_traces*. Pada komponen penilaian esai, sistem menyimpan rincian skor dan hasil perhitungan berbasis rubrik. Informasi tersebut membantu proses audit karena pengelola dapat menelusuri kembali dasar penilaian dan dasar rekomendasi yang diberikan kepada pengguna.

Dengan implementasi ini, MicroCred GenAI Backend tidak memosisikan keluaran sebagai hasil kotak hitam. Setiap rekomendasi memiliki alasan kecocokan, setiap skor memiliki rincian perhitungan, dan setiap proses utama memiliki jejak yang dapat digunakan untuk pemeriksaan lanjutan.

4.3.8 Implementasi Umpan Balik Pengguna

Selain menampilkan hasil asesmen dan rekomendasi, sistem juga mengimplementasikan mekanisme umpan balik pengguna. Melalui endpoint POST `/assessments/{id}/feedback`, pengguna dapat memberikan penilaian terhadap relevansi rekomendasi, penerimaan rekomendasi, serta komentar kualitatif mengenai hasil yang diterima.

Implementasi umpan balik ini penting karena sistem dikembangkan sebagai MVP yang masih membutuhkan validasi awal dari pengguna. Data umpan balik tidak digunakan untuk mengklaim generalisasi luas, tetapi digunakan sebagai indikator awal untuk membaca apakah rekomendasi yang diberikan dianggap relevan, dapat diterima, dan mudah dipahami oleh pengguna.

Umpan balik pengguna kemudian menjadi salah satu sumber evaluasi pada Bagian 4.4. Dengan demikian, implementasi fitur umpan balik menghubungkan bagian implementasi sistem dengan bagian pengujian dan evaluasi.

4.3.9 Ringkasan Hasil Implementasi

Berdasarkan hasil implementasi yang telah dijelaskan, MicroCred GenAI Backend berhasil dibangun sebagai sistem yang menghubungkan asesmen, penilaian esai berbasis GPT, rekomendasi berbasis RAG, penggabungan hasil, keterlacakan, dan umpan balik pengguna. Tabel 4.4 menunjukkan bahwa kebutuhan fungsional utama telah direalisasikan melalui endpoint atau layanan aplikasi. Gambar 4.2 sampai Gambar 4.5 menunjukkan bahwa antarmuka pengguna telah mendukung alur mulai asesmen, pengerjaan soal, pembacaan hasil, dan penerimaan rekomendasi. Sementara itu, Gambar 4.6 sampai Gambar 4.8 menunjukkan implementasi komponen teknis utama, yaitu pemicu pemrosesan asinkron, pemeringkatan RAG, dan perhitungan skor esai berbasis rubrik.

Hasil implementasi ini menjadi dasar bagi proses pengujian dan evaluasi sistem pada Bagian 4.4. Dengan susunan ini, pembaca dapat memahami terlebih dahulu sistem yang telah dibangun sebelum melihat hasil observasi operasional, uji coba pengguna, audit manual, dan skenario pengujian.

4.4 Hasil Pengujian dan Evaluasi Sistem

Bagian ini menyajikan hasil pengujian dan evaluasi terhadap MicroCred GenAI Backend setelah sistem berhasil diimplementasikan. Evaluasi dilakukan untuk melihat kelayakan awal sistem dari sisi operasional, penerimaan pengguna, kualitas rekomendasi, keselarasan penilaian

esai otomatis, keterlaksanaan hasil, serta keberhasilan skenario pengujian. Dengan demikian, bagian ini tidak lagi membahas rancangan atau proses implementasi, tetapi berfokus pada bukti hasil yang diperoleh setelah sistem diuji.

Pengujian dan evaluasi dilakukan melalui beberapa sumber data. Pertama, observasi operasional digunakan untuk membaca waktu penyelesaian, latensi GPT, penggunaan token, dan estimasi biaya pemrosesan. Kedua, uji coba pengguna terbatas digunakan untuk membaca respons awal pengguna terhadap relevansi dan penerimaan rekomendasi. Ketiga, audit manual digunakan untuk menilai kualitas rekomendasi dan keselarasan penilaian esai otomatis. Keempat, *snapshot dump* dan skenario pengujian digunakan untuk memeriksa penyelesaian proses asinkron, kondisi terdegradasi, serta keberhasilan alur ujung-ke-ujung / *end-to-end*.

4.4.1 Hasil Observasi Operasional

Observasi operasional dilakukan terhadap 147 asesmen selesai dalam jendela pengamatan satu hari. Observasi ini bertujuan untuk melihat kelayakan awal sistem dari sisi waktu penyelesaian, latensi pemanggilan GPT, penggunaan token, dan estimasi biaya pemrosesan. Ringkasan hasil observasi operasional ditunjukkan pada Tabel 4.5.

Tabel 4. 5 Hasil observasi operasional sistem

Metrik	Nilai
Jumlah asesmen selesai dalam jendela pengamatan	147 asesmen
Waktu penyelesaian / <i>turnaround time</i> median	13,21 detik
Latensi GPT median	1696,5 ms
Penggunaan token rata-rata	± 420 token per pembelajar

Berdasarkan Tabel 4.5, sistem menunjukkan kelayakan operasional awal yang cukup baik pada konteks pengujian MVP. Median waktu penyelesaian sebesar 13,21 detik menunjukkan bahwa alur asinkron dapat membantu menjaga proses asesmen tetap responsif meskipun sistem melibatkan proses GPT, RAG, dan *fusion*. Nilai ini menunjukkan bahwa pengguna tidak perlu menunggu proses berat diselesaikan dalam satu permintaan yang sama.

Latensi GPT median sebesar 1696,5 ms menunjukkan bahwa pemanggilan model memiliki kontribusi terhadap waktu pemrosesan, tetapi bukan satu-satunya komponen yang menentukan

keseluruhan waktu penyelesaian. Sistem masih menjalankan proses lain, seperti penyimpanan respons, penilaian berbasis aturan, pengambilan kandidat rekomendasi, pemeringkatan, dan penggabungan hasil. Dari sisi biaya, estimasi biaya per asesmen berada pada rentang USD 0.00017–0.00020. Nilai ini menunjukkan bahwa penggunaan GPT pada tahap MVP masih berada pada tingkat biaya yang rendah untuk skenario uji coba terbatas.

Namun, hasil pada Tabel 4.5 tetap perlu dipahami sebagai hasil observasi awal, bukan sebagai bukti uji beban produksi skala besar. Jumlah asesmen, durasi pengamatan, dan kondisi penggunaan masih terbatas pada konteks penelitian.

4.4.2 Hasil Uji Coba Pengguna Terbatas

Uji coba pengguna terbatas dilakukan terhadap 10 partisipan. Setiap partisipan menjalani tiga percobaan asesmen / *attempt*, sehingga total terdapat 30 percobaan asesmen. Identitas partisipan dianonimkan menggunakan kode U1 sampai U10. Uji coba ini digunakan untuk melihat respons awal pengguna terhadap hasil asesmen dan rekomendasi yang diberikan sistem. Kohor pengguna yang terlibat dalam uji coba ditunjukkan pada Tabel 4.6.

Tabel 4. 6 Kohor uji pengguna terbatas

Kode	Audit Rekomendasi	Audit AES	Umpan Balik
U1	Ya	Ya	Ya
U2	Ya	Ya	Ya
U3	Ya	Ya	Ya
U4	Ya	Ya	Ya
U5	Ya	Ya	Ya
U6	Ya	Ya	Ya
U7	Ya	Ya	Ya
U8	Ya	Ya	Ya
U9	Ya	Ya	Ya
U10	Ya	Ya	Ya

Berdasarkan Tabel 4.6, seluruh partisipan dalam uji coba terbatas memiliki data yang dapat digunakan untuk audit rekomendasi, audit penilaian esai otomatis / *Automated Essay Scoring* (AES), dan umpan balik pengguna. Hal ini membuat data evaluasi lebih konsisten karena hasil audit dan tanggapan pengguna berasal dari kohor yang sama. Ringkasan indikator hasil uji pengguna terbatas ditunjukkan pada Tabel 4.7.

Tabel 4. 7 Ringkasan indikator uji pengguna terbatas

Indikator	Nilai
Jumlah partisipan	10 partisipan
Jumlah percobaan asesmen / <i>attempt</i>	30 percobaan
Rata-rata penilaian relevansi rekomendasi	4,3 dari 5
Median penilaian relevansi rekomendasi	4,5 dari 5
Rata-rata penilaian penerimaan rekomendasi	4,3 dari 5
Median penilaian penerimaan rekomendasi	5 dari 5
Rasio penerimaan tinggi, yaitu rating ≥ 4	0,70

Berdasarkan Tabel 4.7, tanggapan awal pengguna terhadap rekomendasi sistem cenderung positif. Rata-rata relevansi dan penerimaan rekomendasi sebesar 4,3 dari 5 menunjukkan bahwa sebagian besar pengguna menilai rekomendasi cukup sesuai dan dapat diterima. Median penerimaan sebesar 5 dari 5 juga menunjukkan bahwa sebagian pengguna memberikan penilaian sangat baik terhadap hasil yang diterima.

Meskipun demikian, hasil uji pengguna ini belum dapat digeneralisasi secara luas karena jumlah partisipan masih terbatas. Hasil tersebut lebih tepat dibaca sebagai indikasi awal bahwa sistem dapat diterima oleh pengguna dalam skenario uji coba MVP.

Selain penilaian numerik, komentar pengguna juga dianalisis untuk melihat tema utama yang muncul dari umpan balik. Ringkasan tema komentar pengguna ditunjukkan pada Tabel 4.8.

Tabel 4. 8 Tema utama komentar pengguna

Tema	Parafrase Tanggapan	Makna bagi Sistem
Kejelasan rincian penilaian / <i>scoring breakdown</i>	Pengguna menilai hasil sistem cukup terstruktur, komponen skor mudah dibaca, dan ambang batas membantu memahami kesiapan belajar.	Transparansi hasil menjadi nilai tambah yang dirasakan pengguna.
Kebutuhan personalisasi lebih tajam	Beberapa pengguna merasa rekomendasi sudah cukup relevan, tetapi masih dapat dibuat lebih spesifik terhadap tujuan belajar.	Sistem perlu memperkuat konteks tujuan belajar, preferensi, dan riwayat pembelajaran pengguna.
Kebutuhan orientasi progres	Pengguna mengusulkan adanya linimasa belajar, arah perkembangan, atau penanda posisi belajar.	Sistem belum cukup jika hanya memberikan daftar rekomendasi tanpa arah perkembangan belajar yang lebih eksplisit.

Berdasarkan Tabel 4.8, pengguna mengapresiasi struktur hasil dan rincian skor, tetapi masih mengharapkan personalisasi yang lebih tajam. Temuan ini menunjukkan bahwa sistem sudah membantu pengguna memahami hasil asesmen dan alasan rekomendasi, tetapi pengembangan berikutnya masih perlu memperkuat aspek jalur belajar, orientasi progres, dan kesesuaian rekomendasi terhadap tujuan belajar yang lebih spesifik.

4.4.3 Hasil Audit Manual Kualitas Rekomendasi

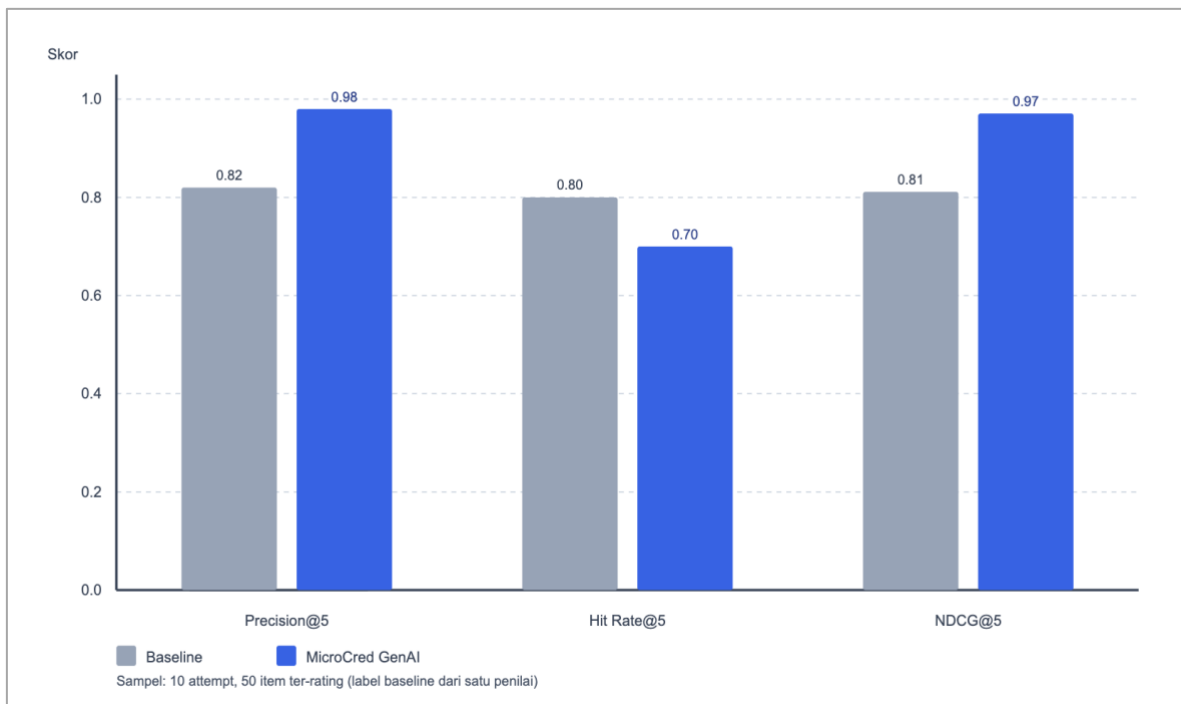
Audit manual kualitas rekomendasi dilakukan untuk menilai kualitas rekomendasi yang dihasilkan MicroCred GenAI dibandingkan dengan tolok ukur awal / *baseline*. Audit ini menggunakan tiga metrik, yaitu *Precision@5*, *HitRate@5*, dan *NDCG@5*. Hasil audit kualitas rekomendasi ditunjukkan pada Tabel 4.9.

Tabel 4. 9 Hasil audit kualitas rekomendasi

Metrik	Baseline	MicroCred GenAI
<i>Precision@5</i>	0,82	0,98
<i>HitRate@5</i>	0,80	0,70
<i>NDCG@5</i>	0,8114	0,9709

Berdasarkan Tabel 4.9, MicroCred GenAI memperoleh nilai *Precision@5* dan *NDCG@5* yang lebih tinggi dibandingkan baseline. Nilai *Precision@5* sebesar 0,98 menunjukkan bahwa sebagian besar rekomendasi pada lima posisi teratas dinilai relevan. Nilai *NDCG@5* sebesar 0,9709 juga menunjukkan bahwa urutan rekomendasi yang dihasilkan sistem cenderung menempatkan item yang lebih relevan pada posisi atas.

Perbandingan kualitas rekomendasi antara baseline dan MicroCred GenAI divisualisasikan pada Gambar 4.10.



Gambar 4. 10 Perbandingan kualitas rekomendasi antara baseline dan MicroCred GenAI

Berdasarkan Gambar 4.10, terlihat bahwa MicroCred GenAI lebih unggul pada *Precision@5* dan *NDCG@5*. Hal ini menunjukkan bahwa sistem lebih kuat dalam menghasilkan rekomendasi yang tepat pada posisi atas dan mengurutkan rekomendasi relevan secara lebih baik. Namun, nilai *HitRate@5* MicroCred GenAI sebesar 0,70 lebih rendah dibandingkan baseline sebesar 0,80. Temuan ini menunjukkan bahwa meskipun rekomendasi yang muncul cenderung lebih tepat dan terurut dengan baik, cakupan temuan relevan pada lima rekomendasi teratas masih dapat ditingkatkan.

Dengan demikian, hasil audit rekomendasi perlu dibaca secara seimbang. MicroCred GenAI menunjukkan peningkatan pada ketepatan rekomendasi dan kualitas peringkat, tetapi masih memiliki ruang perbaikan pada cakupan rekomendasi, keberagaman kandidat, dan personalisasi hasil.

4.4.4 Hasil Audit Penilaian Esai Otomatis

Audit penilaian esai otomatis dilakukan untuk melihat keselarasan skor AES berbasis GPT dengan skor penilai manusia. Audit dilakukan terhadap 30 esai dari 10 partisipan uji. Metrik yang digunakan adalah *Mean Absolute Error* (MAE), korelasi Pearson, serta persentase selisih skor dalam rentang ± 5 dan ± 10 poin. Hasil audit keselarasan AES ditunjukkan pada Tabel 4.10.

Tabel 4. 10 Hasil audit keselarasan AES terhadap penilai manusia

Metrik	Nilai
Jumlah esai diaudit	30 esai
MAE rerata per partisipan	1,1998
Korelasi Pearson	0,9990
Selisih dalam ± 5 poin	100%
Selisih dalam ± 10 poin	100%

Berdasarkan Tabel 4.10, hasil penilaian esai otomatis menunjukkan keselarasan awal yang tinggi dengan penilaian manusia pada sampel uji terbatas. Nilai MAE sebesar 1,1998 menunjukkan bahwa rata-rata selisih absolut antara skor AES dan skor manusia relatif kecil. Korelasi Pearson sebesar 0,9990 menunjukkan hubungan linear yang sangat kuat antara kedua sumber skor pada data yang diuji.

Perbandingan rerata skor AES dan rerata skor penilai manusia divisualisasikan pada Gambar 4.11.



Gambar 4. 11 Perbandingan rerata skor AES dan rerata skor penilai manusia

Berdasarkan Gambar 4.11, pola skor AES terlihat mengikuti pola skor penilai manusia pada sampel uji. Hal ini mendukung temuan pada Tabel 4.10 bahwa penilaian esai otomatis

memiliki kedekatan awal dengan penilaian manusia. Namun, hasil ini tetap perlu dibaca secara hati-hati karena jumlah esai yang diaudit masih terbatas dan belum melibatkan banyak penilai manusia untuk mengukur konsistensi antar-penilai.

Dengan demikian, penilaian esai otomatis pada penelitian ini dapat diposisikan sebagai pendekatan yang layak untuk mendukung MVP, tetapi belum dapat dipahami sebagai pengganti validasi manusia secara menyeluruh. Evaluasi lanjutan dengan jumlah esai yang lebih besar dan penilai manusia yang lebih banyak masih diperlukan.

4.4.5 Evidensi Pendukung dari Snapshot Dump

Selain observasi operasional dan uji coba pengguna terbatas, evaluasi juga didukung oleh analisis *snapshot dump* lokal. Analisis ini digunakan untuk memeriksa data sistem yang tersimpan, terutama pada aspek penyelesaian asesmen, keberhasilan pekerjaan asinkron, kondisi terdegradasi, percobaan ulang, dan duplikasi data. Ringkasan metrik pendukung dari *snapshot dump* ditunjukkan pada Tabel 4.11.

Tabel 4. 11 Metrik pendukung dari snapshot dump lokal

Metrik	Nilai
Jumlah asesmen pada snapshot	152 asesmen
Pekerjaan GPT selesai	152 dari 152
Pekerjaan RAG selesai	152 dari 152
Pekerjaan <i>fusion</i> selesai	152 dari 152
Tingkat penyelesaian pekerjaan GPT, RAG, dan <i>fusion</i>	100%
Total percobaan ulang / <i>retry</i> yang tercatat	15 kali
Percobaan ulang / <i>retry</i> pada jalur GPT	13 kali
Kondisi terdegradasi / <i>degraded</i> atau <i>fallback</i>	29 kasus
Duplikasi <i>submit</i> yang terdeteksi	0 kasus

Catatan: Artefak pendukung snapshot dump tersedia pada tautan berikut: https://drive.google.com/drive/folders/1AY2ePQGPI7QFYF7YD-mP_RZa-rhQP6kt?usp=sharing

Berdasarkan Tabel 4.11, seluruh pekerjaan GPT, RAG, dan *fusion* pada *snapshot dump* berhasil selesai. Hal ini menunjukkan bahwa alur pemrosesan asinkron dapat menyelesaikan pekerjaan utama sistem pada data yang diamati. Adanya 15 percobaan ulang, dengan 13 di antaranya terjadi pada jalur GPT, menunjukkan bahwa sistem menghadapi kegagalan sementara, tetapi mekanisme *retry* membantu proses tetap dapat dilanjutkan.

Selain itu, terdapat 29 kondisi terdegradasi atau *fallback*. Kondisi ini menunjukkan bahwa sistem tidak selalu berjalan dalam kondisi ideal, tetapi tetap memiliki mekanisme cadangan untuk menghasilkan keluaran ketika proses tertentu mengalami gangguan. Tidak ditemukannya duplikasi *submit* juga menunjukkan bahwa mekanisme idempotensi berperan dalam menjaga konsistensi data.

Hasil pada Tabel 4.11 memperkuat bahwa sistem telah memiliki dasar reliabilitas awal. Namun, jumlah kondisi terdegradasi tetap perlu diperhatikan sebagai ruang perbaikan, terutama pada stabilitas layanan eksternal, strategi *retry*, dan pengendalian kondisi *fallback*.

4.4.6 Hasil Skenario Pengujian Sistem

Pengujian skenario dilakukan untuk memeriksa apakah fungsi utama yang telah dirancang pada BAB III dapat berjalan pada sistem yang telah diimplementasikan. Skenario diuji melalui kombinasi pengamatan langsung pada lingkungan *livedan* pembacaan data dari *snapshot dump*. Pemetaan skenario pengujian terhadap bukti ditunjukkan pada Tabel 4.12.

Tabel 4. 12 Pemetaan skenario pengujian terhadap bukti

No.	Skenario Pengujian	Bukti	Status
1	Mulai asesmen	Lingkungan <i>live</i>	Teramati langsung
2	<i>Submit</i> normal	Lingkungan <i>live</i>	Teramati langsung
3	Idempotensi / <i>idempotency</i>	<i>Snapshot dump</i> menunjukkan 0 duplikasi <i>submit</i>	Didukung data snapshot
4	Pekerjaan GPT asinkron sukses	<i>Snapshot dump</i> menunjukkan 152/152 selesai	Didukung data snapshot
5	GPT <i>timeout</i> dan <i>retry</i>	<i>Snapshot dump</i> menunjukkan 13 <i>retry</i> GPT	Didukung data snapshot
6	RAG sukses	<i>Snapshot dump</i> menunjukkan 152/152 selesai	Didukung data snapshot
7	RAG <i>degraded</i> / <i>fallback</i>	<i>Snapshot dump</i> menunjukkan 29 kondisi <i>degraded</i> atau <i>fallback</i>	Didukung data snapshot
8	<i>Fusion</i> sukses	<i>Snapshot dump</i> menunjukkan 152/152 selesai	Didukung data snapshot
9	Pemantauan status	Lingkungan <i>live</i> melalui <i>status polling</i>	Teramati langsung
10	Pengambilan hasil	Lingkungan <i>live</i> melalui <i>payload</i> hasil	Teramati langsung
11	Pengiriman umpan balik	Lingkungan <i>live</i> dengan respons HTTP 201	Teramati langsung

Berdasarkan Tabel 4.12, seluruh skenario pengujian utama memiliki bukti, baik melalui pengamatan langsung maupun melalui data *snapshot*. Skenario yang berkaitan dengan interaksi pengguna, seperti memulai asesmen, mengirim jawaban, memantau status, mengambil hasil, dan mengirim umpan balik, teramati langsung pada lingkungan *live*. Sementara itu, skenario yang berkaitan dengan stabilitas pemrosesan, seperti GPT, RAG, *fusion*, *retry*, *degraded mode*, dan idempotensi, didukung oleh data dari *snapshot dump*. Selain skenario umum, pengujian juga dilakukan terhadap tiga skenario representatif berdasarkan kualitas input pengguna, yaitu input berkualitas campuran / *mixed-quality*, input berkualitas tinggi / *high-quality*, dan input berkualitas rendah / *low-quality*. Hasil skenario representatif ujung-ke-ujung ditunjukkan pada Tabel 4.13.

Tabel 4. 13 Hasil skenario representatif ujung-ke-ujung

Skenario	Karakteristik Input	Hasil yang Diharapkan	Hasil Pengujian
<i>Mixed-quality</i>	Jawaban berisi sebagian informasi relevan, tetapi masih memiliki kekurangan pada kedalaman atau kelengkapan.	Sistem menghasilkan skor menengah dan rekomendasi yang membantu pengguna memperkuat kompetensi.	Sistem dapat menyelesaikan asesmen, menghasilkan skor, rekomendasi, dan alasan rekomendasi.
<i>High-quality</i>	Jawaban relatif lengkap, relevan, dan menunjukkan pemahaman yang lebih baik.	Sistem menghasilkan skor lebih tinggi dan rekomendasi yang lebih dekat dengan jalur target.	Sistem dapat menghasilkan rekomendasi berkualitas, tetapi cakupan target masih dapat diperluas.
<i>Low-quality</i>	Jawaban terbatas, kurang lengkap, atau tidak cukup menunjukkan pemahaman.	Sistem mengarahkan pengguna pada penguatan dasar atau jalur fondasi.	Sistem dapat menyelesaikan alur dan memberikan rekomendasi yang lebih bersifat fondasional.

Berdasarkan Tabel 4.13, sistem dapat menjalankan alur ujung-ke-ujung pada variasi kualitas input yang berbeda. Sistem mampu menghasilkan skor, rekomendasi, alasan kecocokan, dan data keterlacakan pada setiap skenario. Pada skenario input berkualitas rendah, sistem tetap dapat menyelesaikan alur dan menghasilkan rekomendasi yang lebih bersifat fondasional. Pada skenario input berkualitas tinggi, sistem dapat menghasilkan rekomendasi yang lebih dekat dengan jalur target, meskipun cakupan rekomendasi masih dapat diperluas.

Dengan demikian, hasil skenario pengujian menunjukkan bahwa sistem dapat menjalankan fungsi utama sesuai rancangan. Namun, hasil ini tetap diposisikan sebagai pengujian awal karena variasi skenario masih terbatas dan belum mencakup seluruh kemungkinan perilaku pengguna.

4.4.7 Ringkasan Hasil Pengujian dan Evaluasi

Berdasarkan hasil pengujian dan evaluasi, MicroCred GenAI Backend menunjukkan kelayakan awal sebagai sistem asesmen dan rekomendasi kredensial mikro berbasis GenAI. Dari sisi operasional, sistem mampu menyelesaikan 147 asesmen dalam jendela pengamatan dengan median waktu penyelesaian 13,21 detik dan estimasi biaya per asesmen yang rendah. Dari sisi pengguna, rata-rata relevansi dan penerimaan rekomendasi berada pada nilai 4,3 dari 5, yang menunjukkan respons awal yang positif terhadap keluaran sistem.

Dari sisi kualitas hasil, audit rekomendasi menunjukkan peningkatan pada *Precision@5* dan *NDCG@5* dibandingkan baseline, meskipun *HitRate@5* masih lebih rendah. Audit AES juga menunjukkan keselarasan awal yang tinggi terhadap penilai manusia pada sampel terbatas. Sementara itu, *snapshot dump* dan skenario pengujian menunjukkan bahwa alur GPT, RAG, dan *fusion* dapat diselesaikan serta didukung oleh mekanisme *retry*, *degraded mode*, *fallback*, dan idempotensi.

Dengan demikian, hasil pengujian menunjukkan bahwa sistem telah memenuhi tujuan utama MVP, yaitu mendukung asesmen yang lebih cepat, menghasilkan rekomendasi yang relevan pada peringkat atas, dan menyediakan dasar hasil yang dapat ditelusuri. Namun, hasil ini tetap perlu dibaca sebagai evaluasi awal karena jumlah partisipan, cakupan dataset, dan skenario pengujian masih terbatas.

4.5 Pembahasan

Bagian ini membahas hasil penelitian secara terpadu berdasarkan hasil analisis kebutuhan, perancangan, implementasi, serta pengujian dan evaluasi sistem yang telah disajikan pada bagian sebelumnya. Pembahasan tidak diarahkan untuk mengulang seluruh angka evaluasi, tetapi untuk menafsirkan makna hasil tersebut terhadap pertanyaan penelitian.

Pembahasan difokuskan pada tiga arah utama, yaitu kemampuan sistem dalam mendukung pemrosesan asesmen yang lebih cepat, kemampuan mekanisme retrieval-augmented generation (RAG) dengan pemeringkatan hibrida ringan dalam menghasilkan rekomendasi pembelajaran yang relevan, serta kemampuan sistem dalam menyediakan hasil asesmen dan

rekomendasi yang lebih jelas serta dapat ditelusuri. Selain itu, bagian ini juga membahas implikasi hasil penelitian dan batas interpretasi temuan agar hasil penelitian tidak dibaca secara berlebihan di luar ruang lingkup evaluasi awal.

Untuk memperjelas hubungan antara pertanyaan penelitian dan hasil yang diperoleh, pemetaan hasil penelitian ditunjukkan pada Tabel 4.14.

Tabel 4. 14 Pemetaan hasil penelitian terhadap pertanyaan penelitian

Fokus Penelitian	Pertanyaan	Hasil yang Berkaitan	Makna Pembahasan
Rancangan dan implementasi alur asesmen asinkron		Median waktu penyelesaian 13,21 detik pada 147 asesmen selesai, serta 152/152 pekerjaan GPT, RAG, dan fusion selesai pada snapshot dump.	Alur asinkron mendukung pemrosesan asesmen yang lebih responsif pada tahap MVP.
Mekanisme RAG dan relevansi rekomendasi		Rata-rata relevansi dan penerimaan pengguna 4,3 dari 5, Precision@5 sebesar 0,98, dan NDCG@5 sebesar 0,9709.	RAG dengan pemeringkatan hibrida ringan dapat menghasilkan rekomendasi yang cukup relevan pada uji terbatas.
Keterjelasan dan keterlacakan hasil		Tersedianya scoring breakdown, match_reason, rag_query, rag_traces, serta audit AES dengan MAE 1,1998 dan korelasi Pearson 0,9990.	Sistem tidak hanya menghasilkan skor dan rekomendasi, tetapi juga menyediakan dasar hasil yang dapat diperiksa kembali.
Batas interpretasi hasil		Partisipan uji masih terbatas, audit esai menggunakan 30 esai, dan dataset kursus masih berbasis sumber lokal.	Hasil penelitian perlu diposisikan sebagai evaluasi awal, bukan sebagai generalisasi menyeluruh.

Berdasarkan Tabel 4.14, hasil penelitian menunjukkan bahwa MicroCred GenAI Backend telah memenuhi fungsi utama sebagai minimum viable product (MVP). Sistem mampu menjalankan alur asesmen, memproses penilaian esai, menghasilkan rekomendasi kursus, dan menyimpan dasar keterlacakan hasil. Namun, hasil tersebut tetap perlu dibaca secara hati-hati karena evaluasi masih dilakukan pada skala terbatas.

4.5.1 Responsivitas Alur Asesmen Asinkron

Hasil observasi operasional pada Tabel 4.5 menunjukkan bahwa sistem mampu menyelesaikan 147 asesmen dengan median waktu penyelesaian 13,21 detik. Selain itu, Tabel

4.11 menunjukkan bahwa seluruh pekerjaan GPT, RAG, dan fusion pada snapshot dump berhasil mencapai status selesai. Hasil ini menunjukkan bahwa pemisahan proses sinkron dan asinkron menjadi keputusan rancangan yang tepat untuk tahap MVP.

Pada alur sistem yang dikembangkan, proses sinkron digunakan untuk menangani aktivitas yang harus segera diproses, seperti menerima jawaban pengguna, menyimpan respons, dan menghitung komponen penilaian berbasis aturan. Sementara itu, proses yang lebih berat, seperti penilaian esai berbasis GPT, pembentukan rekomendasi RAG, dan penggabungan hasil akhir, dijalankan secara asinkron. Pemisahan ini membuat pengguna tidak harus menunggu seluruh proses berat selesai pada satu permintaan utama.

Hubungan antara rancangan dan hasil tersebut juga didukung oleh Gambar 4.6 yang menunjukkan pemicu alur pemrosesan asinkron setelah pengguna mengirim asesmen. Dengan rancangan tersebut, backend dapat menerima asesmen terlebih dahulu, lalu meneruskan pekerjaan lanjutan ke proses latar. Tabel 4.12 dan Tabel 4.13 juga menunjukkan bahwa skenario pengujian utama, termasuk alur ujung-ke-ujung, dapat dijalankan sesuai rancangan.

Meskipun demikian, hasil ini belum cukup untuk menyimpulkan bahwa sistem telah siap digunakan pada skala besar. Evaluasi dilakukan pada jendela pengamatan dan jumlah data yang terbatas. Oleh karena itu, hasil responsivitas ini lebih tepat dimaknai sebagai bukti awal bahwa rancangan asinkron layak digunakan untuk MVP dan dapat menjadi dasar pengujian beban yang lebih besar pada pengembangan berikutnya.

4.5.2 Kualitas Penilaian Esai Berbasis GPT

Penilaian esai berbasis GPT menjadi salah satu komponen penting karena sistem tidak hanya menilai jawaban objektif, tetapi juga jawaban terbuka. Gambar 4.8 menunjukkan bahwa prompt penilaian dibentuk dengan memisahkan system prompt dan user prompt. System prompt digunakan untuk menetapkan peran model sebagai evaluator esai, sedangkan user prompt membawa data spesifik berupa pertanyaan, jawaban pengguna, jawaban acuan apabila tersedia, serta bobot rubrik penilaian.

Pemisahan tersebut penting karena penilaian esai tidak seharusnya bergantung pada respons bebas model. Model diarahkan untuk menilai berdasarkan konteks yang diberikan sistem, bukan berdasarkan pengetahuan eksternal atau interpretasi yang tidak terkendali. Dengan demikian, rubrik tidak hanya menjadi narasi dalam laporan, tetapi benar-benar menjadi bagian dari masukan yang dikirimkan ke GPT.

Gambar 4.9 menunjukkan bahwa backend tetap melakukan validasi dan perhitungan skor akhir setelah model menghasilkan keluaran. Artinya, GPT berperan sebagai evaluator awal untuk memberi skor per dimensi, sedangkan backend berperan sebagai pengendali akhir yang memeriksa struktur keluaran, membatasi nilai pada rentang yang ditentukan, dan menghitung ulang skor berdasarkan bobot rubrik. Pendekatan ini mengurangi risiko keluaran model digunakan secara mentah tanpa pemeriksaan.

Hasil audit pada Tabel 4.10 menunjukkan bahwa penilaian esai otomatis memiliki kedekatan yang tinggi terhadap penilai manusia pada sampel terbatas. Gambar 4.11 juga memperlihatkan perbandingan rerata skor AES dan skor penilai manusia. Hasil tersebut menunjukkan bahwa pendekatan penilaian berbasis rubrik dengan validasi backend dapat menghasilkan keluaran yang cukup selaras pada konteks pengujian ini.

Namun, hasil tersebut tetap perlu dibaca secara terbatas. Audit hanya dilakukan pada 30 esai, sehingga belum cukup untuk menyimpulkan bahwa kualitas penilaian akan konsisten pada variasi jawaban, topik, atau tingkat kesulitan yang lebih luas. Selain itu, korelasi yang tinggi tidak selalu berarti tidak ada kesalahan penilaian pada level individual. Oleh karena itu, evaluasi lanjutan tetap diperlukan dengan jumlah sampel yang lebih besar, variasi soal yang lebih luas, dan pembandingan penilai manusia yang lebih dari satu orang.

4.5.3 Relevansi Rekomendasi Berbasis RAG dengan Pemeringkatan Hibrida Ringan

Hasil evaluasi rekomendasi menunjukkan bahwa mekanisme RAG dengan pemeringkatan hibrida ringan mampu menghasilkan rekomendasi yang cukup relevan pada uji terbatas. Tabel 4.9 menunjukkan Precision@5 sebesar 0,98 dan NDCG@5 sebesar 0,9709. Nilai tersebut menunjukkan bahwa sebagian besar rekomendasi pada lima posisi teratas dinilai relevan dan urutannya cukup baik berdasarkan audit manual.

Hasil tersebut sejalan dengan rancangan sistem yang tidak menghasilkan rekomendasi secara bebas dari model bahasa, tetapi mengambil kandidat dari dataset kursus yang tersedia. Gambar 4.7 menunjukkan bahwa skor akhir kandidat kursus dibentuk dari beberapa sinyal, seperti relevansi kata kunci, kualitas kursus, kecocokan metadata, feature hashing, tag boost, dan faktor pendukung lain yang berkaitan dengan profil pengguna. Dengan demikian, rekomendasi tidak hanya didasarkan pada satu indikator, tetapi merupakan hasil penggabungan beberapa sinyal relevansi.

Pada konteks kredensial mikro, pendekatan ini penting karena pengguna tidak hanya membutuhkan daftar kursus, tetapi juga membutuhkan rekomendasi yang sesuai dengan

tingkat kesiapan, minat, dan kebutuhan kompetensi. Hasil uji pengguna pada Tabel 4.7 menunjukkan rata-rata relevansi dan penerimaan sebesar 4,3 dari 5, sehingga rekomendasi yang diberikan dinilai cukup dapat diterima oleh pengguna pada skenario terbatas.

Namun, nilai $\text{HitRate}@5$ yang lebih rendah dibandingkan $\text{Precision}@5$ dan $\text{NDCG}@5$ menunjukkan bahwa masih terdapat ruang perbaikan pada cakupan rekomendasi. Hal ini dapat mengindikasikan bahwa sistem sudah cukup baik ketika berhasil menemukan kandidat relevan, tetapi belum selalu mampu menjamin bahwa seluruh kebutuhan pengguna terwakili dalam lima rekomendasi teratas. Keterbatasan ini dapat berkaitan dengan cakupan dataset, variasi tag, kualitas metadata, atau kemampuan sistem dalam menangkap kebutuhan pengguna dari hasil asesmen.

Dengan demikian, mekanisme RAG yang dikembangkan dapat dinilai layak sebagai fondasi awal sistem rekomendasi pembelajaran. Namun, peningkatan dataset, normalisasi metadata, perluasan tag, dan pengujian pada skenario pengguna yang lebih beragam tetap diperlukan agar rekomendasi lebih stabil dan tidak terlalu bergantung pada cakupan dataset awal.

4.5.4 Keterjelasan dan Keterlacakan Hasil

Salah satu fokus penelitian ini adalah menghasilkan sistem yang tidak hanya memberi skor dan daftar rekomendasi, tetapi juga menyediakan dasar yang dapat diperiksa kembali. Hasil implementasi menunjukkan bahwa sistem menyimpan beberapa komponen pendukung keterjelasan, seperti `scoring breakdown`, `match_reason`, `rag_query`, dan `rag_traces`.

Gambar 4.4 menunjukkan bahwa halaman mulai asesmen menampilkan rincian penilaian atau `scoring breakdown`. Informasi ini membantu pengguna memahami bahwa skor akhir tidak muncul sebagai angka tunggal tanpa dasar, tetapi berasal dari beberapa komponen penilaian. Gambar 4.5 juga menunjukkan bahwa daftar rekomendasi kursus dilengkapi nilai kecocokan atau `match score`. Dengan informasi tersebut, pengguna dapat melihat bahwa setiap rekomendasi memiliki tingkat kecocokan tertentu terhadap profil atau hasil asesmen.

Selain tampilan kepada pengguna, keterlacakan juga diperkuat pada sisi backend melalui penyimpanan `rag_query` dan `rag_traces`. Komponen tersebut penting karena rekomendasi berbasis RAG perlu memiliki dasar pengambilan yang dapat ditinjau ulang. Dengan adanya jejak pengambilan, sistem dapat menunjukkan bahwa rekomendasi berasal dari kandidat dalam basis pengetahuan, bukan dari keluaran generatif yang tidak memiliki sumber operasional.

Hasil pada Tabel 4.11 menunjukkan bahwa data pendukung dari snapshot dump tersedia untuk membaca status pekerjaan GPT, RAG, dan fusion. Hal ini memperkuat posisi sistem sebagai backend yang tidak hanya menjalankan proses, tetapi juga menyimpan informasi yang dapat digunakan untuk audit teknis. Dengan demikian, keterlacakan pada penelitian ini tidak hanya dipahami sebagai fitur tampilan, tetapi juga sebagai kemampuan sistem untuk menyimpan jejak proses dan dasar keluaran.

Meskipun demikian, keterlacakan yang dikembangkan masih berada pada tingkat awal. Sistem sudah menyimpan dasar rekomendasi dan alasan kecocokan, tetapi belum sepenuhnya menyediakan antarmuka audit yang lengkap untuk dosen, admin, atau evaluator eksternal. Pada pengembangan berikutnya, keterlacakan dapat diperkuat melalui dashboard audit, riwayat perubahan rekomendasi, log versi dataset, serta penjelasan yang lebih eksplisit mengenai kontribusi setiap sinyal terhadap skor akhir rekomendasi.

4.5.5 Implikasi Hasil Penelitian

Hasil penelitian ini memiliki beberapa implikasi. Pertama, dari sisi pengembangan sistem, penggunaan arsitektur backend berbasis modular monolith cukup sesuai untuk tahap MVP. Pendekatan ini memungkinkan modul asesmen, penilaian esai, rekomendasi, dan fusion dipisahkan secara logis tanpa menambah kompleksitas infrastruktur yang terlalu besar pada tahap awal. Dengan demikian, sistem masih mudah dikembangkan, diuji, dan diperbaiki.

Kedua, dari sisi asesmen pembelajaran, integrasi penilaian berbasis aturan dan GPT menunjukkan bahwa asesmen kredensial mikro dapat dilakukan secara lebih fleksibel. Pertanyaan objektif dan profil dapat dihitung secara deterministik, sedangkan jawaban esai dapat dinilai menggunakan rubrik melalui GPT. Kombinasi ini memberi ruang bagi sistem untuk menilai aspek pengetahuan, pemahaman, dan ekspresi jawaban pengguna secara lebih luas dibandingkan asesmen objektif saja.

Ketiga, dari sisi rekomendasi pembelajaran, penggunaan RAG dengan pemeringkatan hibrida ringan menunjukkan bahwa personalisasi tidak harus langsung bergantung pada collaborative filtering atau data interaksi pengguna dalam jumlah besar. Pada kondisi cold-start, sistem masih dapat menggunakan hasil asesmen, profil pengguna, preferensi, kata kunci, dan metadata kursus sebagai dasar rekomendasi. Hal ini relevan untuk sistem kredensial mikro tahap awal yang belum memiliki riwayat interaksi pengguna yang besar.

Keempat, dari sisi keterlacakan, penyimpanan scoring breakdown, match_reason, rag_query, dan rag_traces memberi dasar bahwa hasil sistem dapat diperiksa kembali. Ini

penting karena sistem berbasis AI dalam pendidikan perlu menghindari keluaran yang tampak meyakinkan tetapi tidak memiliki dasar yang jelas. Dengan adanya informasi pendukung, rekomendasi dan penilaian dapat dijelaskan secara lebih terstruktur.

Secara keseluruhan, implikasi utama penelitian ini adalah bahwa MicroCred GenAI Backend dapat menjadi fondasi awal untuk layanan asesmen dan rekomendasi kredensial mikro yang cepat, relevan, dan lebih dapat ditelusuri. Namun, posisi tersebut masih berada pada tingkat MVP dan evaluasi awal, sehingga pengembangan lebih lanjut tetap diperlukan sebelum sistem digunakan pada skala institusional yang lebih luas.

4.5.6 Keterbatasan Interpretasi Temuan

Hasil penelitian ini perlu dibaca dalam batas ruang lingkup yang jelas. Pertama, jumlah partisipan uji pengguna masih terbatas, yaitu 10 partisipan. Jumlah tersebut cukup untuk memperoleh gambaran awal mengenai penerimaan dan relevansi rekomendasi, tetapi belum cukup untuk mewakili seluruh variasi karakteristik mahasiswa, latar belakang kompetensi, dan kebutuhan pembelajaran.

Kedua, audit penilaian esai dilakukan pada 30 esai. Jumlah ini memberikan indikasi awal mengenai keselarasan AES terhadap penilai manusia, tetapi belum cukup untuk menilai konsistensi penilaian pada skala besar. Variasi gaya bahasa, panjang jawaban, kualitas argumentasi, dan jenis soal dapat memengaruhi hasil penilaian. Oleh karena itu, evaluasi lanjutan perlu menggunakan sampel yang lebih besar dan melibatkan lebih dari satu penilai manusia agar reliabilitas pembandingan lebih kuat.

Ketiga, dataset kursus yang digunakan masih terbatas pada himpunan data yang dikurasi dalam penelitian. Kualitas rekomendasi sangat bergantung pada kualitas metadata, kelengkapan tag, variasi topik, dan kecocokan data kursus terhadap kebutuhan pengguna. Jika dataset memiliki cakupan terbatas, sistem juga berpotensi menghasilkan rekomendasi yang terbatas meskipun mekanisme pemeringkatannya sudah berjalan dengan baik.

Keempat, evaluasi operasional dilakukan pada konteks MVP dan belum mencakup pengujian beban skala besar. Median waktu penyelesaian yang diperoleh menunjukkan indikasi responsivitas awal, tetapi belum dapat digunakan untuk menyimpulkan ketahanan sistem terhadap jumlah pengguna yang jauh lebih besar, variasi koneksi, antrean paralel, atau kegagalan layanan eksternal.

Kelima, penelitian ini belum mengukur dampak pembelajaran jangka panjang. Sistem telah menghasilkan asesmen dan rekomendasi, tetapi belum mengevaluasi apakah pengguna yang

mengikuti rekomendasi tersebut benar-benar mengalami peningkatan kompetensi. Oleh karena itu, penelitian lanjutan perlu mengukur dampak setelah pengguna mengikuti kursus atau jalur belajar yang direkomendasikan.

Dengan mempertimbangkan keterbatasan tersebut, hasil penelitian ini sebaiknya diposisikan sebagai bukti kelayakan awal, bukan sebagai klaim bahwa sistem telah optimal atau siap digeneralisasikan ke seluruh konteks pendidikan tinggi. Temuan penelitian menunjukkan bahwa rancangan dan implementasi MicroCred GenAI Backend memiliki arah yang menjanjikan, tetapi masih membutuhkan pengujian lanjutan, perluasan data, dan validasi yang lebih kuat.

4.6 Ringkasan

BAB IV telah menyajikan hasil penelitian dan pembahasan terhadap pengembangan MicroCred GenAI Backend. Bagian awal bab ini menjelaskan hasil analisis kebutuhan awal pengguna yang diperoleh dari formulir kebutuhan dan umpan balik pemangku kepentingan. Hasil tersebut menunjukkan bahwa pengguna membutuhkan sistem yang dapat membantu memilih kursus secara lebih terarah, memberikan asesmen yang jelas, serta menyediakan rekomendasi pembelajaran yang mudah dipahami.

Selanjutnya, hasil perancangan sistem menunjukkan bahwa MicroCred GenAI Backend dirancang sebagai layanan sisi peladen / *backend* yang menghubungkan proses asesmen, penilaian esai otomatis, rekomendasi berbasis RAG, dan keterlacakan hasil. Rancangan tersebut mencakup alur asesmen asinkron, mekanisme RAG dengan basis pengetahuan kursus, rancangan basis data, serta rancangan antarmuka pengguna sebagai media interaksi dalam uji coba MVP.

Hasil implementasi menunjukkan bahwa sistem berhasil dibangun sesuai rancangan utama. Sistem dapat menerima jawaban asesmen, menjalankan penilaian berbasis aturan, memproses penilaian esai berbasis GPT, menghasilkan rekomendasi kursus melalui mekanisme RAG, menggabungkan hasil melalui proses *fusion*, serta menyimpan skor, alasan rekomendasi, dan jejak pengambilan. Antarmuka pengguna juga digunakan untuk mendukung alur uji coba, mulai dari pengerjaan asesmen hingga tampilan hasil rekomendasi.

Hasil pengujian dan evaluasi menunjukkan bahwa sistem memiliki kelayakan awal sebagai MVP. Observasi operasional terhadap 147 asesmen selesai menunjukkan median waktu penyelesaian 13,21 detik dan biaya per asesmen yang rendah. Uji coba terbatas terhadap 10 partisipan menghasilkan rata-rata relevansi dan penerimaan rekomendasi sebesar 4,3 dari 5.

Audit rekomendasi menunjukkan nilai *Precision@5* sebesar 0,98 dan *NDCG@5* sebesar 0,9709, sedangkan audit penilaian esai otomatis menunjukkan MAE sebesar 1,1998 dan korelasi Pearson sebesar 0,9990 pada sampel uji terbatas.

Pembahasan menunjukkan bahwa MicroCred GenAI Backend mampu mendukung tiga arah utama penelitian, yaitu mempercepat proses asesmen melalui alur asinkron, menghasilkan rekomendasi yang cukup relevan melalui RAG dengan pemeringkatan hibrida ringan, serta menyediakan hasil yang lebih jelas dan dapat ditelusuri melalui skor, *match_reason*, *rag_query*, dan *rag_traces*. Meskipun demikian, hasil penelitian tetap perlu dibaca sebagai evaluasi awal karena jumlah partisipan masih terbatas, cakupan dataset belum luas, personalisasi masih dapat ditingkatkan, dan dampak pembelajaran jangka panjang belum diukur.

Dengan demikian, BAB IV menunjukkan bahwa MicroCred GenAI Backend telah memenuhi fungsi utama sebagai MVP untuk mendukung asesmen kredensial mikro dan rekomendasi pembelajaran berbasis GenAI. Hasil pada bab ini menjadi dasar untuk menarik kesimpulan, menjelaskan keterbatasan penelitian, dan merumuskan saran pengembangan lanjutan pada BAB V.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian, MicroCred GenAI Backend berhasil dikembangkan sebagai *minimum viable product* (MVP) untuk mendukung asesmen kredensial mikro dan rekomendasi pembelajaran berbasis GenAI. Kesimpulan penelitian disusun berdasarkan tiga pertanyaan penelitian sebagai berikut.

1. **Alur asesmen asinkron pada MicroCred GenAI Backend mampu mendukung pemrosesan asesmen yang lebih cepat pada tahap MVP.**
Hal ini dicapai melalui pemisahan proses sinkron dan asinkron. Proses sinkron digunakan untuk menerima jawaban, menyimpan respons, dan menjalankan penilaian berbasis aturan, sedangkan proses asinkron digunakan untuk menjalankan penilaian esai berbasis GPT, rekomendasi RAG, dan penggabungan hasil / *fusion*. Hasil observasi terhadap 147 asesmen menunjukkan median waktu penyelesaian sebesar 13,21 detik, sedangkan *snapshot dump* pada 152 asesmen menunjukkan seluruh pekerjaan GPT, RAG, dan *fusion* selesai 100%.
2. **Mekanisme RAG dengan pemeringkatan hibrida ringan mampu menghasilkan rekomendasi pembelajaran yang relevan terhadap profil kompetensi pembelajar pada uji terbatas.**
Sistem menggunakan hasil asesmen, profil pembelajar, preferensi, kata kunci esai, topik yang belum dikuasai, dan skor asesmen sebagai dasar rekomendasi. Hasil uji pengguna menunjukkan rata-rata relevansi dan penerimaan rekomendasi sebesar 4,3 dari 5. Audit manual juga menunjukkan *Precision@5* sebesar 0,98 dan *NDCG@5* sebesar 0,9709. Namun, nilai *HitRate@5* sebesar 0,70 menunjukkan bahwa cakupan rekomendasi masih perlu ditingkatkan.
3. **Mekanisme skor, alasan rekomendasi, dan jejak pengambilan mampu mendukung keterjelasan serta keterlacakan hasil asesmen dan rekomendasi.**
Sistem tidak hanya menampilkan skor akhir dan daftar kursus, tetapi juga menyimpan rincian skor, *match_reason*, *rag_query*, dan *rag_traces*. Pada komponen penilaian esai, keluaran GPT divalidasi dan dihitung kembali oleh *backend* berdasarkan rubrik. Audit terhadap 30 esai menunjukkan MAE sebesar 1,1998 dan korelasi Pearson sebesar 0,9990 terhadap penilai manusia pada sampel terbatas.

Secara keseluruhan, penelitian ini menunjukkan bahwa MicroCred GenAI Backend layak menjadi fondasi awal layanan asesmen dan rekomendasi kredensial mikro berbasis GenAI. Sistem mampu mempercepat proses asesmen, menghasilkan rekomendasi yang cukup relevan, dan menyediakan dasar hasil yang dapat ditelusuri. Namun, simpulan ini masih terbatas pada konteks evaluasi awal karena jumlah partisipan, cakupan dataset, dan skenario pengujian masih terbatas.

5.2 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, sistem yang dikembangkan masih berada pada tahap *minimum viable product* (MVP), sehingga fitur yang tersedia difokuskan pada fungsi inti, yaitu asesmen, pemrosesan jawaban esai, rekomendasi kursus, dan penyimpanan jejak rekomendasi. Dengan demikian, sistem belum mencakup seluruh kebutuhan ekosistem kredensial mikro secara penuh, seperti integrasi dengan sistem akademik, manajemen sertifikat, atau pemantauan perkembangan pembelajaran dalam jangka panjang.

Kedua, basis pengetahuan rekomendasi masih terbatas pada dataset kursus Udemy berbentuk CSV yang digunakan secara lokal. Field asli dataset mencakup informasi seperti *course_title*, *url*, *is_paid*, *price*, *num_subscribers*, *num_reviews*, *num_lectures*, *level*, *content_duration*, *published_timestamp*, dan *subject*. Meskipun sistem melakukan pengayaan data untuk mendukung pencocokan rekomendasi, cakupan data kursus masih bergantung pada kelengkapan dan karakteristik dataset tersebut. Oleh karena itu, hasil rekomendasi belum dapat dianggap mewakili seluruh pilihan kredensial mikro yang tersedia pada berbagai platform pembelajaran.

Ketiga, evaluasi pengguna masih dilakukan dalam skala terbatas. Uji coba pengguna hanya melibatkan sejumlah partisipan yang terbatas, sehingga hasil penilaian relevansi, penerimaan, dan pengalaman pengguna belum dapat digeneralisasi secara luas. Evaluasi ini lebih tepat dipahami sebagai evaluasi awal untuk melihat kelayakan sistem, bukan sebagai pembuktian menyeluruh terhadap efektivitas sistem pada populasi yang besar.

Keempat, evaluasi kualitas rekomendasi dan penilaian esai masih dilakukan melalui audit manual pada sampel terbatas. Meskipun audit tersebut dapat memberikan gambaran awal mengenai kesesuaian keluaran sistem, jumlah sampel yang terbatas membuat hasil evaluasi masih perlu diuji kembali dengan data yang lebih besar dan lebih beragam. Selain itu, penilaian

esai berbasis GPT tetap memiliki risiko ketidakkonsistenan keluaran, sehingga validasi dan pemantauan kualitas masih diperlukan.

Kelima, penelitian ini belum mengukur dampak pembelajaran jangka panjang. Sistem telah menghasilkan asesmen dan rekomendasi pembelajaran, tetapi penelitian belum menilai apakah rekomendasi tersebut benar-benar meningkatkan pemahaman, penyelesaian kursus, atau capaian kompetensi pengguna setelah mengikuti kursus yang direkomendasikan. Oleh karena itu, hasil penelitian ini lebih menekankan kelayakan awal sistem, bukan dampak pembelajaran secara longitudinal.

5.3 Saran Pengembangan Lanjutan

Berdasarkan keterbatasan tersebut, terdapat beberapa saran pengembangan lanjutan yang dapat dilakukan. Pertama, sistem dapat dikembangkan dari tahap MVP menuju sistem yang lebih lengkap dengan menambahkan fitur manajemen kredensial mikro, integrasi dengan sistem akademik, pemantauan progres belajar, dan dokumentasi capaian pembelajaran. Pengembangan ini penting agar sistem tidak hanya berhenti pada rekomendasi kursus, tetapi juga dapat mendukung siklus pembelajaran yang lebih utuh.

Kedua, basis pengetahuan rekomendasi dapat diperluas dengan menambahkan sumber data dari berbagai platform kursus atau penyedia kredensial mikro lain. Dataset juga dapat diperkaya dengan informasi yang lebih detail, seperti deskripsi kursus, capaian pembelajaran, prasyarat, bahasa, tingkat kesulitan yang lebih terstruktur, dan keterkaitan dengan kompetensi tertentu. Dengan basis pengetahuan yang lebih kaya, rekomendasi yang dihasilkan berpotensi menjadi lebih relevan dan lebih beragam.

Ketiga, mekanisme personalisasi dapat ditingkatkan dengan mempertimbangkan riwayat belajar, interaksi pengguna, preferensi yang berubah dari waktu ke waktu, serta umpan balik setelah pengguna menerima rekomendasi. Pada tahap lanjutan, sistem juga dapat membandingkan pendekatan RAG berbasis *feature hashing* lokal dengan pendekatan lain, seperti *embedding model* eksternal atau model rekomendasi hibrida yang lebih kompleks, selama tetap memperhatikan aspek keterlacakan dan biaya komputasi.

Keempat, evaluasi sistem perlu dilakukan dengan jumlah partisipan yang lebih besar dan karakteristik pengguna yang lebih beragam. Evaluasi juga dapat mencakup perbandingan dengan metode rekomendasi lain, pengujian pada beberapa domain kompetensi, serta pengukuran kualitas rekomendasi menggunakan metrik yang lebih luas. Dengan evaluasi yang

lebih besar, hasil penelitian dapat memberikan gambaran yang lebih kuat mengenai keandalan dan manfaat sistem.

Kelima, penelitian lanjutan dapat mengukur dampak rekomendasi terhadap hasil pembelajaran pengguna. Pengukuran tersebut dapat dilakukan dengan melihat perubahan skor asesmen, penyelesaian kursus, peningkatan kompetensi, atau tingkat kepuasan pengguna setelah mengikuti rekomendasi yang diberikan. Dengan demikian, pengembangan berikutnya tidak hanya menilai apakah sistem dapat menghasilkan rekomendasi, tetapi juga apakah rekomendasi tersebut benar-benar membantu proses belajar pengguna.

DAFTAR PUSTAKA

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>
- Bernardo, V. (2023). *TechDispatch: Explainable artificial intelligence*. Publications Office. <https://data.europa.eu/doi/10.2804/132319>
- Halim, F. (2024). Unravelling Challenges of Higher Education Institutions in Implementing Effective Micro-Credentials: A Multi-Stakeholder Qualitative Study. *Asian Journal of University Education*, 20(1), 114–126. <https://doi.org/10.24191/ajue.v20i1.25698>
- Harizan, D. S. H. (2025). *Artificial Intelligence In Micro-Credentials For Open And Distance Learning: A Technologically Enhanced Systematic Review*.
- Holt, D., Sun, X., & Davies, B. (2024). Assessment Feedback: What Do Students Want and Need? *Journal of University Teaching and Learning Practice*, 21(09). <https://doi.org/10.53761/tv2dfa83>
- Ji, Z., Chen, D., Ishii, E., Cahyawijaya, S., Bang, Y., Wilie, B., & Fung, P. (2024). LLM Internal States Reveal Hallucination Risk Faced With a Query. *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, 88–104. <https://doi.org/10.18653/v1/2024.blackboxnlp-1.6>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>

- Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning*, 16(3). <https://doi.org/10.19173/irrodl.v16i3.2112>
- Kato. (2023). *Micro-credentials for lifelong learning and employability: Uses and possibilities* (OECD Education Policy Perspectives 66; OECD Education Policy Perspectives, Vol. 66). <https://doi.org/10.1787/9c4b7b68-en>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* (Version 4). arXiv. <https://doi.org/10.48550/ARXIV.2005.111401>
- Meyliana, A., Rapiyanta, P. T., & Andriani, A. (2024). Application of the Rapid Application Development (RAD) Method for Web-Based Financial Management and Wood Inventory Using CodeIgniter. *ARRUS Journal of Engineering and Technology*, 4(1), 81–89. <https://doi.org/10.35877/jetech2722>
- Ricci, F., Rokach, L., & Shapira, B. (Ed.). (2015). *Recommender Systems Handbook*. Springer US. <https://doi.org/10.1007/978-1-4899-7637-6>
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 253–260. <https://doi.org/10.1145/564376.564421>
- Shanmughan, P., Joseph, J., S., B. N., S. M., A., C. S., A., & Jose, J. (2024a). The AI revolution in micro-credentialing: Personalized learning paths. *Frontiers in Education*, 9, 1445654. <https://doi.org/10.3389/feduc.2024.1445654>

- Shanmughan, P., Joseph, J., S., B. N., S. M., A., C. S., A., & Jose, J. (2024b). The AI revolution in micro-credentialing: Personalized learning paths. *Frontiers in Education*, 9, 1445654. <https://doi.org/10.3389/feduc.2024.1445654>
- Stein, M. (2025). *Micro-Credentials-Impact-Report-25.pdf*. Coursera. <https://www.luminafoundation.org/wp-content/uploads/2025/05/Micro-Credentials-Impact-Report-25.pdf>
- Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-Aware Recommender Systems for Learning: A Survey and Future Challenges. *IEEE Transactions on Learning Technologies*, 5(4), 318–335. <https://doi.org/10.1109/TLT.2012.11>
- Weinberger, K., Dasgupta, A., Langford, J., Smola, A., & Attenberg, J. (2009). Feature hashing for large scale multitask learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1113–1120. <https://doi.org/10.1145/1553374.1553516>
- Xiao, C., Ma, W., Song, Q., Xu, S. X., Zhang, K., Wang, Y., & Fu, Q. (2025). Human-AI Collaborative Essay Scoring: A Dual-Process Framework with LLMs. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 293–305. <https://doi.org/10.1145/3706468.3706507>