

MODEL AKUNTABILITAS KINERJA INSTANSI PEMERINTAH MENGUNAKAN *RETRIEVAL AUGMENTED GENERATION* (STUDI KASUS : BPPSDMKP)

RIDONA VIJU RAFELIANDI



**PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Model Akuntabilitas Kinerja Instansi Pemerintah Menggunakan *Retrieval Augmented Generation* (Studi Kasus : BPPSDMKP)” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Oktober 2026

Ridona Viju Rafeliandi
M0503241051

RINGKASAN

RIDONA VIJU RAFELIANDI. Model Akuntabilitas Kinerja Instansi Pemerintah Menggunakan *Retrieval Augmented Generation* (Studi Kasus : BPPSDMKP). Dibimbing oleh IRMAN HERMADI dan YANI NURHADRYANI.

Implementasi Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) yang menghasilkan volume dokumen regulasi dan laporan yang masif menyebabkan terlalu banyak informasi, sehingga menghambat efektivitas evaluasi kinerja organisasi dan pengambilan keputusan strategis. Meskipun penggunaan kecerdasan buatan bisa membantu mengatasi masalah tidak efisien tersebut, penggunaannya tetap berisiko menghasilkan halusinasi data yang bisa mengganggu keakuratan dalam menganalisis kebijakan. Oleh karena itu, arsitektur *Retrieval-Augmented Generation* (RAG) diajukan sebagai solusi inovatif yang menggabungkan kemampuan mengambil informasi dari sumber pengetahuan yang terverifikasi dengan *Large Language Model* (LLM), sehingga memastikan hasil evaluasi SAKIP menjadi lebih valid, akurat, dan dapat berinteraksi dengan baik.

Penelitian ini bertujuan untuk: (1) merancang model akuntabilitas kinerja instansi pemerintah menggunakan RAG; (2) menganalisis dan mengukur performa model RAG; serta (3) mengevaluasi manfaat dan memberikan rekomendasi penerapan model RAG. Penelitian ini menggunakan pendekatan RAG, di mana kualitas model dinilai menggunakan *framework* RAGAS, sedangkan manfaatnya dinilai melalui metode *System Usability Scale* (SUS) dan uji statistik *Wilcoxon Signed-Rank Test*.

Penelitian ini menghasilkan model yang menggabungkan sistem evaluasi SAKIP dengan RAG. Hasil penilaian menggunakan metrik RAGAS kemampuan pengambilan informasi secara komprehensif yang dibuktikan melalui skor pada aspek *Context Precision* (0,835) dan *Context Recall* (0,984). Selain itu, respons yang dihasilkan terbukti relevan dan konsisten secara fakta dan makna sesuai dengan jawaban sebenarnya (*ground truth*), yang ditunjukkan oleh skor pada aspek *Faithfulness* (0,979), *Answer Relevancy* (0,816), dan *Answer Correctness* (0,964). Sementara itu, hasil uji statistik menunjukkan adanya perbedaan yang sangat signifikan pada efektivitas kerja sebelum dan sesudah intervensi penggunaan model RAG ($Z = -3,294$; $p < 0,05$). Evaluasi *usability* menghasilkan skor rata-rata 73,21 (kategori *Acceptable*).

Penerapan kecerdasan buatan ini mengubah proses evaluasi dari yang bersifat manual menjadi sistem pendukung keputusan yang mampu mengurangi beban tugas rutin asesor, sehingga asesor dapat lebih berfokus pada analisis konteks strategis yang berorientasi pada hasil. Temuan penelitian ini diharapkan dapat menjadi rujukan dalam mempercepat digitalisasi sistem monitoring dan evaluasi kinerja instansi pemerintah agar lebih cerdas dan adaptif. Meskipun demikian, kecerdasan buatan dalam sistem ini berfungsi sebatas sebagai alat bantu, sedangkan keputusan akhir tetap sepenuhnya berada pada pengguna.

Kata Kunci : SAKIP, Kecerdasan Buatan, *Retrieval Augmented Generation*

SUMMARY

RIDONA VIJU RAFELIANDI. Model of Government Agency Performance Accountability using Retrieval-Augmented Generation (Case Study: BPPSDMKP). Supervised by IRMAN HERMADI and YANI NURHADRYANI.

The implementation of the Government Agency Performance Accountability System (SAKIP), which produces a massive volume of regulatory documents and reports, results in an overload of information, hampering the effectiveness of organizational performance evaluation and strategic decision-making. While the use of artificial intelligence can help address this inefficiency, its use still carries the risk of data hallucinations that can compromise the accuracy of policy analysis. Therefore, the Retrieval-Augmented Generation (RAG) architecture is proposed as an innovative solution that combines the ability to retrieve information from verified knowledge sources with a Large Language Model (LLM), thereby ensuring more valid, accurate, and interactive SAKIP evaluation results.

This study aims to: (1) design a government agency performance accountability model using RAG; (2) analyze and measure the performance of the RAG model; and (3) evaluate the benefits and provide recommendations for implementing the RAG model. This study uses the RAG approach, where model quality is assessed using the RAGAS framework, while its benefits are assessed using the System Usability Scale (SUS) method and the Wilcoxon Signed-Rank Test.

The research resulted in a model that integrates the SAKIP evaluation system with RAG. Assessments using RAGAS metrics demonstrated comprehensive information retrieval capabilities, evidenced by scores in Context Precision (0.835) and Context Recall (0.984). Furthermore, the generated responses proved relevant and consistent in terms of both fact and meaning with the actual answers (ground truth), as indicated by scores in Faithfulness (0.979), Answer Relevancy (0.816), and Answer Correctness (0.964). Meanwhile, statistical analysis revealed a highly significant difference in work effectiveness before and after the implementation of the RAG model ($Z = -3.294$; $p < 0.05$). Usability evaluation yielded an average score of 73.21 (categorized as "Acceptable").

The implementation of artificial intelligence transforms the evaluation process from a manual approach into a decision support system capable of reducing the routine workload of assessors, thereby allowing them to focus more on analyzing the strategic context with an outcome-oriented perspective. The findings of this study are expected to serve as a reference for accelerating the digitalization of government agency performance monitoring and evaluation systems, making them smarter and more adaptive. Nevertheless, the artificial intelligence within this system functions solely as a tool, while the final decision remains entirely with the user.

Keywords : SAKIP, Artificial Intelligence, Retrieval Augmented Generation



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

MODEL AKUNTABILITAS KINERJA INSTANSI PEMERINTAH MENGUNAKAN *RETRIEVAL AUGMENTED GENERATION* (STUDI KASUS : BPPSDMKP)

RIDONA VIJU RAFELIANDI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Komputer pada
Program Studi Ilmu Komputer

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:
Rina Trisminingsih S.Komp., M.T., Ph.D.

Judul Tesis : Model Akuntabilitas Kinerja Instansi Pemerintah Menggunakan
Retrieval Augmented Generation (Studi Kasus : BPPSDMKP)
Nama : Ridona Viju Rafeliandi
NIM : M0503241051

Disetujui oleh

Pembimbing 1:
Irman Hermadi, S.Kom., M.S., Ph.D



Pembimbing 2:
Dr. Yani Nurhadryani, S.Si, M.T



Diketahui oleh

Ketua Program Studi:
Dr. Karlisa Priandana, S.T, M.Eng
NIP. 19851121 201212 2 002



Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom
NIP. 19660702 199302 1 001



Tanggal Ujian:
10 Agustus 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Penelitian dilaksanakan sejak bulan Juli 2025 sampai bulan Juli 2026, dengan judul penelitian “Model Akuntabilitas Kinerja Instansi Pemerintah Menggunakan *Retrieval Augmented Generation* (Studi Kasus : BPPSDMKP)”.

Terima kasih penulis ucapkan kepada para pembimbing yaitu Bapak Irman Hermadi, S.Kom., MS, Ph.D. selaku ketua pembimbing dan Ibu Dr. Yani Nurhadryani, S.Si., MT selaku anggota pembimbing yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada komisi pembimbing, moderator ujian tesis Ibu Dr. Shelve Nidya Neyman, S.Kom., M.T., dan penguji luar komisi pembimbing Ibu Rina Trisminingsih S.Kom., M.T., Ph.D. Di samping itu, penghargaan penulis sampaikan kepada rekan-rekan Tim Kerja Pemantauan, Evaluasi dan Pelaporan, Sekretariat BPPSDMKP beserta staf Laboratorium Server Komputasi IPB yang telah membantu selama pengumpulan dan pengolahan data. Ungkapan terima kasih juga disampaikan kepada Ibu Epih Hendraningsih, Bapak Ujun Junaedi, serta istri Fahriyah Hastuti, anak Hideo Rakauna Hafiedz, Shana Rainun Hafiedza yang telah memberikan dukungan, doa, dan kasih sayangnya.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Oktober 2026

Ridona Viju Rafeliandi

DAFTAR ISI

DAFTAR ISI	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN	vii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II TINJAUAN PUSTAKA	5
2.1 Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan	5
2.2 Kinerja	6
2.3 Sistem Akuntabilitas Kinerja Instansi Pemerintah	8
2.4 Penilaian SAKIP	9
2.5 Aplikasi e-SAKIP Reviu	10
2.6 <i>Dashboard</i>	11
2.7 <i>Large Language Model</i>	13
2.8 <i>Retrieval Augmented Generation</i>	15
2.9 <i>Retrieval Augmented Generation Assessment</i>	16
2.10 Model Gemini 2.5 Pro	18
2.11 <i>Facebook AI Similarity Search (FAISS)</i>	18
2.12 Evaluasi Model	19
2.13 <i>System Usability Scale (SUS)</i>	20
2.14 <i>Wilcoxon Signed-Rank Test</i>	21
2.15 Penelitian Sebelumnya	22
III METODE	29
3.1 Lokasi Penelitian	29
3.2 Kebutuhan Pengembangan	29
3.3 Tahapan Penelitian	30
IV HASIL DAN PEMBAHASAN	35
4.1 Analisis Sistem Evaluasi SAKIP	35
4.2 <i>Use Case Diagram</i>	39
4.3 <i>Activity Diagram</i>	41
4.4 <i>Sequence Diagram</i>	42
4.5 Arsitektur RAG	43
4.6 Hasil Praproses Data	44
4.7 Hasil Pembangunan RAG	47
4.8 Hasil Pembangunan <i>Dashboard</i> Evaluasi SAKIP	50
4.9 Evaluasi Model RAG	52
4.10 Evaluasi Manfaat	57

4.11 Evaluasi <i>Dashboard</i>	59
V SIMPULAN DAN SARAN	62
5.1 Simpulan	62
5.2 Saran	62
DAFTAR PUSTAKA	64
LAMPIRAN	68
RIWAYAT HIDUP	103



DAFTAR TABEL

1	Pembobotan Nilai Evaluasi AKIP	10
2	<i>Evaluation Matrix</i>	19
3	Penelitian Terkait SAKIP	22
4	Penelitian Terkait RAG dan <i>Dashboard</i>	27
5	Dokumen Kinerja	45
6	Hasil <i>Chunking</i>	45
7	Waktu Respon <i>Knowledge Ingestion</i>	50
8	Waktu Respon Interaksi AI	50
9	Dataset Evaluasi	54
10	Hasil Survey <i>Pretest</i> dan <i>Posttest</i>	60
11	Hasil Uji <i>Wilcoxon Signed-Rank</i>	60

DAFTAR GAMBAR

1	Bagan Susunan Organisasi Jabatan Struktural Eselon I dan Jabatan Struktural Eselon II Kementerian Kelautan dan Perikanan	5
2	Sebaran Unit Kerja Lingkup BPPSDMKP	6
3	<i>Leveling of Objective</i>	7
4	Contoh representatif proses RAG yang diterapkan pada tanya jawab	15
5	<i>Confussion Matrix</i>	19
6	Contoh Pengukuran <i>System Usability</i>	21
7	<i>Adjective Rating SUS</i>	21
8	Tahapan Kegiatan Penelitian	34
9	Siklus Pengelolaan Kinerja Organisasi	35
10	<i>Timeline</i> Pengelolaan Kinerja Organisasi	35
11	<i>Use Case Diagram</i>	40
12	Tampilan Menu Buat Baru	46
13	Tampilan Menu Tambah Data	46
14	Tampilan Menu Hapus Data	46
15	Tampilan Menu <i>Preview</i>	47
16	Contoh <i>Prompt Template</i> untuk <i>Query Transformation</i>	47
17	Contoh Hasil <i>Retrieval</i>	48
18	Contoh <i>Prompt Template</i> untuk <i>Reranking</i> Dokumen	48
19	Contoh <i>Prompt Template</i> Akhir	49
20	Contoh Tampilan Pengisian <i>Ground Truth</i>	49
21	Tampilan Grafik Perbandingan Hasil Penilaian Mandiri SAKIP	50
22	Contoh Tampilan Perbandingan Nilai SAKIP Satker Berdasarkan Komponen dan Sub-Komponen SAKIP.	51
23	Contoh Tampilan Perbandingan Nilai Satker (Tertinggi - Terendah)	51
24	Contoh Tampilan Rekomendasi Hasil Penilaian Mandiri SAKIP	51
25	Contoh Tampilan Laporan Hasil Penilaian Mandiri SAKIP	52
26	Hasil perhitungan model RAG (sebelum dan sesudah perbaikan)	53
27	Hasil perhitungan model RAG berdasarkan komponen SAKIP	55
28	Contoh Percakapan (Skor <i>Faithfulness</i>)	56
29	Contoh Percakapan (Skor <i>Context Precision</i>)	56

30	Contoh Percakapan (Skor <i>Context Recall</i>)	56
31	Contoh Percakapan (Skor <i>Answer Correctness</i>)	57
32	Contoh Percakapan (Skor <i>Answer Relevancy</i>)	57

DAFTAR LAMPIRAN

1	Peta Kriteria Evaluasi AKIP	69
2	Gradasi Nilai Jawaban LKE AKIP	73
3	Predikat Nilai AKIP	74
4	<i>Curated Dashboard</i>	75
5	<i>Data Collection Dashboards</i>	76
6	Contoh Tampilan <i>Data Information, Meta Information, Visual Encoding, Interaction, Layout</i> , dan Penggunaan Warna pada <i>Dashboard</i>	77
7	<i>Activity Diagram</i>	80
8	<i>Sequence Diagram</i>	81
9	Arsitektur RAG	82
10	<i>Pseudocode</i> untuk <i>Preprocessing</i>	83
11	<i>Pseudocode</i> untuk <i>Knowledge Ingestion</i>	84
12	<i>Pseudocode</i> untuk <i>Retrieval, Reranking, dan Generation</i>	86
13	<i>Pseudocode</i> untuk <i>Evaluation</i>	88
14	Kuesioner Evaluasi <i>System Usability Scale (SUS)</i>	90
15	Kuesioner Evaluasi Manfaat Aplikasi e-SAKIP Reviu (eSR Kemen PANRB)	95
16	Kuesioner Evaluasi Manfaat Aplikasi DSMS SAKIP BPPSDMKP	98
17	Hasil Evaluasi Manfaat Aplikasi e-SAKIP Reviu (eSR Kemen PANRB)	101
18	Hasil Evaluasi Manfaat Aplikasi DSMS SAKIP BPPSDMKP	102



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

I PENDAHULUAN

1.1 Latar Belakang

Akuntabilitas kinerja adalah perwujudan kewajiban suatu instansi pemerintah untuk mempertanggungjawabkan keberhasilan dan kegagalan pelaksanaan misi organisasi dalam mencapai sasaran dan tujuan yang telah ditetapkan melalui sistem pertanggungjawaban secara periodik. Sedangkan, Akuntabilitas Kinerja Instansi Pemerintah (AKIP) adalah pertanggungjawaban dan peningkatan kinerja instansi pemerintah melalui implementasi Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP). Penguatan akuntabilitas kinerja merupakan salah satu strategi yang dilaksanakan dalam rangka mempercepat pelaksanaan Reformasi Birokrasi, untuk mewujudkan pemerintahan yang bersih dan akuntabel, pemerintahan yang kapabel, serta meningkatnya kualitas pelayanan publik kepada masyarakat.

Dalam Instruksi Presiden Nomor 7 Tahun 1999 tentang Akuntabilitas Kinerja Instansi Pemerintah (AKIP), pada setiap tahun anggaran, mulai tahun anggaran 2000/2001, setiap instansi pusat dan daerah wajib menyampaikan laporan akuntabilitas kinerja instansi pemerintah kepada Presiden dan salinannya kepada Kepala Badan Pengawasan Keuangan dan Pembangunan dengan menggunakan pedoman penyusunan sistem akuntabilitas kinerja. Peraturan Presiden Nomor 29 Tahun 2014 tentang SAKIP mencabut Instruksi Presiden Nomor 7 Tahun 1999 tentang Akuntabilitas Kinerja Instansi Pemerintah. Hal ini semakin memperkuat implementasi SAKIP secara lebih komprehensif di Indonesia.

Dalam menindaklanjuti perintah Presiden tentang peningkatan efektivitas dan efisiensi penggunaan anggaran, penyederhanaan penyampaian laporan kinerja dengan menggunakan *e-government* serta meningkatkan kualitas pelaksanaan akuntabilitas kinerja instansi pemerintah, Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (Kemen PANRB) telah membangun sistem aplikasi Reviu AKIP dan penyampaian Laporan Kinerja Instansi Pemerintah yaitu e-SAKIP Reviu. Melalui sistem ini pembinaan/pendampingan peningkatan kualitas pelaksanaan akuntabilitas kinerja dan penyampaian laporan kinerja dapat dilakukan secara *online*. Hal ini seperti tertuang dalam Surat Edaran Kemen PANRB Nomor B/05/M.RB.00/2018 tanggal 19 Januari 2018.

Untuk mengetahui sejauh mana implementasi SAKIP dilaksanakan, serta untuk mendorong peningkatan pencapaian kinerja yang tepat sasaran dan berorientasi hasil, maka perlu dilakukan evaluasi AKIP atau evaluasi atas implementasi SAKIP. Kemen PANRB telah menerbitkan Peraturan Menteri (Permen) PANRB Nomor 88 Tahun 2021 tentang Evaluasi Akuntabilitas Kinerja Instansi Pemerintah, yang merupakan penyesuaian terhadap kompleksitas pengukuran kinerja instansi pemerintah dan pesatnya kemajuan teknologi, yang sebelumnya diatur dalam Permen PANRB Nomor 12 Tahun 2015 tentang Pedoman Evaluasi atas Implementasi SAKIP.

Selama tahun 2024-2025 terjadi peningkatan nilai SAKIP Kementerian Kelautan dan Perikanan (KKP) dari 84,01 menjadi 84,45, begitu juga dengan Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan (BPPSDMKP) yang mengalami peningkatan nilai dari 88,79 menjadi 89,12. Implementasi SAKIP berhasil mencegah potensi pemborosan APBN/APBD sebesar ± Rp128,5 Triliun selama tahun 2023-2024. Hal ini disampaikan oleh Plt. Asisten Deputi Koordinasi Pelaksanaan Kebijakan dan Evaluasi Reformasi

Birokrasi, Akuntabilitas Aparatur, dan Pengawasan II, Kemen PANRB pada acara *Kick Off Meeting* Evaluasi RB, SAKIP dan Zona Integritas Tahun 2025 pada tanggal 1 Agustus 2025 di Jakarta. Namun demikian, sebagian besar pemerintah kabupaten/kota masih memperoleh nilai di bawah provinsi dan kementerian/lembaga dalam evaluasi SAKIP, yang mengindikasikan kualitas analisis dan pemanfaatan data kinerja untuk perbaikan berkelanjutan masih belum optimal. Capaian akuntabilitas kinerja pemerintah kabupaten di berbagai daerah masih berada pada kategori cukup, dengan skor rata-rata sekitar 61,6 (Hendrik *et al.* 2022). Salah satu penyebabnya yaitu keterbatasan kapasitas sumber daya manusia pengelola SAKIP (Salomo dan Rahmayanti 2023).

Kondisi tersebut memunculkan fenomena kelebihan beban informasi di kalangan pengelola kinerja, di mana volume regulasi, capaian indikator kinerja utama, dan dokumen pendukung yang harus dibaca, disintesis, dan dianalisis jauh melampaui kapasitas kognitif manusia dalam rentang waktu penyusunan laporan yang terbatas. Menurut Motzfeldt Jensen *et al.* (2025), ekstraksi manual memerlukan banyak tenaga, memakan waktu, dan mudah membuat kesalahan. Hal ini berdampak pada kemampuan untuk mengulangi hasil dan menyulitkan dalam menangani kumpulan data yang besar.

Untuk mengatasi keterbatasan telaah secara manual, diperlukan perbaikan dengan memanfaatkan AI. Dalam Jana *et al.* (2024), *Large Language Model* (LLM) membantu sistem pencarian informasi dan tanya jawab yang memudahkan pengguna untuk mencari informasi yang relevan serta menjawab pertanyaan yang diajukan dengan bahasa yang alami. LLM memiliki kemampuan yang luar biasa namun masih memiliki keterbatasan, salah satunya yaitu masalah halusinasi (memberikan respon seolah serius tetapi sebenarnya tidak relevan dengan *query* yang diminta oleh pengguna). Dalam konteks inilah, *Retrieval-Augmented Generation* (RAG) hadir sebagai arsitektur yang menggabungkan kemampuan pengambilan informasi (*retrieval*) dari basis pengetahuan (*knowledge base*) yang terverifikasi dengan kemampuan generasi bahasa alami dari LLM (Lewis *et al.* 2021). RAG meningkatkan LLM dengan mengambil potongan dokumen yang relevan dari basis pengetahuan eksternal melalui perhitungan kesamaan semantik (Gao *et al.* 2024).

Dalam konteks SAKIP, implementasi RAG masih sangat terbatas, sebagian besar studi tentang pemanfaatan teknologi informasi pada SAKIP melalui pendekatan sistem informasi berbasis *website* dan belum dilengkapi dengan aspek kecerdasan buatan (Yassir *et al.* 2024; Marjan *et al.* 2021; Sadewa *et al.* 2020; Saputra *et al.* 2017; Pradana dan Maghfiroh 2024). Hal ini selaras dalam mendukung terwujudnya pelaksanaan Reformasi Birokrasi dengan menindaklanjuti salah satu rekomendasi perbaikan implementasi SAKIP yaitu pemanfaatan *Artificial Intelligence* (AI) untuk memperkuat perumusan kebijakan berbasis data.

Penggunaan RAG ini sangat penting, karena mempertimbangkan jumlah pertanyaan dalam penilaian SAKIP yang tidak sedikit (Permen PANRB Nomor 88 Tahun 2021), begitu juga dengan jumlah satuan kerja yang dinilai. Untuk menjawab setiap pertanyaan tersebut harus dilengkapi dengan data dan dokumen kinerja yang relevan, sehingga diperlukan sebuah proses penelaahan dokumen kinerja yang lebih cepat, akurat dan efisien untuk memberikan *insight* yang mendalam dalam melakukan penilaian tersebut.

1.2 Rumusan Masalah

Saat ini BPPSDMKP memiliki jumlah satuan kerja (satker) sebanyak 43 satker di seluruh Indonesia dengan target IKU PM SAKIP yang beragam. Dengan rentang kendali yang luas sehingga diperlukan metode yang lebih efisien dan efektif dalam menganalisis data dan dokumen kinerja yang dibutuhkan dalam penilaian SAKIP. Berdasarkan permasalahan tersebut maka pertanyaan penelitian yang pertama adalah “Bagaimana mengintegrasikan data dan dokumen kinerja instansi pemerintah dari berbagai sumber secara otomatis dan akurat menggunakan *Retrieval Augmented Generation* (RAG) ?”.

Tahapan selanjutnya yaitu melakukan analisis terhadap performa model yang dibuat apakah telah dapat memberikan penilaian yang lebih cepat, akurat, dan kontekstual dibandingkan metode konvensional. Berdasarkan kondisi tersebut maka pertanyaan penelitian yang ke-2 adalah “Bagaimana tingkat performa model RAG yang dikembangkan dalam hal akurasi, relevansi jawaban, dan efisiensi proses pencarian informasi berdasarkan data kinerja instansi pemerintah ?”.

Penggunaan RAG pada model akuntabilitas kinerja instansi pemerintah diharapkan dapat mempercepat proses analisis terhadap dokumen atau laporan kinerja, memperkuat narasi analitik dalam bentuk *dashboard* dan memperkuat saran atau rekomendasi perbaikan kinerja yang diberikan oleh evaluator. Berdasarkan kondisi tersebut maka pertanyaan penelitian yang ke-3 adalah “Bagaimana manfaat penerapan model RAG terhadap peningkatan efektivitas analisis kinerja instansi pemerintah serta rekomendasi strategis yang dapat diberikan ?”.

1.3 Tujuan

Berdasarkan latar belakang dan rumusan masalah, maka tujuan penelitian ini adalah:

1. Merancang model kinerja instansi pemerintah menggunakan *Retrieval Augmented Generation* (RAG).
2. Menganalisis dan mengukur performa model RAG.
3. Mengevaluasi manfaat dan memberikan rekomendasi penerapan model RAG.

1.4 Manfaat

Manfaat dilakukannya penelitian ini adalah sebagai berikut:

1. Bagi Pemerintah (BPPSDMKP): Penelitian ini diharapkan dapat meningkatkan efisiensi dan akurasi analisis dokumen kinerja instansi pemerintah menggunakan RAG secara otomatis.
2. Bagi Disiplin Ilmu: Penelitian ini diharapkan sebagai salah satu acuan akademik studi kasus empiris penerapan RAG di bidang pemerintahan.
3. Manfaat Teoritis: Penelitian ini diharapkan dapat memberikan kontribusi teoritis berupa implementasi RAG yang menghubungkan antara *AI reasoning* dan *governance analytics*.

1.5 Ruang Lingkup

Ruang lingkup dari penelitian yang dilakukan meliputi:

1. Mekanisme penilaian SAKIP pada penelitian ini mengacu pada Permen PANRB Nomor 88 Tahun 2021 dan Permen KP Nomor 35 Tahun 2023.

2. Data yang digunakan pada penelitian ini diperoleh dari hasil pendampingan dan verifikasi penilaian mandiri SAKIP oleh tim evaluator Tahun 2022-2025 (mengacu pada pemberlakuan aturan Evaluasi AKIP terbaru).
3. Pengguna dari model ini adalah tim pengelola kinerja dari seluruh satker lingkup BPPSDMKP dan Inspektorat Jenderal sebagai Tim Evaluator Penilaian Mandiri SAKIP.
4. Model ini merupakan modul terpisah dengan aplikasi kinerja lainnya.

@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang

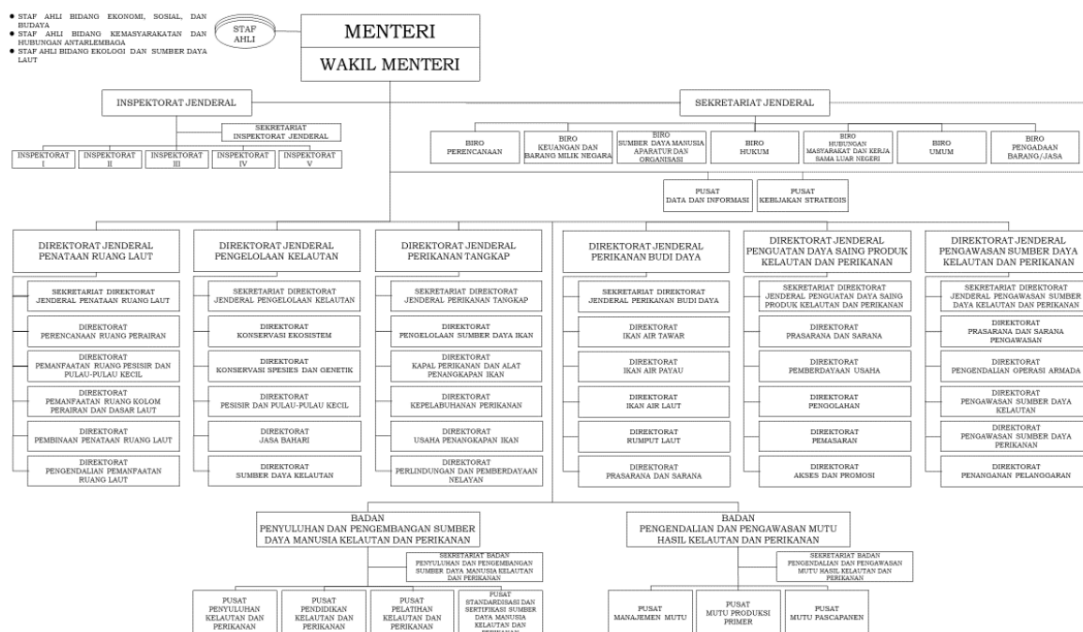
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

II TINJAUAN PUSTAKA

2.1 Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan

Dalam Peraturan Menteri Kelautan dan Perikanan (Permen KP) Nomor 2 Tahun 2025 tentang Organisasi dan Tata Kerja Kementerian Kelautan dan Perikanan (KKP), menyebutkan tugas Kementerian Kelautan dan Perikanan yaitu menyelenggarakan urusan pemerintahan di bidang kelautan dan perikanan untuk membantu Presiden dalam menyelenggarakan pemerintahan negara. Struktur Organisasi KKP dapat dilihat pada Gambar 1, yang terdiri dari:

1. Sekretariat Jenderal
2. Direktorat Jenderal Penataan Ruang Laut
3. Direktorat Jenderal Pengelolaan Kelautan
4. Direktorat Jenderal Perikanan Tangkap
5. Direktorat Jenderal Perikanan Budi Daya
6. Direktorat Jenderal Penguatan Daya Saing Produk Kelautan dan Perikanan
7. Direktorat Jenderal Pengawasan Sumber Daya Kelautan dan Perikanan
8. Inspektorat Jenderal
9. Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan
10. Badan Pengendalian dan Pengawasan Mutu Hasil Kelautan dan Perikanan
11. Staf Ahli Bidang Ekonomi, Sosial, dan Budaya
12. Staf Ahli Bidang Kemasyarakatan dan Hubungan Antarlembaga
13. Staf Ahli Bidang Ekologi dan Sumber Daya Laut



Gambar 1 Bagan Susunan Organisasi Jabatan Struktural Eselon I dan Jabatan Struktural Eselon II Kementerian Kelautan dan Perikanan

Salah satu Unit Eselon I di KKP mempunyai tugas menyelenggarakan penyuluhan dan pengembangan sumber daya manusia kelautan dan perikanan adalah Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan (BPPSDMKP) yang didukung oleh salah satu Unit Eselon II yaitu Sekretariat BPPSDMKP yang mempunyai tugas melaksanakan pelayanan teknis dan pelaksanaan administrasi kepada seluruh unit organisasi pada BPPSDMKP, dimana salah satu fungsinya yaitu fungsi koordinasi penyusunan rencana, program, dan anggaran, pengelolaan kinerja organisasi, serta pemantauan, evaluasi, dan pelaporan pelaksanaan rencana, program, dan anggaran pada BPPSDMKP. Untuk mendukung hal tersebut, maka Sekretariat BPPSDMKP melalui Tim Kerja Monitoring, Evaluasi dan Pelaporan mendukung penyelenggaraan fungsi dalam rangka mendukung tercapainya target Indikator Kinerja Utama (IKU) Penilaian Mandiri (PM) Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) pada 43 Satuan Kerja (Satker) Lingkup BPPSDMKP. Setiap tahunnya dilaksanakan PM SAKIP melalui proses pendampingan kepada seluruh satker dengan lokasi seperti terlihat pada Gambar 2.



Gambar 2 Sebaran Unit Kerja Lingkup BPPSDMKP

2.2 Kinerja

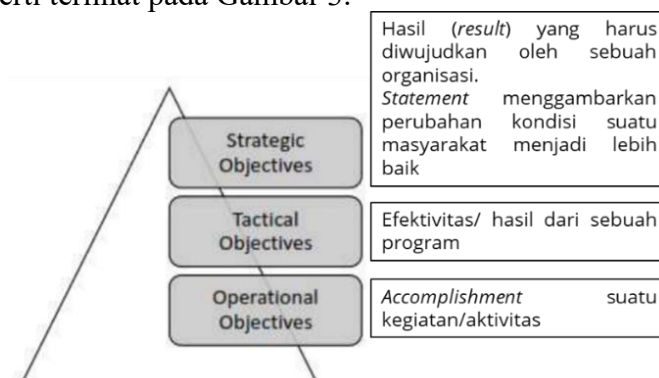
Menurut Peraturan Menteri PANRB Nomor 89 Tahun 2021, kinerja adalah keluaran/hasil dari kegiatan/program yang telah atau hendak dicapai sehubungan dengan penggunaan anggaran dengan kuantitas dan kualitas terukur. Ukuran keberhasilan yang akan dicapai dari kinerja program dan kegiatan yang telah direncanakan yaitu indikator kinerja. Penjenjangan kinerja adalah proses penjabaran dan penyelarasan sasaran strategis, indikator kinerja, dan target kinerja organisasi kepada unit organisasi sampai dengan individu pegawai. Penyelarasan tujuan organisasi dengan kinerja setiap individu dan dengan desain strategi/program/kegiatan, maka didasarkan pada proses membangun model logis (*logic model*) maupun kerangka berpikir logis (*logical framework*) yang tepat.

Pohon kinerja adalah alat bantu bagi organisasi untuk mengawal struktur logika sebab akibat atas berbagai kondisi yang diperlukan organisasi dalam menghasilkan *outcome* yang diinginkan. Menyusun pohon kinerja sebenarnya seperti mengurai rute atau jalur kinerja yang secara logis dianggap paling terkuat dan dibutuhkan dalam pencapaian *outcome* yang diinginkan. Prinsip-prinsip umum dalam membangun pohon kinerja:

1. Logis, yakni suatu pohon kinerja harus menggambarkan hubungan “sebab-akibat” ataupun “jika-maka”.
2. Empiris, yakni suatu pohon kinerja harus berdasarkan kondisi/isu strategis/permasalahan faktual yang terjadi.
3. Antisipatif, yakni suatu pohon kinerja harus disusun dengan mempertimbangkan kondisi masa depan karena akan digunakan dalam jangka waktu lebih dari satu tahun.
4. Dinamis, yakni suatu pohon kinerja harus mengikuti perubahan lingkungan strategis.
5. Holistik, yakni suatu pohon kinerja harus mempertimbangkan keterkaitan dengan urusan lainnya.
6. *Out of The Box*, yakni suatu pohon kinerja harus mengedepankan kerangka logis untuk mendapatkan upaya/strategis terbaik, bukan untuk mempertahankan *status quo*.
7. Materialitas, yakni suatu pohon kinerja harus diisi oleh kondisi-kondisi yang penting, strategis, dan paling berdampak.

Model logis terdiri dari tahapan kondisi yang saling berkaitan/berhubungan dalam membentuk *outcome*/hasil yang diharapkan. Tahapan kondisi ini membentuk suatu alur logis yang tersistem yang sering dinamai dengan rantai nilai (*value chain*), yang model sederhananya terdiri dari *input*, proses, *output*, *outcome*. *Input* merupakan besaran sumber daya yang dibutuhkan sebuah organisasi untuk memproduksi *output* (keluaran) baik barang maupun jasa. Proses merupakan aktivitas atau upaya yang dilakukan untuk mengolah input menjadi *output*. *Output* merupakan barang/jasa yang dihasilkan oleh sebuah organisasi dalam rangka mencapai *outcome*. Sedangkan *outcome* adalah hasil dari berfungsinya *output*. Dalam kenyataannya, tahapan/alur logis yang terdiri lebih kompleks, bervariasi dan lebih panjang. Sebuah model logis dari sebuah program seringkali membutuhkan tahapan kondisi yang lebih panjang. Setiap *output* pun, tidak selalu langsung menghasilkan *outcome* yang diinginkan. Terdapat kondisi antara output dan outcome yang ingin dicapai, kondisi tersebut sering disebut juga dengan *outcome* antara (*intermediate outcome*), dan *outcome* pendahuluan (*initial/intermediate outcome*).

Dalam sebuah organisasi, setidaknya dikenal tiga level kinerja (sasaran kinerja), yaitu sasaran/kinerja level strategis (*strategic objectives*), sasaran/kinerja level taktis (*tactical objectives*), dan sasaran/kinerja level operasional (*operational objectives*) seperti terlihat pada Gambar 3.



Gambar 3 Leveling of Objective

2.3 Sistem Akuntabilitas Kinerja Instansi Pemerintah

Menurut Peraturan Presiden Nomor 29 Tahun 2014, Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) adalah serangkaian proses sistematis dari beragam aktivitas, alat, dan metode yang dirancang untuk mengidentifikasi dan mengukur, mengumpulkan data, mengklasifikasikan, merangkum, dan melaporkan kinerja lembaga pemerintah, sebagai bagian dari pertanggungjawaban dan peningkatan kinerja lembaga pemerintah.

Menurut Raharja (2018) dalam Pratama (2024), indikator-indikator utama dalam penilaian AKIP menurut Perpres Nomor 29 Tahun 2014 tentang SAKIP meliputi:

1. Perencanaan kinerja

Perencanaan kinerja diterjemahkan ke dalam dokumen perjanjian kinerja yang sesuai dengan jabatan masing-masing. Perjanjian kinerja merupakan dokumen yang terkait dengan penugasan pimpinan dalam melaksanakan program atau kegiatan beserta indikator kinerjanya. Berbagai informasi yang terdapat dalam rencana kinerja diantaranya: 1) tujuan yang akan dicapai dalam periode tertentu; 2) program, kegiatan dan berbagai indikator serta target kinerja; 3) informasi yang menjelaskan hubungan antara kegiatan dan sasaran, kebijakan dan pogram, dan hubungan berbagai macam kegiatan yang dilakukan oleh organisasi.

2. Pengukuran kinerja

Pengukuran kinerja merupakan suatu proses sistematis dan berkelanjutan dalam mengevaluasi keberhasilan dan kegagalan implementasi kegiatan yang mengacu pada program, kebijakan dan sasaran serta tujuan yang diterjemahkan dalam perwujudan visi, misi dan strategi organisasi. Proses ini bertujuan dalam mengevaluasi capaian seluruh indikator kinerja untuk menyajikan gambaran keberhasilan maupun kegagalan dalam mencapai tujuan dan sasaran. Selain itu juga dilaksanakan analisis akuntabilitas kinerja yang menyajikan hubungan antara capaian kinerja kegiatan dengan program dan kebijakan untuk mencapai tujuan dan sasaran serta visi misi yang telah termuat dalam rencana strategis. Pada penerapan SAKIP, pengukuran kinerja dilaksanakan melalui perbandingan realisasi dengan target kinerja kinerja yang dituangkan dalam dokumen perjanjian kinerja pada konteks pelaksanaan anggaran tahun berjalan sehingga bisa terlihat hasil capaian kinerjanya seperti apa.

3. Pelaporan kinerja

Seluruh instansi pemerintah wajib untuk menyusun dan menyampaikan laporan kinerja sesuai dengan alokasi anggaran yang telah ditetapkan. Laporan kinerja merupakan ikhtisar pencapaian kinerja secara ringkas dan lengkap yang dibuat sesuai dengan rencana kerja yang telah ditetapkan dalam konteks pelaksanaan anggaran. Laporan kinerja adalah sarana akuntabilitas bagi lembaga pemerintah dalam pertanggungjawaban dan penjelasan atas keberhasilan maupun kegagalan pencapaian kinerja. Output laporan kinerja merupakan hasil analisis atas pengukuran kinerja yang telah dilaksanakan.

4. Evaluasi kinerja

Tahap terakhir adalah evaluasi kinerja sehingga informasi kinerja dapat digunakan untuk meningkatkan kinerja berkelanjutan. Evaluasi kinerja adalah kegiatan mengevaluasi keberhasilan dan kegagalan suatu instansi

pemerintah dalam melakukan tugas dan fungsi yang diamanatkan. Evaluasi kinerja memiliki tujuan untuk memastikan kemajuan yang dicapai dan hambatan yang dihadapi dalam rangka peningkatan implementasi program atau kegiatan di masa depan. Evaluasi kinerja dapat dilakukan dengan membandingkan rencana dengan capaian kinerja guna menyajikan informasi mengenai adanya kesenjangan kinerja.

2.4 Penilaian SAKIP

Permen KP Nomor 35 Tahun 2023 tentang Pengelolaan Kinerja Organisasi di Lingkungan Kementerian Kelautan dan Perikanan, menyebutkan salah satu bentuk evaluasi kinerja adalah evaluasi implementasi SAKIP, yang bertujuan untuk mengetahui implementasi SAKIP dan mendorong peningkatan pencapaian kinerja yang tepat sasaran dan berorientasi hasil yang dilakukan melalui penilaian mandiri SAKIP oleh (1) Inspektur Jenderal untuk level 0 dan level I, dan (2) pimpinan Unit Organisasi Eselon I untuk level II dan level III.

Pelaksanaan evaluasi implementasi SAKIP secara khusus bertujuan untuk:

1. memperoleh informasi mengenai implementasi SAKIP;
2. menilai tingkat implementasi SAKIP;
3. menilai tingkat akuntabilitas kinerja;
4. memberikan saran perbaikan untuk peningkatan AKIP; dan
5. memonitor tindak lanjut rekomendasi hasil evaluasi periode sebelumnya.

Dalam pedoman Evaluasi AKIP yang diterbitkan Inspektorat Jenderal KKP, evaluasi implementasi SAKIP harus dapat memberikan simpulan hasil penilaian beberapa variabel, antara lain kriteria-kriteria yang ada dalam penerapan komponen-komponen manajemen kinerja yang meliputi perencanaan kinerja, pengukuran kinerja, pelaporan kinerja, dan evaluasi akuntabilitas kinerja internal sebagai fakta obyektif unit kerja mengimplementasikan SAKIP. Komponen-komponen tersebut kemudian dituangkan dalam Lembar Kerja Evaluasi (LKE) sesuai dengan kriteria masing-masing komponen. Variabel-variabel tersebut, yaitu:

1. Komponen
Terdiri dari Perencanaan Kinerja, Pengukuran Kinerja, Pelaporan Kinerja, dan Evaluasi Akuntabilitas Kinerja Internal.
2. Sub-komponen
Dibagi dengan gradasi Keberadaan, Kualitas, dan Pemanfaatan pada setiap komponen.
3. Kriteria
Merupakan gambaran kondisi yang perlu dicapai di setiap sub-komponen untuk dapat dinilai apakah kondisi tersebut sudah atau belum dicapai dan dapat digambarkan atau tidak.

LKE menyajikan komponen, sub-komponen, serta dilengkapi dengan kriteria penilaian, dengan bobot sebagaimana Tabel 1. Berdasarkan data dan informasi yang dikumpulkan, LKE kemudian dianalisis, dan digunakan sebagai bahan dasar dalam menyusun Lembar Hasil Evaluasi (LHE). Setiap kriteria yang dinilai pada LKE ini membutuhkan “*Professional Judgements*” dari Evaluator, karena terkait dengan penilaian kualitatif. LKE disusun dengan maksud sebagai kertas kerja evaluator dalam melakukan pengumpulan, penilaian, analisis, serta penyimpulan data dan informasi. Variabel dalam LKE AKIP disesuaikan dengan level organisasi

sebagaimana disajikan pada LKE AKIP masing-masing level organisasi dan secara garis besar dipetakan sebagaimana disajikan pada Lampiran 1.

Tabel 1 Pembobotan Nilai Evaluasi AKIP

Komponen	Sub-komponen			Total Bobot
	Keberadaan (20%)	Kualitas (30%)	Pemanfaatan (50%)	
Perencanaan Kinerja	6	9	15	30
Pengukuran Kinerja	6	9	15	30
Pelaporan Kinerja	3	4,5	7,5	15
Evaluasi Akuntabilitas Kinerja Internal	5	7,5	12,5	25
Nilai Akuntabilitas Kinerja	20	30	50	100

Penilaian dilakukan pada sub-komponen evaluasi AKIP, dan setiap sub-komponen dinilai berdasarkan pemenuhan kualitas dari kriteria. Pemenuhan kualitas dari kriteria dilakukan dengan mengacu LKE pada Lampiran 1. Sub-komponen akan dinilai dengan pilihan jawaban AA/A/BB/B/CC/C/D/E, jika kondisi atau gambaran kriteria sesuai dengan gradasi nilai pada Lampiran 2.

Setiap sub-komponen kriteria yang telah diberikan nilai dalam pemenuhan kriteria harus dilengkapi dengan catatan berupa keterangan beserta bukti daftar dokumen yang mendukung dan relevan. Nilai pada sub-komponen yang telah terisi akan terakumulasi sehingga diperoleh nilai total (hasil akhir) di setiap komponen.

Nilai hasil akhir dari penjumlahan komponen-komponen memberikan gambaran tingkat AKIP, dengan kategori predikat sebagaimana disajikan pada Lampiran 3.

Data hasil evaluasi SAKIP yang tersedia saat ini adalah data hasil penilaian mandiri 43 satker lingkup BPPSDMKP selama tahun 2022-2025. Hasil penilaian tersebut telah melalui rangkaian proses pendampingan, verifikasi dan validasi bukti data dukung capaian kinerja.

2.5 Aplikasi e-SAKIP Reviu

Kemen PANRB telah membangun sistem aplikasi Reviu AKIP dan penyampaian Laporan Kinerja Instansi Pemerintah melalui Aplikasi e-SAKIP Reviu yang dapat diakses melalui <https://esr.menpan.go.id/>. Pada Aplikasi e-SAKIP Reviu versi 1.0 terdapat tujuh dokumen yang wajib diunggah K/L. Pada tahun 2024, Kemen PANRB merilis Aplikasi e-SAKIP Reviu (Versi 2.0) dan terdapat 15 dokumen yang wajib diunggah K/L.

Penggunaan aplikasi e-SAKIP Reviu yang dimiliki oleh Kemen PANRB menjadi salah satu kebijakan secara nasional, namun masih memiliki beberapa keterbatasan, antara lain:

1. Aplikasi hanya digunakan untuk mengunggah sebagian dokumen yang digunakan dalam penilaian SAKIP.
2. Belum tersedia *dashboard analytics* yang memvisualisasikan hasil penilaian SAKIP, melainkan hanya pendataan dokumen SAKIP yang sudah dilaporkan.

3. Belum memanfaatkan kecerdasan buatan seperti arah kebijakan Pemerintah Digital ke depan.

2.6 Dashboard

Dashboard merupakan tampilan visual dari informasi terpenting yang dibutuhkan untuk mencapai satu atau beberapa tujuan, yang tergabung dan tersusun dalam satu layar tunggal sehingga bisa dipantau secara sekilas (Few 2006). *Dashboard* memberikan tampilan antarmuka dengan berbagai bentuk seperti diagram, laporan, indikator visual, mekanisme *alert*, yang dipadukan dengan informasi yang dinamis dan relevan (Gonzales 2006).

Menurut Rasmussen *et al.* (2009) dalam Sihombing *et al.* (2019), terdapat tiga jenis *dashboard*, yaitu:

1. *Dashboard Strategis (Strategic Dashboard)*
Dashboard strategis digunakan untuk mendukung level manajemen level strategis untuk memberikan informasi dalam membuat keputusan bisnis, memprediksi peluang, dan memberikan arahan pencapaian tujuan strategis.
2. *Dashboard Taktis (Tactical Dashboard)*
Dashboard tipe taktis berfokus pada proses analisis untuk menentukan penyebab dari suatu kondisi atau kejadian tertentu.
3. *Dashboard Operasional (Operational Dashboard)*
Dashboard operasional digunakan untuk memantau proses bisnis, aktivitas bisnis, dan hal yang kompleks. *Dashboard operasional* akan memberikan *update* harian atau mingguan atau grafik *real time* yang menggambarkan status proses bisnis dari suatu organisasi.

Menurut Bach *et al.* (2022), terdapat dua jenis *dashboard*, yaitu:

1. *Curated Dashboards*
Karakteristiknya yaitu sangat selektif terhadap data dan representasi visual; memiliki tujuan tertentu, misalnya untuk memberi tahu pemirsa tentang sesuatu yang khusus; dan bisa dianggap sebagai cerita yang digerakkan oleh penulis. Contohnya yaitu:
 - a. *Static Dashboard*
Merujuk pada konsep tradisional dasbor sebagai tampilan informasi terstruktur yang datar dan *non-interaktif*, dengan pola desain umum sebagai berikut. Pola desain dan contoh tampilan *static dashboard* dapat dilihat pada Lampiran 4.
 - b. *Magazine Dashboard*
Merupakan bagian integral dari artikel jurnalistik dan menyerupai visualisasi genre *majalah*. Pola desain dan contoh tampilan *magazine dashboard* dapat dilihat pada Lampiran 4.
 - c. *Infographic Dashboards*
Sering digunakan untuk merepresentasikan kumpulan data statis; misalnya, menyajikan cuplikan data penting secara bulanan atau tahunan. Pola desain dan contoh tampilan *infographic dashboard* dapat dilihat pada Lampiran 4.
 - d. *Embedded Mini Dashboards*
Dapat disematkan ke aplikasi lain seperti situs *web* berita. *Dashboard* mini yang ringkas ini hanya menempati area kecil di layar dan biasanya dilengkapi berbagai fitur interaktif untuk navigasi, atau untuk

memparameterisasi konten. Pola desain dan contoh tampilan *embedded mini dashboard* dapat dilihat pada Lampiran 4.

2. *Data Collection Dashboards*

Karakteristiknya yaitu bertujuan untuk mengirimkan informasi dalam jumlah besar; memungkinkan pemirsa mencari informasi yang paling relevan dengan kebutuhan mereka; dan bisa dianggap sebagai cerita yang digerakkan oleh pembaca.

a. *Analytic Dashboard*

Jenis ini umumnya menggunakan visualisasi lengkap (*signature charts* dan *trend arrows* yang lebih sederhana). Banyak elemen *dashboard* yang sepenuhnya interaktif, menyediakan fitur *pan and zoom*, *focus and context*, *tooltips*, *brushing and linking*, dan strategi eksplorasi serta navigasi lainnya. Pola desain dan contoh tampilan *analytic dashboard* dapat dilihat pada Lampiran 5.

b. *Repository Dashboard*

Memberikan interaksi dan biasanya menyediakan tautan untuk menjelajahi, memilih, dan akhirnya mengunduh data mentah. Data dan visualisasi diperbarui, dengan memilih visualisasi dan angka yang sangat umum untuk memvisualisasikan data. Pola desain dan contoh tampilan *repository dashboard* dapat dilihat pada Lampiran 5.

Lebih lanjut dalam Bach *et al.* (2022) menjelaskan pola desain komponen dalam *dashboard* meliputi:

1. *Data Information*

Kelompok ini mencakup jenis informasi yang ditampilkan di *dashboard*, contoh tampilan *data information* pada *dashboard* dapat dilihat pada Lampiran 6.

2. *Meta Information*

Kelompok ini menangkap informasi tambahan untuk memberikan konteks dan penjelasan tentang data, contoh tampilan *meta information* pada *dashboard* dapat dilihat pada Lampiran 6.

3. *Visual Representation*

Kelompok ini menjelaskan solusi umum untuk menyajikan informasi data dalam *dashboard*. Contoh tampilan *visual encoding* pada *dashboard* dapat dilihat pada Lampiran 6.

4. *Interaction*

Kelompok ini menjelaskan pendekatan interaksi umum yang terdapat dalam *dashboard*. Interaksi dapat terjadi melalui data interaktif, elemen antarmuka pengguna, dan interaksi tingkat jendela, contoh tampilan *interactions* pada *dashboard* dapat dilihat pada Lampiran 6.

Dalam Bach *et al.* (2022) juga mengidentifikasi beberapa kelompok pola komposisi: penggunaan ruang layar yang tersedia untuk merepresentasikan informasi (termasuk strategi untuk mengurangi jumlah informasi yang ditampilkan sekaligus), struktur beberapa halaman *dashboard* dan hubungan di antara halaman-halaman tersebut, *layout* komponen dalam satu halaman *dashboard*, dan penggunaan warna serta skema warna di seluruh *dashboard*. Contoh *layout* pada *dashboard* dan contoh penggunaan warna pada *dashboard* dapat dilihat pada Lampiran 6.

2.7 Large Language Model

Menurut Jana *et al.* (2024), *Large Language Model* (LLM) adalah model kecerdasan buatan (AI) canggih yang dirancang untuk memahami dan menghasilkan teks seperti manusia berdasarkan sejumlah besar data pelatihan. Model-model ini memanfaatkan teknik pembelajaran mendalam, khususnya jaringan saraf tiruan, untuk menganalisis pola dalam bahasa dan menghasilkan teks yang koheren dan relevan secara kontekstual. Arsitektur LLM biasanya terdiri dari beberapa lapisan unit jaringan saraf tiruan, dengan arsitektur transformator menjadi yang paling menonjol karena kemampuannya untuk menangani dependensi jarak jauh dalam teks. Model-model ini dilatih menggunakan teknik-teknik seperti *supervised learning*, *self-supervised learning*, dan *transfer learning*, di mana mereka belajar memprediksi kata berikutnya dalam suatu urutan berdasarkan konteks. Contoh-contoh LLM yang terkenal termasuk seri GPT OpenAI, BERT Google, dan Roberta Facebook. LLM telah menunjukkan kemampuan luar biasa dalam berbagai tugas *Natural Language Processing* (NLP), termasuk pembuatan teks, peringkasan, penerjemahan, analisis sentimen, dan tanya jawab. LLM telah diintegrasikan ke dalam berbagai aplikasi, mulai dari asisten virtual dan *chatbot* hingga alat pembuatan konten dan sistem rekomendasi. Teknologi dan arsitektur LLM terdiri dari:

1. Neural Network Architecture

a. Transformer

Arsitektur *Transformer*, yang diperkenalkan oleh Vaswani dkk. dalam makalah penting "*Attention is All You Need*", telah menjadi standar *de facto* untuk sistem *Large Language Model* (LLM). Model *Transformer* mengandalkan mekanisme *self-attention* untuk menangkap ketergantungan antar kata dalam suatu urutan, sehingga memungkinkannya memproses konteks jarak jauh secara efisien. Model ini terdiri dari beberapa lapisan jaringan saraf *self-attention* dan *feedforward*, yang memungkinkannya mempelajari representasi kontekstual kata dan hubungannya dalam sebuah teks. *Transformer* telah menunjukkan kinerja yang unggul dalam berbagai tugas NLP dan merupakan tulang punggung LLM mutakhir seperti seri GPT OpenAI dan BERT Google.

b. Long Short-Term Memory (LSTM)

Jaringan LSTM adalah jenis arsitektur *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi masalah gradien menghilang dalam RNN tradisional. LSTM menggabungkan *gated units*, termasuk gerbang *input*, *forget*, dan *output*, yang mengatur aliran informasi melalui jaringan dari waktu ke waktu. Hal ini memungkinkan *LSTM* untuk menangkap dependensi jangka panjang dalam data sekuensial, sehingga cocok untuk tugas-tugas yang melibatkan pembuatan teks, analisis sentimen, dan pemodelan bahasa. Meskipun LSTM telah banyak digunakan dalam LLM di masa lalu, namun sebagian besar telah dilampaui oleh arsitektur *Transformer* karena kinerjanya yang unggul pada kumpulan data berskala besar.

c. Gated Recurrent Unit (GRU)

Serupa dengan LSTM, GRU merupakan varian RNN yang dirancang untuk mengatasi kekurangan RNN tradisional. GRU menggabungkan

fungsi gerbang masukan dan gerbang lupa menjadi satu gerbang pembaruan, menyederhanakan arsitektur dan mengurangi kompleksitas komputasi. Meskipun GRU kurang kompleks dibandingkan LSTM, GRU telah terbukti mencapai kinerja yang sebanding dalam tugas-tugas NLP tertentu, menjadikannya alternatif yang layak untuk LLM skala kecil atau aplikasi dengan sumber daya komputasi terbatas.

2. Teknik Pelatihan

a. *Supervised Learning*

Dalam konteks LLM, *supervised learning* sering digunakan untuk menyempurnakan model yang telah dilatih sebelumnya pada tugas-tugas hilir tertentu, seperti klasifikasi teks, pengenalan entitas bernama, atau analisis sentimen. Penyempurnaan terawasi memungkinkan LLM untuk beradaptasi dengan domain atau tugas tertentu dengan memperbarui parameter model berdasarkan data pelatihan spesifik tugas.

b. *Self-Supervised Learning*

Dalam LLM, pembelajaran mandiri biasanya digunakan untuk pra-pelatihan, di mana model belajar memprediksi token yang tersamar atau rusak dalam korpus teks. Tujuan pra-pelatihan ini mendorong model untuk mempelajari representasi bahasa yang bermakna yang dapat digeneralisasi ke berbagai tugas selanjutnya.

c. *Transfer Learning*

Dalam konteks LLM, *transfer learning* umumnya digunakan untuk menginisialisasi parameter model menggunakan bobot yang telah dilatih sebelumnya dari tugas pemodelan bahasa skala besar, seperti memprediksi kata berikutnya dalam urutan teks. Dengan mentransfer pengetahuan dari model yang telah dilatih sebelumnya ke tugas-tugas selanjutnya, LLM dapat mencapai kinerja yang lebih baik dengan data berlabel dan sumber daya komputasi yang lebih sedikit.

3. *Preprocessing and Postprocessing Methods*

a. *Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs)*

GPU dan TPU adalah akselerator perangkat keras khusus yang dirancang untuk pemrosesan paralel operasi matriks, yang umum dalam pelatihan jaringan saraf tiruan. Akselerator ini dapat mempercepat waktu pelatihan secara signifikan dibandingkan dengan CPU tradisional, sehingga sangat cocok untuk pelatihan LLM. TPU, khususnya, dioptimalkan untuk beban kerja pembelajaran mesin dan telah digunakan oleh organisasi seperti Google untuk melatih model bahasa skala besar seperti BERT dan T5.

b. *Distributed Training*

Pelatihan terdistribusi melibatkan paralelisasi proses pelatihan di beberapa node komputasi atau perangkat untuk mempercepat pelatihan dan menangani set data yang lebih besar. Hal ini biasanya melibatkan partisi data pelatihan dan parameter model di beberapa node dan koordinasi komunikasi serta sinkronisasi pembaruan selama pelatihan. Kerangka kerja pelatihan terdistribusi seperti TensorFlow dan PyTorch menyediakan API untuk pelatihan terdistribusi, yang memungkinkan peneliti dan teknisi untuk meningkatkan skala pelatihan LLM ke kluster GPU atau TPU.

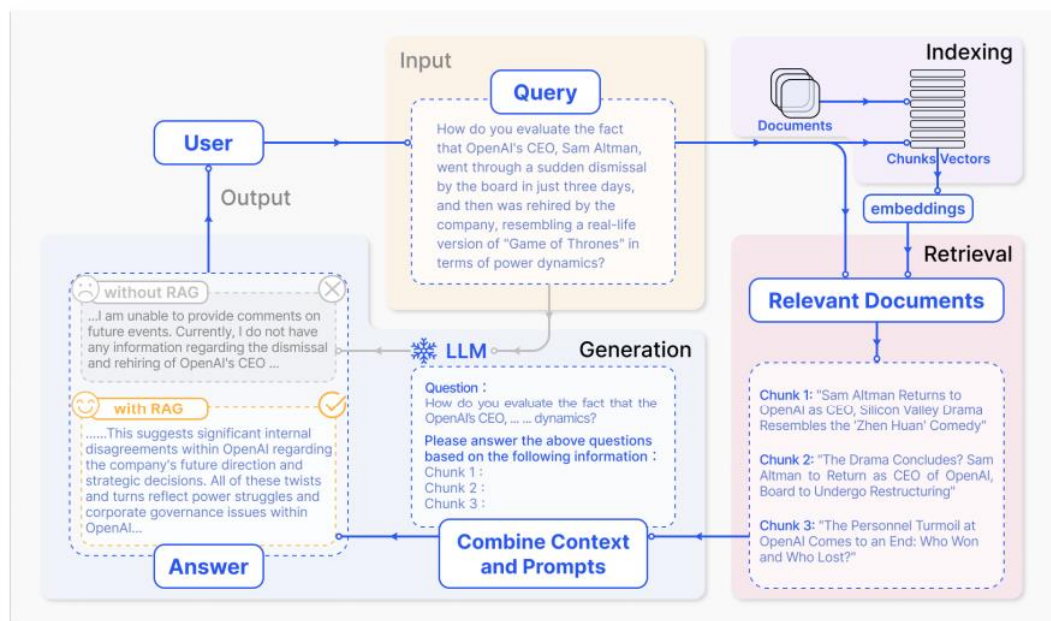
c. *Model Parallelism and Data Parallelism*

Paralelisme model melibatkan pembagian lapisan-lapisan jaringan saraf tiruan ke beberapa perangkat, yang memungkinkan setiap perangkat menghitung lintasan maju dan mundur untuk sebagian model. Di sisi lain, paralelisme data melibatkan replikasi model ke beberapa perangkat dan memparalelkan proses pelatihan dengan mendistribusikan sejumlah data ke setiap perangkat. Kedua pendekatan ini dapat dikombinasikan untuk memanfaatkan sumber daya komputasi yang tersedia dalam lingkungan pelatihan terdistribusi secara efektif.

2.8 Retrieval Augmented Generation

LLM memiliki kemampuan yang luar biasa namun masih memiliki keterbatasan, salah satunya yaitu masalah halusinasi (memberikan respon seolah serius tetapi sebenarnya tidak relevan dengan query yang diminta oleh pengguna). Menurut Gao *et al.* (2024), *Retrieval-Augmented Generation* (RAG) meningkatkan LLM dengan mengambil potongan dokumen yang relevan dari basis pengetahuan eksternal melalui perhitungan kesamaan semantik.

Aplikasi RAG diilustrasikan seperti terlihat pada Gambar 4.



Gambar 4 Contoh representatif proses RAG yang diterapkan pada tanya jawab
Sumber: Gao *et al.* (2024)

Berdasarkan gambar di atas, seorang pengguna mengajukan pertanyaan kepada ChatGPT tentang berita terbaru yang banyak dibicarakan. Mengingat ChatGPT bergantung pada data pra-pelatihan, awalnya ChatGPT tidak memiliki kapasitas untuk memberikan informasi terbaru tentang perkembangan terkini. RAG menjembatani kesenjangan informasi ini dengan mencari dan menggabungkan pengetahuan dari basis data eksternal. Dalam hal ini, RAG mengumpulkan artikel berita relevan yang terkait dengan pertanyaan pengguna. Artikel-artikel ini, dikombinasikan dengan pertanyaan awal, membentuk prompt komprehensif yang memungkinkan LLM untuk menghasilkan jawaban yang tepat.

Paradigma penelitian RAG terus berkembang, dan Gao *et al.* (2024) mengkategorikannya menjadi tiga tahap:

1. *Naive RAG*

Naive RAG terdiri dari tiga komponen, antara lain:

- a. *Indexing*

Dokumen dipecah menjadi potongan-potongan kecil, dikodekan menjadi vektor, dan disimpan dalam basis data vektor.

- b. *Retrieval*

Mengambil k potongan teratas yang paling relevan dengan pertanyaan berdasarkan kesamaan semantik.

- c. *Generation*

Masukkan pertanyaan awal dan potongan-potongan yang diambil bersama-sama ke dalam LLM untuk menghasilkan jawaban akhir.

2. *Advanced RAG*

Advanced RAG mengusulkan beberapa strategi optimasi seputar *pre-retrieval* dan *post-retrieval*, dengan proses yang mirip dengan *Naive RAG*, namun tetap mengikuti *chain-like structure*.

3. *Modular RAG*

Modular RAG mewarisi dan mengembangkan paradigma sebelumnya, yang secara keseluruhan menunjukkan fleksibilitas yang lebih besar. Hal ini terlihat dari pengenalan beberapa modul fungsional spesifik dan penggantian modul yang sudah ada. Proses keseluruhannya tidak terbatas pada pengambilan dan pembangkitan sekuensial; tetapi juga mencakup metode seperti pengambilan iteratif dan adaptif.

2.9 Retrieval Augmented Generation Assessment

Menurut Es *et al.* (2024), *Retrieval Augmented Generation Assessment* atau RAGAS merupakan sebuah *framework* yang digunakan untuk melakukan alur kerja dari RAG yang berfokus pada pengaturan di mana referensi jawaban mungkin tidak tersedia, jadi perlu menilai "kebenaran" jawaban dengan cara lain selain hanya menilai apakah dokumen yang diambil (*retrieved passages*) berguna atau tidak. Fokus pada lima dimensi kualitas utama yang diukur, antara lain:

1. *Faithfulness* (F), yaitu prinsip bahwa jawaban harus didasarkan pada konteks yang tersedia untuk menghindari halusinasi dan memastikan konteks yang diambil dapat menjadi dasar membenaran jawaban. Hal ini sangat penting pada sistem RAG, terutama dalam bidang yang menuntut konsistensi fakta tinggi seperti hukum, di mana informasi terus berkembang.

Jika diberikan pernyataan jawaban $a_s(q)$ berdasarkan *query* q , kemudian pecah setiap kalimat menjadi beberapa pernyataan, $S(a_s(q))$. Lalu tentukan apakah setiap $S(a_s(q))$ dapat disimpulkan berdasarkan konteks yang didapatkan berdasarkan *query*, $c(q)$, seperti terlihat pada rumus persamaan 1 berikut:

$$F = \frac{|V|}{|S|} \quad (1)$$

Dimana:

$|V|$ adalah banyaknya pernyataan yang didukung oleh konteks,

$|S|$ adalah banyaknya pernyataan

2. *Answer Relevancy* (AR), yaitu jawaban yang dihasilkan harus benar-benar menjawab pertanyaan yang diajukan.

Jawaban $a_s(q)$ dianggap relevan jika jawaban tersebut secara langsung menjawab pertanyaan dengan cara yang tepat. Secara khusus, penilaian terhadap relevansi jawaban tidak memperhitungkan faktor-faktor kualitas, tetapi memberikan penalti jika jawabannya tidak lengkap atau berisi informasi yang berlebihan. Untuk memperkirakan relevansi jawaban, untuk jawaban yang diberikan $a_s(q)$, maka LLM diminta untuk menghasilkan n pertanyaan potensial q_i berdasarkan $a_s(q)$. Untuk setiap q_i , dihitung kesamaan $sim(q, q_i)$ dengan pertanyaan awal q , sebagai *cosine* antara *embedding* yang sesuai. Skor relevansi jawaban, AR, untuk pertanyaan q kemudian dihitung dengan rumus persamaan 2 berikut:

$$AR = \frac{1}{n} \sum_{i=1}^n sim(q, q_i) \quad (2)$$

3. *Context Precision* (CP), yaitu metrik yang mengevaluasi apakah semua item relevan yang ada di dalam data sebenarnya (*contexts*) diberi peringkat lebih tinggi atau tidak. Idealnya, semua bagian yang relevan harus muncul di peringkat teratas. Metrik ini dihitung menggunakan *question*, *ground_truth* dan *contexts*, dengan nilai berkisar antara 0 dan 1, di mana skor yang lebih tinggi menunjukkan presisi yang lebih baik, seperti terlihat pada rumus persamaan 3 berikut:

$$CP@K = \frac{\sum_{k=1}^K (P@k \times v_k)}{\text{Total number of relevant items in the top } K \text{ results}} \quad (3)$$

Nilai *precision* (P) untuk masing-masing k dihitung menggunakan rumus persamaan 4 berikut:

$$P = \frac{TP}{(TP + FP)} \quad (4)$$

Di mana K adalah jumlah total potongan di dalam *contexts* dan $v_k \in \{0,1\}$ adalah indikator relevansi pada peringkat.

4. *Context Recall* (CR), mengukur kelengkapan konteks yang diambil, yaitu apakah semua informasi yang diperlukan untuk menjawab pertanyaan berhasil ditemukan, yang dihitung menggunakan rumus persamaan 5 berikut:

$$CR = \frac{|GT \text{ claims that can be attributed to context}|}{|Number of claims in GT|} \quad (5)$$

5. *Answer Correctness* (AC), membandingkan jawaban sistem dengan *ground truth* untuk menilai kebenaran faktual. Nilai *Answer Correctness* adalah rata-rata tertimbang (*weighted average*) dari kesamaan faktual dan kesamaan semantik antara jawaban sistem dengan jawaban referensi (*ground truth*), yang dihitung menggunakan rumus persamaan 6 berikut:

$$AC = (\omega_1 \times FS) + (\omega_2 \times SS) \quad (6)$$

Rumus persamaan 6 di atas dapat dijelaskan sebagai berikut:

- a. *Factual Similarity* (FS)

Komponen ini dihitung dengan mengekstraksi fakta-fakta tunggal dari jawaban yang dihasilkan (sistem) dan jawaban referensi (*ground truth*), lalu mengukurnya menggunakan rumus *F1-Score*. RAGAS menggunakan LLM untuk mengkategorikan fakta-fakta tersebut menjadi tiga kelompok:

- TP (*True Positives*): Fakta yang ada di jawaban yang dihasilkan dan ada di *ground truth*.
- FP (*False Positives*): Fakta yang ada di jawaban yang dihasilkan, tetapi tidak ada di *ground truth* (berpotensi halusinasi).

- FN (*False Negatives*): Fakta yang ada di *ground truth*, tetapi hilang di jawaban yang dihasilkan (informasi yang terlewat).

Factual Similarity dihitung menggunakan rumus persamaan 7 berikut:

$$F1 = \frac{TP}{(TP + 0,5 \times (FP + FM))} \quad (7)$$

b. *Semantic Similarity* (SS)

Komponen ini tidak mempedulikan fakta spesifik, melainkan mengukur apakah makna atau intisari dari jawaban sistem sama dengan jawaban referensi, terlepas dari perbedaan kata-kata (*parafrase*).

Biasanya, ini dihitung dengan mengubah kedua jawaban menjadi bentuk vektor angka menggunakan model *embedding*, lalu menghitung jarak atau sudut di antara kedua vektor tersebut. *Cosine Similarity* dihitung menggunakan rumus persamaan 8 berikut:

$$SS = \frac{A \cdot B}{||A|| \times ||B||} \quad (8)$$

Dimana A adalah *vector embedding* dari jawaban sistem, dan B adalah *vector embedding* dari *ground truth*.

2.10 Model Gemini 2.5 Pro

Model Gemini 2.5 Pro berevolusi dari sekadar model prediksi linguistik berbasis probabilitas menjadi *thinking model* (model bernalar) yang secara natif mengintegrasikan penalaran kognitif sebelum menghasilkan keluaran (Comanici *et al.* 2025). Keunggulan analitis dari Gemini 2.5 Pro secara empiris terlihat pada kemampuannya untuk mengungguli berbagai tolok ukur evaluasi kecerdasan buatan terdepan pada pemahaman level lanjut, khususnya dalam ranah matematika, sains STEM, dan penalaran logis (Team *et al.* 2025). Manfaat langsung dari arsitektur bernalar ini berdampak pada tingginya akurasi faktual dan mitigasi fenomena halusinasi secara drastis jika dibandingkan dengan arsitektur standar (Widhani *et al.* 2026). Sementara itu, studi Leutz-Schmidt *et al.* (2025) yang membandingkan performa ChatGPT dan Gemini pada pertanyaan manajemen kerja di bidang radiologi turut memperkuat gambaran bahwa keunggulan Gemini bersifat kontekstual, bergantung pada domain dan jenis tugas yang diberikan.

Menurut Bayer *et al.* (2024), ChatGPT yang dibuat oleh OpenAI dan Gemini yang dibuat oleh Google adalah dua alat bahasa kecerdasan buatan yang paling populer saat ini. Penelitian Rane *et al.* (2024) menunjukkan bahwa ChatGPT cocok untuk interaksi yang menyenangkan, sementara Gemini unggul dalam tugas-tugas yang membutuhkan akurasi. Keunggulan jendela konteks Gemini 2.5 Pro yang mencapai satu juta token juga menjadi salah satu faktor kunci yang membedakannya dari model kompetitor seperti Claude 3.7 Sonnet maupun DeepSeek R1, sehingga memungkinkan model memproses dokumen atau basis kode berukuran besar dalam satu kali proses tanpa harus memecahnya menjadi potongan-potongan kecil.

2.11 Facebook AI Similarity Search (FAISS)

Dalam ekosistem RAG, FAISS merupakan salah satu pustaka pencarian kemiripan vektor yang paling banyak diadopsi untuk mendukung proses *retrieval* pada tahap augmentasi pengetahuan eksternal. Menurut Douze *et al.* (2025), FAISS dirancang sebagai *toolkit* metode pengindeksan dan primitif terkait untuk melakukan pencarian, klusterisasi, kompresi, serta transformasi vektor berdimensi

tinggi, dengan menawarkan ruang kompromi (*trade-off space*) yang fleksibel antara tingkat akurasi pencarian, kecepatan kueri, dan efisiensi penggunaan memori baik melalui pencarian eksak (*IndexFlat*) maupun metode aproksimasi seperti *Inverted File Index* (IVF) dan *Hierarchical Navigable Small World* (HNSW). Kelebihan arsitektural ini menjadikan FAISS mampu menangani miliaran vektor dalam satu server tunggal, sekaligus mendukung akselerasi GPU untuk mempercepat proses pencarian pada skala data yang sangat besar.

Sebagian besar implementasi RAG di lingkungan *enterprise* (sekitar 80,5% dari 63 studi yang dianalisis) mengandalkan kerangka *retrieval* standar seperti FAISS dan menegaskan posisinya sebagai salah satu infrastruktur *retrieval* paling matang dan banyak digunakan dalam praktik industri saat ini (Karakurt dan Akbulut 2026). Manfaat FAISS dalam konteks RAG juga tervalidasi secara empiris melalui studi implementasi langsung oleh Burbano *et al.* (2026) dimana integrasi FAISS dengan model bahasa Mistral-7B-Instruct dalam sistem *chatbot* RAG berbasis akses basis data terstruktur berhasil mencapai tingkat keberhasilan respons sebesar 80%, dengan kegagalan yang tersisa umumnya disebabkan oleh keterbatasan *retrieval* dalam menemukan konteks yang paling relevan terhadap kueri pengguna. Pada domain yang menuntut akurasi tinggi seperti kesehatan, tinjauan komprehensif oleh Neha *et al.* (2025) turut menempatkan FAISS sebagai komponen arsitektural kunci dalam berbagai varian sistem RAG (baik yang bersifat naif, lanjutan, maupun modular) untuk mendukung aplikasi dukungan diagnostik, ringkasan rekam medis elektronik, dan penjawaban pertanyaan medis.

2.12 Evaluasi Model

Ada beberapa metrik yang digunakan untuk mengevaluasi kinerja model, yang paling umum adalah akurasi, *precision*, *recall*, dan *F1-score*. Metrik-metrik ini didasarkan pada *confusion matrix*, yaitu tabel yang mengukur hasil prediksi model terhadap data sebenarnya, seperti terlihat pada Gambar 5.

		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Gambar 5 *Confussion Matrix*

Penjelasan *Evaluation Matrix* dapat dilihat pada Tabel 2.

Tabel 2 *Evaluation Matrix*

Metrix Evaluasi	Uraian
<i>Accuracy</i>	Mengukur seberapa sering model memprediksi dengan benar secara keseluruhan yang dihitung melalui rumus persamaan 9.
$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$	

Tabel 2 *Evaluation Matrix (lanjutan)*

Metrix Evaluasi	Uraian
<i>Precision</i>	Menilai berapa banyak prediksi positif yang benar dari total prediksi positif. Penting saat kita ingin mengurangi <i>false positive</i> yang dihitung melalui rumus persamaan 10. $Precision = \frac{TP}{TP+FP} \quad (10)$
<i>Recall</i>	Mengukur seberapa banyak sampel positif yang berhasil ditemukan oleh model dari total sampel positif. Penting saat <i>false negative</i> harus diminimalkan, yang dihitung melalui rumus persamaan 11. $Recall = \frac{TP}{TP+FN} \quad (11)$
<i>F1-Score</i>	Menggabungkan <i>precision</i> dan <i>recall</i> , sangat cocok digunakan ketika kita menghadapi ketidakseimbangan data dan ingin mendapatkan keseimbangan, yang dihitung melalui rumus persamaan 12. $F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$

Dalam konteks sistem RAG, konsep *Confusion Matrix* diadaptasi untuk menilai baik tahap pengambilan (*retrieval*) maupun tahap pembuatan (*generation*) dari model-model tersebut. Afzal *et al.* (2025) menyusun *confusion matrix* berdasarkan fitur-fitur yang sama dengan yang dievaluasi oleh LLM. Studi tersebut juga menyebutkan bahwa sebuah jawaban dianggap dapat diterima jika skor evaluasi yang diberikan melampaui ambang batas tertentu. Dengan demikian, *Confusion Matrix* RAG mengevaluasi:

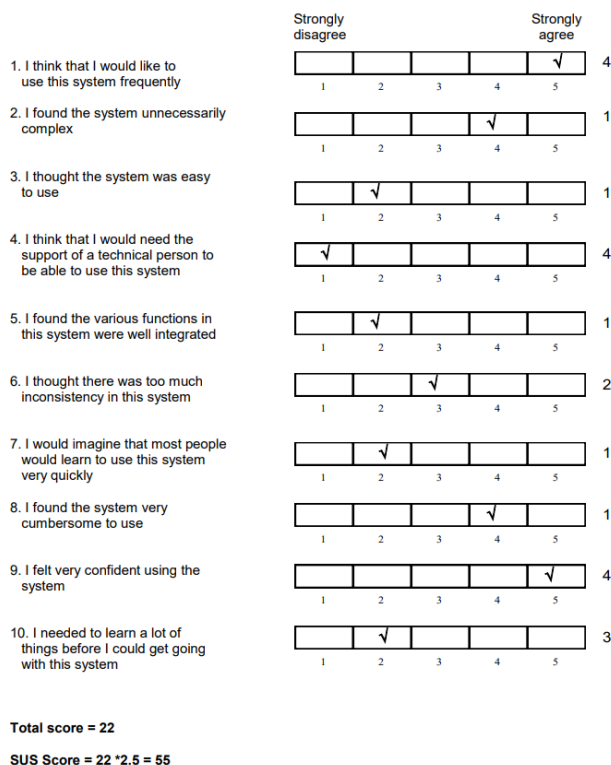
1. *True Positive* (TP): Dokumen relevan berhasil diambil dan respons yang dapat diterima berhasil dibuat.
2. *True Negative* (TN): Dokumen relevan tidak diambil dan respons yang dapat diterima tidak dibuat.
3. *False Positive* (FP): Dokumen relevan tidak diambil namun respons yang dapat diterima berhasil dibuat.
4. *False Negative* (FN): Dokumen relevan berhasil diambil namun respons yang dapat diterima tidak dibuat.

2.13 *System Usability Scale* (SUS)

Menurut Brooke (1996), *System Usability Scale* (SUS) adalah skala yang mudah digunakan dengan sepuluh pertanyaan yang memberikan gambaran umum mengenai penilaian pribadi terhadap seberapa bergunanya sebuah sistem. SUS menghasilkan angka tunggal yang mewakili ukuran gabungan dari keseluruhan kegunaan sistem yang diteliti. Perlu dicatat bahwa skor untuk masing-masing item tidak bermakna secara terpisah. Untuk menghitung skor SUS, pertama-tama jumlahkan kontribusi skor dari setiap item. Kontribusi skor setiap item berkisar antara 0 hingga 4. Untuk item 1, 3, 5, 7, dan 9, kontribusi skor adalah posisi skala dikurangi 1. Untuk item 2, 4, 6, 8, dan 10, kontribusinya adalah 5 dikurangi posisi skala. Kalikan jumlah skor dengan 2,5 untuk memperoleh nilai keseluruhan SU.

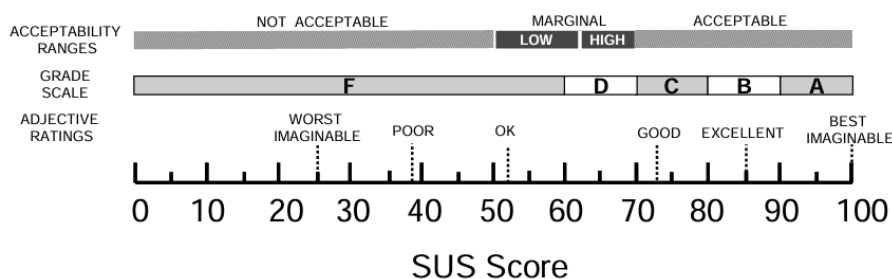
Skor SUS memiliki rentang 0 hingga 100. Contoh skala SU yang dinilai dapat dilihat pada Gambar 6.

© Digital Equipment Corporation, 1986.



Gambar 6 Contoh Pengukuran *System Usability*

Untuk membantu praktisi menginterpretasikan skor SUS individual dan membantu menjelaskan hasilnya kepada para profesional di luar bidang faktor manusia, maka Bangor *et al.* (2009) melakukan perbandingan peringkat kata sifat (*adjective ratings*), skor penerimaan (*acceptability scores*), dan skala penilaian (*school grading score*) dalam kaitannya dengan skor SUS rata-rata, seperti terlihat pada Gambar 7.



Gambar 7 *Adjective Rating SUS*

2.14 Wilcoxon Signed-Rank Test

Dalam Woolson (2008) menyampaikan bahwa *Wilcoxon Signed-Rank Test* yang dikembangkan Wilcoxon (1945) merupakan prosedur uji non-parametrik yang digunakan untuk analisis data berpasangan atau untuk masalah satu sampel. Dalam

pengaturan berpasangan, uji ini digunakan untuk menguji hipotesis bahwa distribusi probabilitas sampel pertama sama dengan distribusi probabilitas sampel kedua.

Dalam Altarazi *et al.* (2023), mengusulkan *framework* untuk mengembangkan dasbor interaktif dan mengevaluasi efektivitas penggunaannya dalam meningkatkan manajemen data, analisis data, visualisasi data, dan berbagi informasi. *Framework* ini terdiri dari empat fase utama antara lain pra-penilaian untuk menilai situasi terkini, standarisasi format basis data, pembuatan dasbor interaktif, dan pasca-penilaian untuk mengevaluasi perbaikan yang disarankan. Pengujian hipotesa menggunakan *Uji Wilcoxon Signed-Rank*, dengan terlebih dahulu menghitung nilai W melalui rumus persamaan 13 berikut:

$$W = \sum R_i \quad (13)$$

di mana R_i adalah peringkat dari selisih positif antara pasangan data.

Nilai Z dihitung melalui rumus persamaan 14 berikut:

$$Z = \frac{W - \left[\frac{n(n+1)}{4} \right]}{\sqrt{\frac{(n)(n+1)(2n+1)}{24} - \frac{t^3 - t}{48}}} \quad (14)$$

Untuk signifikansi ditentukan jika probabilitas (*Asymp.Sig*) < 0,05 maka H_0 ditolak artinya terdapat perbedaan, tetapi jika probabilitas (*Asymp.Sig*) > 0,05 maka H_0 diterima artinya tidak terdapat perbedaan.

2.15 Penelitian Sebelumnya

Salah satu dasar penelitian ini mengacu pada Permen PANRB Nomor 88 Tahun 2021, sehingga penelitian SAKIP dan aplikasinya tergolong masih terbatas, terutama yang membahas penilaian SAKIP secara rinci menggunakan teknologi informasi termasuk kaitannya dengan bantuan kecerdasan buatan.

Beberapa studi literatur yang ditemui dan terkait dengan penelitian dapat dilihat pada Tabel 3.

Tabel 3 Penelitian Terkait SAKIP

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Salomo dan Rahmayanti 2023)	Pemerintah Pusat dan Daerah	Penelitian ini membahas kemajuan dan tantangan kelembagaan dalam reformasi Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) di tingkat pemerintah daerah di Indonesia. Penelitian ini menganalisis sejauh mana reformasi SAKIP telah diimplementasikan (berdasarkan data tahun 2016–2019), serta mengeksplorasi faktor-faktor institusional yang memengaruhi efektivitas penerapannya.	Penelitian ini menemukan bahwa: 1. Kemajuan implementasi SAKIP di daerah cenderung lambat dan dominan formalistik, terutama di tingkat kabupaten/kota dibanding provinsi dan pusat. 2. Faktor-faktor kelembagaan sangat berpengaruh terhadap efektivitas reformasi: a. Regulatorif: koordinasi antar-regulasi masih belum sinkron, dan belum ada sanksi kuat bagi pemerintah daerah dengan nilai SAKIP rendah.	Penelitian ini relevan karena menyediakan <i>evidence-based insight</i> bahwa reformasi SAKIP saat ini mengalami stagnasi di tataran implementasi dan pemahaman sistemik dan mengindikasikan <i>dashboard</i> aplikasi SAKIP yang dirancang dengan baik dapat menjadi alat bantu transformasional.

Tabel 3 Penelitian Terkait SAKIP (*lanjutan*)

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Salomo dan Rahmayanti 2023)	Pemerintah Pusat dan Daerah	Penelitian ini menggunakan pendekatan <i>mixed methods</i> , yaitu analisis data kuantitatif dari evaluasi Kementerian PAN-RB serta wawancara mendalam dengan pejabat pusat dan daerah.	b. Normatif: persepsi yang salah bahwa serapan anggaran adalah ukuran kinerja utama. c. Budaya Kognitif: budaya feodalisme dan rendahnya <i>sense of ownership</i> terhadap SAKIP menghambat internalisasi nilai akuntabilitas. d. Manajerial: kelemahan dalam fungsi perencanaan, kapasitas SDM, dan belum terintegrasinya janji politik ke dalam dokumen Renstra. 3. Efisiensi anggaran akibat implementasi SAKIP pada 2018 diperkirakan mencapai lebih dari Rp65 triliun, namun perubahan tersebut belum menjangkau perubahan budaya atau pemahaman mendalam.	Kelemahan yang ditemui pada penelitian ini yaitu tidak membahas penggunaan <i>dashboard</i> , infografis, atau visualisasi data interaktif sebagai solusi penyajian hasil evaluasi SAKIP.
(Lestari dan Salomo 2022)	Penulis meninjau lebih dari 40 penelitian (baik nasional maupun internasional)	Penelitian ini membahas kajian literatur terkait implementasi Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) pada pemerintah daerah di Indonesia, serta perbandingannya dengan praktik manajemen kinerja daerah di negara lain. Penulis meninjau lebih dari 40 penelitian (baik nasional maupun internasional) untuk mengidentifikasi tantangan utama dalam penerapan SAKIP, serta faktor-faktor yang mendukung atau menghambat efektivitas sistem ini di lingkungan pemerintah daerah.	Temuan utama dari kajian ini: 1. Permasalahan dominan dalam penerapan SAKIP di daerah berasal dari rendahnya: a. Kompetensi dan profesionalisme SDM. b. Kapasitas organisasi (struktur birokrasi, koordinasi, dan dukungan TI). c. Ketersediaan indikator kinerja yang valid dan dapat diukur. d. Integrasi sistem pelaporan dan evaluasi dalam proses manajemen kinerja. 2. Faktor pendukung keberhasilan implementasi SAKIP: a. Komitmen pimpinan daerah dalam mendukung reformasi kinerja. b. Tekanan kelembagaan eksternal, termasuk dari legislatif dan partisipasi publik, dalam meningkatkan transparansi dan akuntabilitas.	Penelitian ini relevan karena menunjukkan adanya fragmentasi dan keterbatasan data dalam sistem kinerja saat ini yang tidak didukung oleh aplikasi yang efektif. Kelemahan yang ditemui pada penelitian ini yaitu hanya berfokus pada tinjauan literatur dan belum mengusulkan solusi konkret, terutama dalam bentuk teknologi visualisasi data atau sistem berbasis digital.

Tabel 3 Penelitian Terkait SAKIP (*lanjutan*)

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Lestari dan Salomo 2022)	Penulis meninjau lebih dari 40 penelitian (baik nasional maupun internasional)	Penelitian ini memetakan permasalahan dalam implementasi manajemen kinerja publik, terutama pada aspek perencanaan strategis, pengukuran kinerja, pelaporan, evaluasi, dan pemanfaatan teknologi informasi dalam konteks akuntabilitas.	Secara global, tantangan serupa juga ditemukan di negara lain, namun terdapat lebih banyak perhatian pada integrasi data, penggunaan <i>dashboard</i> visualisasi kinerja, dan strategi komunikasi berbasis teknologi.	
(Pratama 2024)	Kementerian Kesehatan Republik Indonesia	Tujuan utamanya adalah untuk menganalisis efektivitas implementasi SAKIP, menguji pengaruh penganggaran berbasis kinerja, sistem pengendalian intern, dan sistem pelaporan keuangan terhadap SAKIP, serta merumuskan strategi peningkatan implementasi sistem tersebut. Penelitian ini mengisi kesenjangan dalam literatur terkait tata kelola SAKIP di tingkat instansi pemerintah pusat. Sebelumnya, banyak penelitian dilakukan pada instansi pemerintah daerah, tetapi penelitian mengenai implementasi dan efektivitas SAKIP di tingkat kementerian belum banyak dilakukan	Hasil penelitian menunjukkan bahwa efektivitas SAKIP sangat dipengaruhi oleh penganggaran berbasis kinerja dan sistem pengendalian intern, sedangkan sistem pelaporan keuangan tidak memiliki dampak yang signifikan. Oleh karena itu, strategi peningkatan SAKIP perlu difokuskan pada perbaikan dalam dua aspek tersebut	Penelitian ini relevan karena dapat memberikan wawasan tentang faktor-faktor yang mempengaruhi implementasi SAKIP di tingkat kementerian. Selain itu, hasil studi ini dapat digunakan sebagai dasar untuk pengembangan kebijakan peningkatan akuntabilitas kinerja instansi pemerintah. Kelemahan yang ditemui pada penelitian ini yaitu belum menunjukkan solusi konkret perbaikan SAKIP berbasis teknologi informasi.
(Raharja <i>et al.</i> 2018)	Pemerintah Kota Bogor	Penelitian ini bertujuan untuk menganalisis kelemahan dalam penerapan sistem akuntabilitas kinerja dan merumuskan strategi yang dapat meningkatkannya.	Hasil penelitian menunjukkan bahwa strategi yang dapat diterapkan untuk meningkatkan akuntabilitas kinerja meliputi: 1. Penguatan komitmen dari pimpinan hingga staf. 2. Penerapan anggaran berbasis kinerja untuk memastikan efisiensi dan efektivitas penggunaan dana. 3. Pembinaan SAKIP melalui pelatihan dan bimbingan teknis.	Penelitian ini relevan karena kita dapat menggunakan metode IPA dan AHP dapat digunakan untuk menilai efektivitas sistem yang ada serta menentukan langkah-langkah peningkatan yang lebih terarah. Kelemahan yang ditemui pada penelitian ini yaitu tidak ada bukti pengembangan alat bantu atau sistem digital yang bisa digunakan oleh pemerintah daerah.

Tabel 3 Penelitian Terkait SAKIP (*lanjutan*)

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Raharja <i>et al.</i> 2018)	Pemerintah Kota Bogor	Metode <i>Importance Performance Analysis</i> (IPA) untuk mengetahui persepsi responden terhadap tingkat kepentingan dan kinerja mengenai implemementasi SAKIP. Hasil analisis tersebut, akan digunakan sebagai dasar perumusan strategi peningkatan akuntabilitas kinerja menggunakan metode <i>Analytic Hierarchy Process</i> (AHP)	4. Peningkatan kualitas Aparat Pengawasan Intern Pemerintah (APIP) untuk meningkatkan pengawasan dan evaluasi	
(Anggraini dan Akbar 2022)	Pemerintah Kota Yogyakarta	Penelitian ini membahas penerapan SAKIP di Pemerintah Kota Yogyakarta, khususnya mengapa skor akuntabilitas kerjanya belum optimal. Penelitian ini menggunakan pendekatan kualitatif studi kasus untuk menganalisis berbagai aspek pelaksanaan, mulai dari perencanaan hingga evaluasi kinerja. Penelitian ini juga menganalisis faktor internal dan eksternal yang memengaruhi efektivitas implementasi SAKIP	Penelitian menemukan bahwa skor SAKIP belum optimal karena sejumlah faktor, seperti: 1. Tekanan regulasi, 2. Kurangnya kompetensi dan jumlah SDM, 3. Mutasi pegawai yang menyebabkan ketidakteraturan pelaksanaan, 4. Ketersediaan data yang belum update (khususnya dari BPS), 5. Hambatan teknis di pengukuran <i>outcome</i>, 6. Penggunaan indikator kinerja yang belum tepat sasaran, 7. Masih kuatnya orientasi pada aktivitas (<i>effort</i>) ketimbang dampak (<i>effect</i>), 8. Namun juga terdapat upaya perbaikan seperti pelatihan SDM, penguatan pengawasan Inspektorat, dan inovasi sistem terintegrasi antar-OPD.	Penelitian ini relevan karena memperkuat urgensi bahwa terdapat kebutuhan sistem yang mampu menyatukan data perencanaan, pengukuran, dan pelaporan kinerja secara lebih efektif dan <i>real-time</i>. Kelemahan yang ditemui pada penelitian ini yaitu tidak mengembangkan atau menyaranakan penggunaan <i>dashboard</i> visual atau sistem interaktif dalam menyajikan hasil SAKIP. Penyampaian data masih konvensional.
(Yassir <i>et al.</i> 2024)	Kabupaten Polewali Mandar	Penelitian ini membahas pengembangan dan implementasi sistem informasi berbasis <i>web</i> untuk mendukung penyusunan Laporan Akuntabilitas Kinerja Instansi Pemerintah (LAKIP) di Kabupaten Polewali Mandar. Sistem ini dibangun dengan metode <i>Waterfall</i> dan menggunakan alat bantu seperti UML, PostgreSQL, dan Dreamweaver untuk perancangan	Sistem informasi yang dibangun: 1. Memiliki tampilan yang <i>user-friendly</i> dan <i>dashboard</i> interaktif. 2. Memungkinkan input data indikator, target, realisasi anggaran, dan evaluasi capaian SKPD. 3. Diuji menggunakan <i>black box testing</i> dan menunjukkan semua fungsi berjalan dengan baik. 4. Mampu meningkatkan transparansi, integrasi data, efisiensi kerja ASN, dan kualitas evaluasi kinerja.	Penelitian ini relevan karena memberikan landasan praktis, dimana sistem <i>dashboard</i> yang dibangun dapat menunjukkan bahwa <i>dashboard</i> kinerja bisa memperkuat penerapan SAKIP melalui integrasi data dan pengambilan keputusan berbasis bukti. Kelemahan pada penelitian ini yaitu penilaian SAKIP masih secara sederhana belum mengacu pada Permen PANRB Nomor 88 Tahun 2021.

Tabel 3 Penelitian Terkait SAKIP (*lanjutan*)

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Marjan <i>et al.</i> 2021)	Kabupaten Kepulauan Selayar	Penelitian ini mengisi kesenjangan dalam pengembangan E-SAKIP dengan pendekatan berbasis pengalaman pengguna (UX). Sebelumnya, pelaporan kinerja instansi di Kabupaten Kepulauan Selayar masih dilakukan secara manual dan memiliki banyak kendala seperti keterlambatan pengumpulan dan informasi yang tidak jelas. Penelitian ini menggunakan pendekatan <i>Human-Centered Design</i>	Hasil penelitian menunjukkan bahwa pendekatan <i>Human-Centered Design</i> efektif dalam meningkatkan pengalaman pengguna terhadap aplikasi E-SAKIP. <i>Prototype</i> yang dikembangkan telah sesuai dengan kebutuhan calon pengguna dan memiliki tingkat kepuasan yang tinggi, sehingga dapat mengurangi hambatan dalam pelaporan kinerja instansi pemerintah	Penelitian ini relevan karena pendekatan <i>Human-Centered Design</i> yang digunakan dapat membantu dalam mengembangkan aplikasi pemerintahan lain agar lebih <i>user-friendly</i> dan efisien. Kelemahan yang ditemui pada penelitian ini yaitu sistem yang dibangun belum mengacu pada Permen PANRB Nomor 88 Tahun 2021. Penilaian SAKIP masih secara umum (belum disajikan secara rinci)
(Sadewa <i>et al.</i> 2020)	Kabupaten Buleleng	Penelitian ini bertujuan untuk mengevaluasi efektivitas, efisiensi, dan kepuasan pengguna dalam menggunakan aplikasi, serta memberikan rekomendasi perbaikan berdasarkan hasil pengujian. Penelitian ini mengisi kesenjangan dalam evaluasi <i>usability</i> E-SAKIP yang sebelumnya belum dilakukan secara menyeluruh. Beberapa masalah yang ditemukan dalam aplikasi, seperti kurangnya <i>feedback</i> informatif, kesalahan input tanpa peringatan, dan tidak adanya fitur untuk mengoreksi kesalahan, menunjukkan perlunya evaluasi <i>usability</i> yang lebih mendalam	Hasil penelitian menunjukkan bahwa aplikasi E-SAKIP Kabupaten Buleleng cukup efektif, efisien, dan memuaskan pengguna. Namun, masih terdapat beberapa aspek yang perlu diperbaiki, seperti penyederhanaan menu dan tampilan, pencegahan kesalahan input, pemberian <i>feedback</i> yang lebih informatif, serta peningkatan konsistensi dalam penggunaan istilah	Penelitian ini relevan karena pendekatan <i>usability testing</i> yang digunakan dalam penelitian ini dapat dijadikan acuan untuk mengukur efektivitas, efisiensi, dan kepuasan pengguna dalam sistem lain. Kelemahan yang ditemui pada penelitian ini yaitu sistem yang dibangun belum mengacu pada Permen PANRB Nomor 88 Tahun 2021, karena penelitian dilakukan sebelum tahun 2021.
(Saputra <i>et al.</i> 2017)	Dinas Perindustrian Kota Malang	Pengembangan perangkat lunak ini menggunakan metode <i>waterfall</i> .	Hasil penelitian menunjukkan bahwa penerapan sistem informasi berbasis <i>web</i> dapat meningkatkan efisiensi dalam pelaksanaan SAKIP.	Penelitian ini relevan karena pendekatan <i>waterfall</i> dalam pengembangan perangkat lunak dan pengujian sistem dapat menjadi model untuk implementasi sistem serupa di instansi pemerintah.

Tabel 3 Penelitian Terkait SAKIP (*lanjutan*)

Penulis	Instansi/Kasus	Uraian	Temuan Utama	Relevansi dan Kelemahan/Gap
(Saputra <i>et al.</i> 2017)	Dinas Perindustrian Kota Malang	Tujuan penelitian ini adalah untuk meningkatkan efisiensi dalam pencatatan, pengukuran, dan pelaporan kinerja, yang sebelumnya dilakukan secara manual melalui pengembangan sistem informasi berbasis <i>web</i> untuk mengelola pelaksanaan SAKIP di Dinas Perindustrian Kota Malang.	Dengan sistem ini, kepala dinas dan pegawai dapat memantau capaian kinerja secara <i>real-time</i> , serta mengurangi waktu yang diperlukan untuk pengumpulan dan pelaporan data kinerja	Kelemahan yang ditemui pada penelitian ini yaitu sistem yang dibangun belum mengacu pada Permen PANRB Nomor 88 Tahun 2021, karena penelitian dilakukan sebelum tahun 2021.
(Pradana dan Maghfiroh 2024)	Badan Pusat Statistik (BPS)	Penelitian ini membahas upaya peningkatan implementasi Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) melalui pembangunan aplikasi evaluasi AKIP (Akuntabilitas Kinerja Instansi Pemerintah) yang dirancang khusus untuk kebutuhan Badan Pusat Statistik (BPS). Metodologi <i>System Development Life Cycle (SDLC)</i> model <i>iterative waterfall</i>	Penelitian ini berhasil membangun sebuah aplikasi berbasis <i>web</i> : 1. Diuji menggunakan metode <i>black-box testing</i> dan menunjukkan bahwa semua fungsi berjalan sesuai harapan. 2. Dievaluasi melalui <i>Computer System Usability Questionnaire (CSUQ)</i> dengan hasil tingkat kepuasan pengguna sangat tinggi (nilai rata-rata 6,78 dari 7). 3. Mendapatkan hasil Tingkat Kesiapterapan Teknologi (TKT) pada level 8, menunjukkan aplikasi siap digunakan secara operasional. 4. Aplikasi ini memungkinkan digitalisasi proses evaluasi dari pengumpulan data, pengisian dokumen evaluasi (LKE), hingga penyusunan laporan hasil evaluasi (LHE), sehingga lebih cepat, efisien, dan mengurangi ketergantungan pada <i>platform</i> pihak ketiga.	Penelitian ini relevan karena kita bisa mengembangkan lebih lanjut dari sisi tampilan visual, interaktivitas <i>dashboard</i> , integrasi data secara <i>real-time</i> , serta kapabilitas analitik lanjutan untuk membantu pengambilan keputusan dalam manajemen kinerja instansi pemerintah. Sistem yang dibangun sudah mengacu pada Permen PANRB Nomor 88 Tahun 2021, namun belum ada pemanfaatan teknologi kecerdasan buatan dalam proses evaluasi, penyusunan laporan, atau analisis dokumen.

Beberapa hasil penelitian terkait RAG dan *dashboard* dapat dilihat pada Tabel 4.

Tabel 4 Penelitian Terkait RAG dan *Dashboard*

No	Penulis	Uraian
1	(Matheus <i>et al.</i> 2020)	Penelitian ini menyajikan manfaat dari penggunaan <i>dashboard</i> dari sisi strategis dan politis serta teknis. Meskipun <i>dashboard</i> sering digunakan untuk evaluasi kebijakan, <i>dashboard</i> juga dapat mendukung siklus pembuatan kebijakan secara keseluruhan, termasuk formulasi, implementasi, dan evaluasi kebijakan.

Tabel 4 Penelitian Terkait RAG dan *Dashboard* (lanjutan)

No	Penulis	Uraian
2	(Bach <i>et al.</i> 2022)	Penelitian ini membahas pola-pola desain <i>dashboard</i> yang umum digunakan dalam berbagai konteks. Penulis menganalisis 144 <i>dashboard</i> dari berbagai bidang (kesehatan, pemerintahan, bisnis, dsb.) dan mengidentifikasi 42 pola desain, yang dikelompokkan ke dalam dua kategori besar: konten (isi) dan komposisi (tata letak dan interaksi).
3	(Kobi 2024)	Penelitian ini membahas bagaimana <i>dashboard</i> analitik dan alat visualisasi data dapat digunakan untuk meningkatkan manajemen kinerja dan perbaikan proses secara berkelanjutan, khususnya di industri penerbangan..
4	(Rane <i>et al.</i> 2024)	Penelitian ini menyajikan hasil analisis <i>Strengths, Weaknesses, Opportunities, and Threats</i> (SWOT) dari dua model AI yaitu Gemini dan ChatGPT.
5	(Gao <i>et al.</i> 2024)	Penelitian ini membahas bagaimana RAG unggul dalam lingkungan dinamis dengan menawarkan pembaruan pengetahuan secara <i>real-time</i> dan pemanfaatan sumber pengetahuan eksternal yang efektif dengan interpretabilitas tinggi.
6	(Fan <i>et al.</i> 2024)	Penelitian ini memberikan tinjauan komprehensif tentang <i>Retrieval-Augmented Large Language Models</i> (RA-LLMs), yang mengintegrasikan RAG dengan LLM untuk mengatasi masalah halusinasi dan pengetahuan usang. Keunggulannya adalah kerangka analisis sistematis yang mencakup tiga aspek: arsitektur (<i>retrieval, generation, augmentation</i>), strategi pelatihan (<i>training-free, independent, sequential, joint</i>), dan aplikasi (NLP, <i>downstream tasks, domain-specific</i>).
7	(Fanelli <i>et al.</i> 2025)	Penelitian ini membahas pengembangan PerioGPT, yaitu versi khusus dari model GPT-4o yang dioptimalkan untuk bidang periodontologi (kesehatan gigi dan gusi), menggunakan teknik <i>Retrieval-Augmented Generation</i> (RAG). Sistem ini dirancang agar mampu membaca ribuan dokumen PDF, memahami isi klinisnya, dan menjawab pertanyaan secara akurat dan relevan berdasarkan sumber informasi yang diperoleh dari dokumen tersebut.

Berdasarkan uraian di atas maka penelitian ini hadir untuk mengisi kesenjangan melalui:

1. Penilaian mandiri SAKIP secara lengkap menggunakan teknologi informasi secara efektif dan efisien mengacu pada Permen PANRB Nomor 88 Tahun 2021 dan Permen KP Nomor 35 Tahun 2023. Hal ini untuk mengatasi sistem yang ada belum secara rinci mengacu pada regulasi terbaru.
2. Integrasi RAG pada sistem yang dirancang, berperan dalam merangkum dan menganalisis dokumen kinerja sehingga memberikan *insight* yang mendalam secara otomatis. Hal ini untuk mengatasi minimnya pemanfaatan kecerdasan buatan di bidang pemerintahan.
3. *Dashboard analytics* hasil penilaian mandiri SAKIP yang mampu menyajikan data kinerja secara *real-time*, interaktif, dan mudah dipahami. Hal ini untuk mengatasi keterbatasan dalam visualisasi dan digitalisasi sistem.

Sehingga pada akhirnya tujuan awal dari adanya evaluasi dapat tercapai yaitu memastikan bahwa implementasi SAKIP berjalan efektif dalam mencapai tujuan yang telah ditetapkan, yaitu meningkatkan akuntabilitas, efisiensi, dan efektivitas kinerja instansi pemerintah.

III METODE

3.1. Lokasi Penelitian

Kegiatan pengukuran IKU PM SAKIP merupakan agenda rutin tahunan yang dilaksanakan oleh BPPSDMKP. Hasil penilaian selama 4 tahun (2022-2025) tersedia untuk 43 satker. Adapun lokasi yang akan dijadikan objek penelitian yaitu Sekretariat BPPSDMKP dengan alamat Gedung Mina Bahari III Lantai 6, Jalan Medan Merdeka Timur Nomor 16 Jakarta Pusat.

3.2. Kebutuhan Pengembangan

Adapun alat dan bahan yang digunakan, meliputi:

1. Dokumen Peraturan dan Kebijakan, meliputi:
 - a. Peraturan Presiden Nomor 29 Tahun 2014 tentang Sistem Akuntabilitas Kinerja Instansi Pemerintah
 - b. Peraturan Menteri PANRB Nomor 88 Tahun 2021 tentang Evaluasi Akuntabilitas Kinerja Instansi Pemerintah
2. Dokumen Kinerja Organisasi, meliputi:
 - a. Dokumen perencanaan kinerja, meliputi:
 - 1) Keputusan Menteri Kelautan dan Perikanan Nomor 83 Tahun 2025 tentang Rencana Kerja Kementerian Kelautan dan Perikanan Tahun 2026
 - 2) Keputusan Menteri Kelautan dan Perikanan Nomor 35 Tahun 2023 tentang Pengelolaan Kinerja Organisasi di Lingkungan Kementerian Kelautan dan Perikanan Tahun 2026
 - 3) Manual Indikator Kinerja Utama BPPSDMKP Tahun 2026
 - 4) Matriks Cascading BPPSDMKP Tahun 2026
 - 5) Peraturan Kepala BPPSDMKP Nomor 25 Tahun 2025 tentang Rencana Strategis BPPSDMKP Tahun 2025-2029
 - 6) Peraturan Kepala BPPSDMKP Nomor 148 Tahun 2025 tentang Indikator Kinerja Tahun 2026
 - 7) Perjanjian Kinerja BPPSDMKP Tahun 2026
 - 8) Rekap DIPA BPPSDMKP Tahun 2026
 - 9) Rencana Aksi Kinerja BPPSDMKP Tahun 2026
 - 10) Rincian Target Indikator Kinerja Utama BPPSDMKP Tahun 2026
 - 11) Data Rekapitulasi Indikator Kinerja Utama BPPSDMKP Tahun 2026
 - b. Dokumen pengukuran kinerja, meliputi:
 - 1) Hasil Verifikasi Capaian Indikator Kinerja Level II Lingkup BPPSDMKP Triwulan I Tahun 2026
 - 2) Keputusan Kepala BPPSDM Nomor 177 Tahun 2026 tentang Tim Pengelolaan Kinerja di Lingkungan BPPSDMKP Tahun 2026
 - 3) Capaian Nilai Kinerja Organisasi (NKO) Satuan Kerja Lingkup BPPSDMKP Triwulan I Tahun 2026
 - 4) Hasil Pengukuran Kinerja dan Capaian KKP Triwulan I Tahun 2026
 - 5) Standar Operasional Prosedur Pengukuran dan Pelaporan Level 1 BPPSDMKP
 - 6) Standar Operasional Prosedur Pengukuran dan Pelaporan Level 2 Pusat

- 7) Standar Operasional Prosedur Pengukuran dan Pelaporan Sekretariat BPPSDMKP
- 8) Standar Operasional Prosedur Pengukuran dan Pelaporan UPT
- 9) Standar Operasional Prosedur Pengukuran dan Pelaporan, Verifikasi dan Validasi Online BPPSDMKP
- 10) Standar Operasional Prosedur Verifikasi Online BPPSDMKP
- 11) Manual Indikator Kinerja Utama BPPSDMKP Tahun 2026
- c. Dokumen pelaporan kinerja, meliputi:
 - 1) Laporan Kinerja BPPSDMKP Tahun 2025
 - 2) Hasil Reviu Laporan Kinerja BPPSDMKP Tahun 2025
- d. Dokumen evaluasi akuntabilitas kinerja internal, meliputi:
 - 1) Hasil Evaluasi AKIP BPPSDMKP Tahun 2021
 - 2) Hasil Evaluasi AKIP BPPSDMKP Tahun 2022
 - 3) Hasil Evaluasi AKIP BPPSDMKP Tahun 2023
 - 4) Hasil Evaluasi AKIP BPPSDMKP Tahun 2024
 - 5) Hasil Evaluasi AKIP BPPSDMKP Tahun 2025
 - 6) Laporan Pelaksanaan SAKIP BPPSDMKP Tahun 2025
 - 7) Pedoman Evaluasi AKIP Tahun 2025
3. Data hasil penilaian mandiri SAKIP Satker Lingkup BPPSDMKP oleh Eselon I Tahun 2022-2025.
4. Perangkat Pendukung:
 - a. Perangkat keras berupa:
 - 1) Server Ubuntu 22.04, CUDA 12.5, cuDNN 9.1, R 4.3, Python 3.10, a Ryzen 7 5800X CPU, 128 GB RAM, 500 GB storage, and an RTX-3060 12 GB GPU
 - 2) Laptop dengan spesifikasi Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz (4 CPUs), ~2.7GHz, RAM 8 GB, Harddisk 224 GB dengan Sistem Operasi Windows 10 Pro 64-bit
 - b. Perangkat lunak berupa:
 - 1) Bahasa pemrograman Python 3.10.12
 - 2) RStudio Server 2023.09.1
 - 3) Model LLM: models/Gemini 2.5 Pro
 - 4) Model Embedding: models/gemini-embedding-001
 - 5) LangChain Core 1.2.7
 - 6) LangChain Google 4.2.0
 - 7) *Framework Retrieval Augmented Generation Assessment (RAGAS)*

3.3. Tahapan Penelitian

Penelitian ini dilakukan dalam beberapa tahapan sebagai berikut:

1. Pengumpulan Data

Dalam mengumpulkan informasi dan data yang dibutuhkan dalam pembuatan model akuntabilitas kinerja instansi pemerintah ini dilakukan dengan cara yaitu:

a. Studi Literatur

Studi literatur dilakukan dengan cara membaca dan mempelajari buku-buku, artikel, literatur, catatan kuliah, *browsing* di internet dan sumber-sumber lain yang berkaitan dengan proses perancangan model. Hasil studi literatur telah disampaikan pada bab sebelumnya.

b. Observasi Langsung

Observasi atau pengamatan merupakan salah satu teknik pengumpulan data atau fakta yang cukup efektif untuk mempelajari suatu sistem. Observasi adalah pengamatan langsung para pembuat keputusan berikut lingkungan fisiknya dan atau pengamatan langsung suatu kegiatan yang sedang berjalan. Observasi langsung dilakukan dengan cara peninjauan ke tempat objek yang diteliti untuk melihat dan mengumpulkan informasi secara langsung terkait dengan perancangan sistem. Observasi dilakukan dengan mengikuti kegiatan dalam siklus pengelolaan kinerja organisasi.

c. Wawancara

Wawancara dilakukan dengan mengadakan tanya jawab secara langsung kepada para pengelola kinerja. Hal ini dilakukan untuk mengumpulkan data yang ada dan untuk mengetahui kebutuhan awal sistem yang direncanakan berdasarkan kebutuhan. Pertanyaan yang diajukan saat wawancara mengacu pada lembar kerja evaluasi penilaian mandiri SAKIP dan rekonsiliasi kinerja.

d. Analisis Sistem

Analisis sistem adalah suatu proses mengumpulkan dan menginterpretasikan kenyataan-kenyataan yang ada, mendiagnosis persoalan dan menggunakan keduanya untuk memperbaiki sistem. Tahap yang dilakukan antara lain:

- 1) Mengidentifikasi masalah untuk mendapat pengertian sebenarnya dari masalah yang dihadapi yaitu dengan mengidentifikasi penyebab masalah dan bagaimana langkah untuk menyelesaikannya.
- 2) Memahami sistem kerja yang ada dengan mengumpulkan data hasil penelitian dan memahami sistem yang ada.
- 3) Menganalisis kelemahan sistem dan kebutuhan informasi.

Dari hasil tahapan di atas, diperoleh dokumen akuntabilitas kinerja instansi pemerintah sebanyak 31 dokumen dan data hasil penilaian mandiri SAKIP Satker Lingkup BPPSDMKP oleh Eselon I Tahun 2022-2025 dalam bentuk *file sql* yang berasal dari Aplikasi Penilaian Mandiri SAKIP BPPSDMKP.

2. Praproses Data

Prinsip kerja praproses data dan pembangunan model RAG mengacu pada konsep RAG pada umumnya yang diperkenalkan oleh Lewis *et al.* (2021); Gao *et al.* (2024) pada bab sebelumnya, dengan tahapan praproses data sebagai berikut:

a. *Parsing*

Tahap pertama yang dilakukan yaitu mengurai dokumen kinerja instansi pemerintah yang akan menjadi fakta tambahan dalam mendukung kinerja sistem. Pustaka *Python* yang akan digunakan yaitu *pdfplumber*, yang akan digunakan untuk membaca dan mengekstrak dokumen *pdf* yang akan digunakan sebagai fakta tambahan. Tahap ini membuka file PDF dan membacanya lembar demi lembar (*parsing*) untuk kemudian diekstrak bagian teks, tabel dan gambar pada setiap halamannya dengan memilih metode pembacaan dokumen.

b. *Preprocessing*

Setelah data mentah berhasil diambil, program melakukan tahap pembersihan dan *preprocessing*. Untuk tabel yang ditemukan, program

akan menghapus baris kosong, menyambungkan judul kolom yang terpotong, lalu mengubah strukturnya menjadi bentuk narasi atau tabel teks sederhana (*markdown*). Terakhir, program akan menghapus baris-baris kosong berlebih pada teks, menggabungkannya dengan tabel yang sudah dirapikan, lalu menyimpannya sebagai satu hasil bacaan utuh yang dilengkapi dengan informasi nama file dan nomor halaman.

c. *Chunking*

Tahap ini merupakan proses memotong beberapa bagian teks yang telah diurai tadi menggunakan pustaka dari LangChain yaitu *langchain.text_splitter* dan alat yang digunakan yaitu *RecursiveCharacterTextSplitter* yang digunakan untuk memotong dokumen teks yang besar menjadi bagian-bagian yang lebih kecil. Proses ini harus memastikan bahwa teks yang dipotong tidak menghilangkan makna sebenarnya, oleh karena itu kita perlu mengontrol bagaimana pemisahan dilakukan dengan mengatur parameter *chunk_size* (panjang *chunk* maksimum) dan *chunk_overlap* (jumlah karakter yang tumpang tindih antara *chunk* yang berurutan).

d. *Vectorize*

Dalam pencarian teks secara tradisional kita hanya dapat menggunakan kata kunci yang cocok, namun pada *vector database* memungkinkan kita melakukan pencarian berdasarkan makna atau konteks untuk mendapatkan hasil yang relevan meskipun tidak menggunakan kata-kata yang persis. Kita menggunakan *Facebook AI Similarity Search* (FAISS) untuk meningkatkan performa *indexing* dan *retrieval*, terutama jika jumlah dokumen mulai banyak. FAISS jauh lebih cepat dan ringan untuk pencarian vektor lokal, selaras dengan yang disampaikan oleh Douze *et al.* (2025). Untuk menuju tahapan itu, akan digunakan model *embedding* dari *GoogleGenerativeAIEmbeddings* yaitu *gemini-embedding-001*. Hasil *vectorize* ini kemudian disimpan dalam sebuah *vector database* yang dirancang untuk arsitektur RAG dan memiliki dukungan penuh terhadap LangChain. Dari sisi pengguna yang melakukan *query* menggunakan *prompt engineer* terhadap *vector database* dibantu dengan penggunaan pustaka *ChatGoogleGenerativeAI*, dengan terlebih dahulu dilakukan proses *vectorize*. Model *embedding* ini dioptimalkan untuk pencarian semantik, pengambilan dokumen (RAG), dan sistem rekomendasi dengan efisiensi tinggi serta mendukung lebih dari 100 bahasa (Google Cloud Documentation 2026).

3. Pembangunan Model RAG

a. *Augmented*

Pada tahap selanjutnya proses *retrieval* dilakukan dengan mengambil konteks yang paling relevan dan berasal dari *vector database* sesuai dengan *query* dari pengguna. Untuk memperkuat hasil *retrieval* dilakukan tiga tahap yaitu *query transformation*, *retrieval (similarity search)* dan *reranking*. Hasil proses *retrieval* selanjutnya dikombinasikan dengan *prompt* dan *query* yang diberikan oleh pengguna tadi untuk diolah dengan bantuan model LLM. Proses ini disebut dengan augmentasi. Pada tahap ini kuncinya ada di konteks, bukan sekedar memberikan *prompt* tetapi juga fakta tambahan yang kita berikan di awal.

b. *Generate*

Model LLM yang digunakan yaitu *gemini-2.5-pro* dengan menambahkan parameter *temperature* untuk mengurangi adanya halusinasi. Halusinasi pada LLM menghasilkan jawaban yang seperti meyakinkan tapi sebenarnya salah, tidak akurat, atau tidak ada dasarnya. Hal ini wajar karena LLM tidak “tahu” fakta, melainkan memperkirakan kata berikutnya berdasarkan pola bahasa dari data latihnya.

4. Pembangunan *Dashboard* Evaluasi SAKIP

Sumber utama datanya berasal dari data hasil penilaian mandiri SAKIP Satker Lingkup BPPSDMKP oleh Eselon I Tahun 2022-2025 dalam bentuk *file sql* yang diolah untuk ditampilkan dalam bentuk *dashboard* dengan informasi yang ditampilkan meliputi grafik perbandingan hasil penilaian mandiri SAKIP pada satker BPPSDMKP, sebaran predikat hasil nilai dan perbandingan nilai tertinggi sampai terendah untuk setiap satker (per komponen dan per sub-komponen dalam SAKIP). Pembuatan tampilan *dashboard* ini mengacu pada hasil penelitian Bach *et al.* (2022). Sedangkan untuk laporan menampilkan laporan evaluasi SAKIP per satker secara detail berupa skor, penjelasan, dan data dukung setiap pertanyaan SAKIP.

5. Evaluasi Model RAG

Pengujian kualitas model RAG yang dibangun menggunakan *framework* RAGAS. Studi Es *et al.* (2024) memperkenalkan *framework* RAGAS menggunakan tiga skor metrik yaitu *faithfulness*, *answer relevancy*, dan *context relevance*. Untuk melengkapi kebutuhan evaluasi penulis menggunakan lima skor metrik yang digunakan untuk mengevaluasi kinerja RAG menggunakan *framework* RAGAS, antara lain: *faithfulness*, *answer relevancy*, *context precision*, *context recall*, dan *answer relevancy*. Untuk mempermudah tahapan ini maka digunakan pustaka dari *ragas* yaitu *ragas.metrics* dengan tahapan utama yaitu *setup LLM*, *create metric* dan *evaluate*. Evaluasi tersebut juga sekaligus mengukur nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk melengkapi tahapan evaluasi.

6. Evaluasi Manfaat

Tahapan ini merupakan evaluasi terhadap manfaat dari sistem yang dibangun dengan menggunakan metode *System Usability Scale* (SUS) yang telah diperkenalkan oleh Brooke (1996) dan dilakukan pengembangan dalam interpretasi skor SUS oleh Bangor *et al.* (2008).

7. Evaluasi *Dashboard*

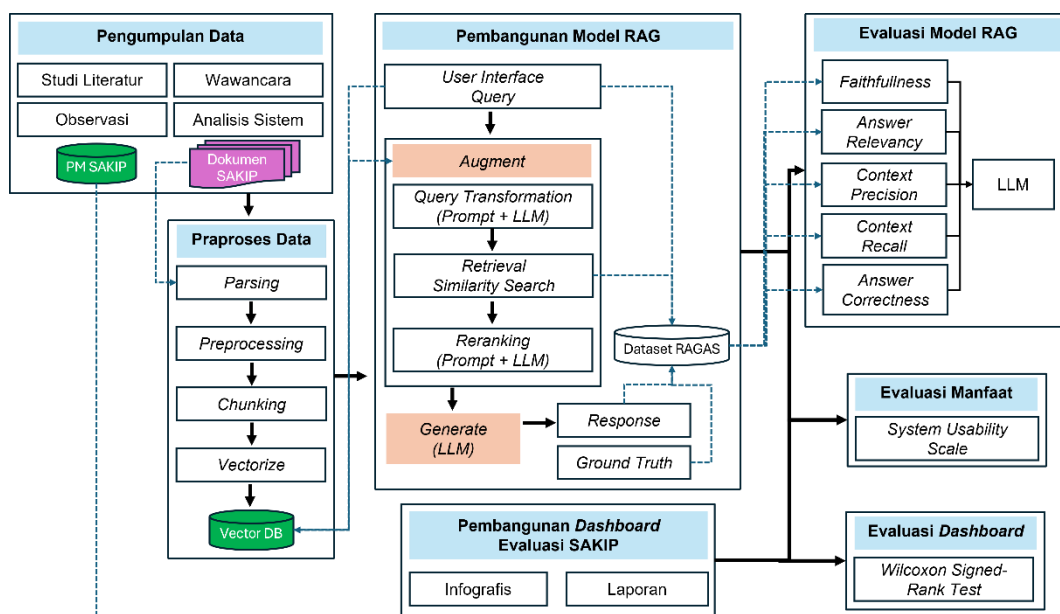
Tahapan ini merupakan evaluasi terhadap tampilan model RAG dan *dashboard* menggunakan metode *Wilcoxon Signed-Rank Test* untuk menilai signifikansi kondisi sebelum dan sesudah tersedia *dashboard* penilaian mandiri SAKIP. Rumus *Wilcoxon Signed-Rank Test* ini berasal dari Wilcoxon (1945) yang kemudian diadopsi dalam penelitian Woolson (2008) dan Altarazi *et al.* (2023) yang menggunakan metode tersebut untuk mengusulkan *framework* untuk mengembangkan dasbor interaktif dan mengevaluasi efektivitas penggunaannya dalam meningkatkan manajemen data, analisis data, visualisasi data, dan berbagi informasi. Sedangkan untuk kuesionernya terdiri dari:

- a. Manfaat *dashboard* perbandingan sebanyak tujuh pertanyaan, selaras dengan studi Yigitbasioglu dan Velcu (2012) yang mengkaji bagaimana

fitur visualisasi (*dashboard*) dalam manajemen kinerja pemerintah/perusahaan dapat meningkatkan kecepatan pemrosesan informasi pengguna dibandingkan membaca dokumen terpisah secara manual.

- b. Pencarian informasi kinerja sebanyak 6 pertanyaan, dengan mengukur kelancaran alur tanya jawab, rasa percaya pengguna pada sistem, serta efisiensi fungsi *query* Borsci *et al.* (2023). Hal ini juga selaras dengan studi Ren *et al.* (2022) tentang indikator efektivitas dan efisiensi sistem *Natural Language Processing* (NLP) dalam menjawab pertanyaan spesifik, selaras dengan metode kerja RAG yang digunakan.
- c. Perbandingan regulasi atau dokumen kinerja sebanyak 6 pertanyaan, yang mengevaluasi dokumen SAKIP, dengan menampilkan metode analisis dokumen kinerja yang dilakukan secara manual dan membandingkannya dengan AI. Selaras dengan studi Zhou dan Hu (2025) yang membandingkan efisiensi waktu pemrosesan 91 artikel oleh manusia dan dua model LLM (ChatGPT-4.0 dan DeepSeek R1). Rata-rata waktu pemrosesan per artikel oleh LLM hanya 33,09 detik, sedangkan manusia membutuhkan 1.582,50 detik. Hal ini menunjukkan peningkatan efisiensi sebesar 47,80 kali lipat.
- d. Manfaat sistem secara umum sebanyak 2 pertanyaan, yang diadopsi dari Bangor *et al.* (2009) sebagai landasan utama dalam memvalidasi keandalan instrumen evaluasi sistem.

Ilustrasi tahapan kegiatan penelitian dapat dilihat pada Gambar 8.



Gambar 8 Tahapan Kegiatan Penelitian

IV HASIL DAN PEMBAHASAN

4.1 Analisis Sistem Evaluasi SAKIP

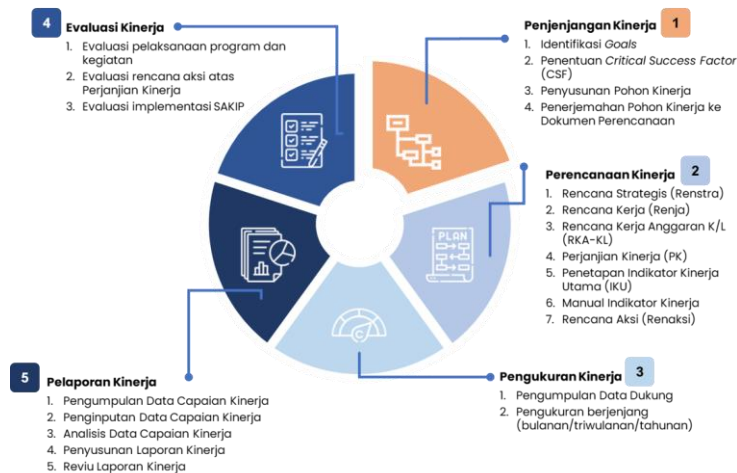
Berdasarkan hasil studi literatur, wawancara, dan observasi yang telah dilakukan selama kurun waktu Juni 2025 s.d Juli 2026, diperoleh informasi terkait hasil analisis sistem evaluasi SAKIP yang saat ini berjalan.

1. Siklus Pengelolaan Kinerja Organisasi

Pengelolaan kinerja organisasi di KKP dilakukan untuk menyelaraskan tujuan dan sasaran setiap level organisasi ke dalam dokumen kinerja organisasi (Indonesia 2023). Level organisasi terdiri atas:

- level 0 untuk Kementerian;
- level I untuk Unit Organisasi Eselon I;
- level II untuk Unit Organisasi Eselon II dan UPT yang bertanggung jawab kepada pimpinan Unit Organisasi Eselon I; dan
- level III untuk UPT yang bertanggung jawab kepada pimpinan Unit Organisasi Eselon II.

Siklus dan *timeline* pengelolaan kinerja organisasi di KKP dapat dilihat pada Gambar 9 dan 10.



Gambar 9 Siklus Pengelolaan Kinerja Organisasi



Gambar 10 Timeline Pengelolaan Kinerja Organisasi

Dalam siklus pengelolaan kinerja terdapat tahapan Evaluasi Kinerja dimana di dalamnya terdapat kegiatan evaluasi implementasi SAKIP (Lihat Bagian 2.4). Pada pelaksanaannya kegiatan pengelolaan kinerja organisasi seluruh dokumen kinerja pada setiap tahapan menjadi sumber data utama pada evaluasi implementasi SAKIP. Proses evaluasi implementasi SAKIP dilaksanakan secara berjenjang sesuai dengan level organisasinya. Dalam proses persiapan penelitian, penulis ikut berpartisipasi langsung melakukan wawancara dan observasi dalam beberapa tahapan pengelolaan kinerja organisasi, meliputi:

a. Penetapan Surat Keputusan Tim Pengelolaan Kinerja

Dalam Surat Keputusan Kepala BPPSDMKP Nomor 177 Tahun 2026 tentang Tim Pengelolaan Kinerja di Lingkungan BPPSDMKP, posisi penulis sebagai Sekretaris yang bertugas membantu Ketua dalam mengoordinasikan perencanaan, pengukuran, pelaporan, evaluasi, dan verifikasi kinerja organisasi di lingkungan BPPSDMKP.

b. Dialog Kinerja Organisasi

Dialog kinerja organisasi merupakan salah satu mekanisme dalam penyusunan perjanjian kinerja. Salah satu tujuannya yaitu melakukan reviu dan penyesuaian target kinerja berdasarkan monitoring dan evaluasi capaian kinerja termasuk kendala, masalah hingga rencana kerja terkait kinerja yang dilakukan secara berjenjang.

c. Penandatanganan Perjanjian Kinerja

Perjanjian kinerja adalah lembar/dokumen yang berisikan penugasan dari pimpinan instansi yang lebih tinggi kepada pimpinan instansi yang lebih rendah untuk melaksanakan program/kegiatan yang disertai dengan indikator kinerja.

d. Pengumpulan Data Capaian Kinerja

Pengumpulan data capaian kinerja dilakukan oleh masing-masing satker dan UPT lingkup BPPSDM, baik untuk capaian level 0 sampai level III tiap triwulannya, berdasarkan target realisasi yang telah dicantumkan pada dokumen perjanjian kinerja yang telah ditandatangani pimpinan satker maupun UPT. Setiap triwulan, sebanyak sembilan satker telah dikumpulkan data capaian kinerjanya oleh penulis.

e. Pengukuran Capaian Kinerja Triwulan

Pengukuran kinerja dilakukan dengan membandingkan antara realisasi kinerja dengan target kinerja secara berjenjang mulai dari level III sampai level 0, dengan berpedoman pada manual indikator kinerja. Pengukuran kinerja dilakukan tiap triwulan dengan memasukkan data realisasi indikator kinerja dan dokumen pendukungnya pada sistem Aplikasi Pengelolaan Kinerja. Sedangkan untuk pengukuran kinerja *mandatory*, dilakukan oleh penanggung jawab indikator kinerja dan disampaikan kepada pimpinan unit organisasi penerima *mandatory* melalui naskah dinas paling lambat tanggal 10 bulan berikutnya sesuai dengan periode pelaporan indikator kinerja (Indonesia 2023). Setiap triwulan, sebanyak sembilan satker telah diukur capaian kinerjanya oleh penulis.

f. Penginputan Capaian Kinerja

Penginputan capaian kinerja dilakukan juga setiap Triwulan pada aplikasi Kinerjaku, sekaligus memasukkan data dukung capaian dimaksud pada

aplikasi dimaksud berdasarkan target realisasi yang telah dicantumkan pada dokumen perjanjian kinerja yang telah ditandatangani pimpinan satker maupun UPT.

g. Validasi dan Verifikasi Capaian Kinerja Triwulan

Validasi dan verifikasi data dukung capaian kinerja dilakukan berdasarkan data realisasi Indikator Kinerja dan dokumen pendukung. Verifikasi dilakukan oleh Sekretariat Jenderal untuk data capaian kinerja level I, dan Pimpinan level I untuk data capaian kinerja level II dan level III. Verifikasi dilakukan dalam jangka waktu paling lama tanggal 19 terhitung sejak triwulan berakhir. Verifikasi akan menghasilkan kertas kerja hasil verifikasi yang menyatakan capaian kinerja valid atau capaian kinerja tidak valid (Indonesia 2023). Setiap triwulan, sebanyak sembilan satker telah divalidasi dan diverifikasi data capaian kerjanya oleh penulis.

h. Penyusunan Laporan Kinerja Triwulan

Pelaporan kinerja dilakukan secara berjenjang dari level III sampai level 0, pelaporan kinerja terdiri dari Laporan Kinerja Interim, merupakan laporan kinerja setiap triwulan; dan Laporan Kinerja tahunan, merupakan laporan kinerja dalam jangka waktu satu tahun. Bentuk dan format Laporan Kinerja tercantum dalam lampiran dan ditandatangani oleh pimpinan unit organisasi (Indonesia 2023). Penulis ikut membantu mengisi narasi akuntabilitas keuangan pada Laporan Kinerja Triwulan BPPSDMKP dan Sekretariat BPPSDMKP.

i. Evaluasi Rencana Aksi

Evaluasi rencana aksi atas Perjanjian Kinerja dilakukan setiap triwulan:

1) level I oleh:

- a) Kepala Biro Perencanaan untuk Sekretariat Jenderal; dan
- b) Sekretaris Direktorat Jenderal/Inspektorat Jenderal/Badan untuk Direktorat Jenderal/Inspektorat Jenderal/Badan, dalam jangka waktu paling lama minggu ketiga terhitung sejak triwulan berakhir;

2) level II oleh pimpinan Unit Organisasi Eselon I, dalam jangka waktu paling lama minggu kedua terhitung sejak triwulan berakhir; dan

3) level III oleh pimpinan Unit Organisasi Eselon II, dalam jangka waktu paling lama minggu kedua terhitung sejak triwulan berakhir.

Penulis ikut membantu mengisi narasi evaluasi rencana aksi atas capaian keuangan pada BPPSDMKP dan Sekretariat BPPSDMKP.

j. Penyampaian Laporan Kinerja

Laporan Interim disampaikan kepada Menteri untuk level I dalam jangka waktu paling lama empat minggu terhitung sejak triwulan berakhir, dan atasan langsung untuk level II dan level III dalam jangka waktu paling lama tiga minggu terhitung sejak triwulan berakhir. Sementara itu Laporan kinerja tahunan untuk level II dan level III disampaikan kepada atasan dalam jangka waktu paling lama tiga minggu terhitung sejak tahun berakhir. Laporan Kinerja Triwulan 1, 2, 3 baik untuk level I sudah disampaikan sesuai prosedur dan dibuktikan dengan adanya nota dinas yang disampaikan ke Biro Perencanaan KKP. Sementara Laporan Kinerja Triwulan 1, 2 dan 3 Sekretariat BPPSDM juga sudah disampaikan dan dibuktikan dengan adanya nota dinas penyampaian Laporan Kinerja dari



Sekretaris BPPSDMKP ke Kepala BPPSDMKP. Penulis ikut membantu mengisi narasi akuntabilitas keuangan pada Laporan Kinerja Triwulan BPPSDMKP dan Sekretariat BPPSDMKP.

k. **Reviu Laporan Kinerja Tahun 2025**

Reviu Laporan Kinerja Tahun 2025 level I dilakukan oleh Inspektorat Jenderal, sementara Laporan Kinerja Tahun 2025 level II dilakukan oleh Sekretariat BPPSDMKP dan Laporan Kinerja Tahun 2025 level III dilakukan oleh level II, dalam hal ini Sekretariat BPPSDMKP bersama Satker Pusat yang melingkupinya. Hasil reviu berupa kertas kerja disampaikan kepada pimpinan level I, II dan III sebagai bahan penyempurnaan Laporan Kinerja. sebanyak tujuh satker telah direviu laporan kinerjanya oleh penulis.

2. **Evaluasi Implementasi SAKIP**

Evaluasi implementasi SAKIP dilakukan melalui:

a. **Rekonsiliasi Kinerja**

Rekonsiliasi kinerja dilakukan terhadap level II dan level III oleh pimpinan level I menggunakan kertas kerja rekonsiliasi kinerja. Kegiatan ini bertujuan untuk meningkatkan kualitas evaluasi akuntabilitas agar lebih menggambarkan tingkat akuntabilitas unit kerja yang dievaluasi dan dapat dimanfaatkan sebagai umpan balik (*feedback*) perbaikan perencanaan kinerja, penerapan manajemen kinerja, dan peningkatan capaian kinerja unit kerja secara berkelanjutan. Disamping itu, hasil reviu ini melengkapi hasil evaluasi penilaian mandiri SAKIP yang dilakukan Inspektorat Jenderal (*Self Assessment*) dengan harapan KKP memiliki gambaran kualitas implementasi SAKIP secara menyeluruh di seluruh Unit Kerja. Sebanyak tujuh satker telah dilakukan rekonsiliasi capaian kinerjanya oleh penulis.

b. **Penilaian Mandiri SAKIP**

Penilaian mandiri SAKIP dilakukan oleh: Inspektur Jenderal untuk level 0 dan level I; dan pimpinan Unit Organisasi Eselon I untuk level II dan level III. Penilaian Mandiri SAKIP dilakukan berdasarkan pedoman evaluasi akuntabilitas kinerja instansi pemerintah. Adapun tahapan dalam penilaian mandiri SAKIP meliputi:

- 1) **Penyiapan Surat Keputusan Tim Pengelolaan Kinerja**
- 2) ***Kick Off Meeting* Penilaian Mandiri SAKIP**
- 3) **Surat Tugas Tim Pendamping dan Verifikasi Penilaian Mandiri SAKIP**
- 4) **Pendampingan dan Verifikasi Penilaian Mandiri SAKIP.**
Teknik evaluasi meliputi *checklist* pengumpulan data dan informasi, komunikasi melalui tanya jawab sederhana, observasi, dan studi dokumentasi. Mekanisme selengkapnya telah dijelaskan pada sub bab 2.4.
- 5) **Penyusunan Berita Acara Hasil Penilaian Mandiri SAKIP Level II dan III**
- 6) **Penyampaian Hasil Penilaian Mandiri SAKIP Level II dan III.**

Dalam proses penilaian mandiri SAKIP di Level 0 oleh Kemen PAN RB, BPPSDMKP mendukung pengisian Aplikasi eSR Kemen PANRB. Selain itu, BPPSDMKP juga menggunakan Aplikasi Penilaian Mandiri SAKIP

berbasis *website* yang dikelola secara mandiri (terpisah dari aplikasi lainnya) untuk memudahkan pengisian instrumen penilaian. Sebanyak 11 satker telah dilakukan pendampingan dan verifikasi penilaian mandiri SAKIP oleh penulis.

Berdasarkan studi literatur dan observasi terhadap praktik tata kelola pemerintahan saat ini, masalah utama yang dihadapi adalah belum optimalnya proses evaluasi SAKIP dalam mendorong efisiensi dan akuntabilitas. Dari diagnosis yang dilakukan, ditemukan tiga penyebab utama, antara lain:

1. Ketidaksesuaian regulasi, dimana instrumen penilaian yang digunakan pada sejumlah instansi seringkali tertinggal dan belum dipetakan secara presisi ke dalam sistem informasi sesuai dengan mandat dalam Permen PANRB Nomor 88 Tahun 2021.
2. Kelebihan beban kognitif, dimana evaluator harus membaca dan merangkum ratusan halaman dokumen kinerja secara manual, sehingga proses penggalian wawasan (*insight*) berjalan lambat dan rentan terhadap bias atau *human error*.
3. Kebutaan data, dimana hasil evaluasi seringkali berakhir sebagai laporan statis. Pimpinan instansi tidak memiliki visibilitas *real-time* terhadap kinerja organisasi karena keterbatasan instrumen visualisasi.

Sistem monitoring dan evaluasi konvensional yang berjalan saat ini masih bersifat mekanis dan terfragmentasi. Dalam alur kerja saat ini, instansi pemerintah mengumpulkan dokumen kinerja ke dalam Aplikasi e-SAKIP Reviu (eSR Kemen PANRB). Evaluator kemudian mengunduh dokumen tersebut, membacanya satu per satu, dan melakukan skoring menggunakan lembar kerja yang bersifat tertutup. Sistem ini mengumpulkan data, tetapi seringkali gagal dalam memahami data tersebut.

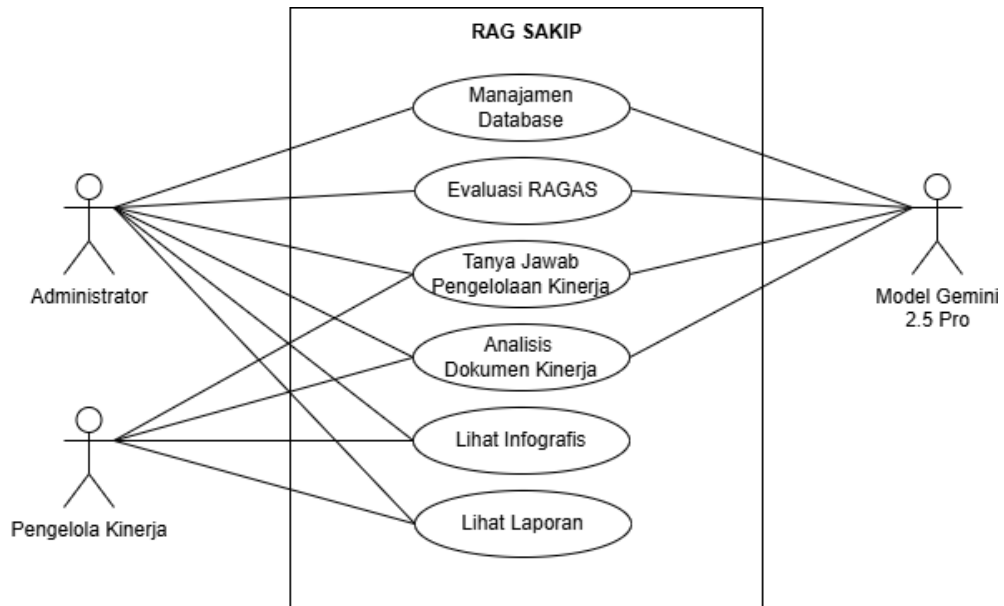
4.2 Use Case Diagram

Berdasarkan penjelasan analisis sistem evaluasi SAKIP, terdapat beberapa kebutuhan yang mendasari *use case diagram* ini dibuat, misalnya untuk *use case* Tanya Jawab Pengelolaan Kinerja digunakan mengakomodir kebutuhan penilaian mandiri SAKIP sedangkan *use case* Analisis Dokumen Kinerja digunakan untuk mengakomodir kegiatan Rekonsiliasi Kinerja yang mendukung peningkatan kualitas evaluasi akuntabilitas agar lebih menggambarkan tingkat akuntabilitas unit kerja yang dievaluasi. Kedua *use case* tersebut diharapkan dapat mengurangi beban kognitif para evaluator.

Use case Lihat Infografis digunakan untuk mengakomodir kebutuhan perbandingan hasil penilaian mandiri SAKIP antar satker, antar komponen dan sub-komponen sehingga dapat memberikan gambaran secara utuh terhadap kondisi pengelolaan kinerja di BPPSDMKP.

Use case Lihat Laporan digunakan untuk mengakomodir kebutuhan informasi tentang hasil penilaian mandiri SAKIP pada satker tertentu secara rinci meliputi nilai yang diberikan, penjelasan atas nilai dan data dukung terhadap setiap pertanyaan, kriteria, sub-komponen, komponen dan rekomendasi serta tindak lanjut dalam penilaian mandiri SAKIP.

Diagram *use case* pada Gambar 11 mendefinisikan interaksi antara aktor (pengguna) dengan fungsionalitas sistem.



Gambar 11 Use Case Diagram

Terdapat tiga aktor utama yang berinteraksi dengan sistem ini, antara lain:

1. *User* (Pengelola Kinerja): Pengguna akhir aplikasi yang bertugas memantau dan mengelola kinerja instansi. Mereka memiliki akses ke fitur-fitur analitik dan informasi.
2. Admin (Administrator): Pengelola sistem yang memiliki hak akses penuh, termasuk mengelola basis pengetahuan (*database*) dan mengevaluasi performa sistem AI.
3. AI (Model Gemini 2.5 Pro): Aktor sistem (berupa *Large Language Model*) yang bertugas memproses data, memberikan jawaban cerdas, menganalisis dokumen, dan mengevaluasi metrik.

Adapun penjelasan setiap *use case* sebagai berikut:

1. UC1: *Manajemen Database*
Aktor yang terlibat adalah Admin. Admin mengunggah dan mengelola dokumen kinerja pemerintah ke dalam *vector database*. Sistem akan berinteraksi dengan AI untuk "*Membuat Knowledge*", yaitu mengubah teks dokumen menjadi format vektor (*embedding*) agar bisa dipahami dan dicari oleh AI dengan cepat. Selain itu Admin juga mengelola data hasil penilaian mandiri SAKIP.
2. UC2: *Lihat Laporan*
Aktor yang terlibat adalah Admin dan *User*. Pengguna dan admin dapat membuka dan melihat laporan kinerja yang sudah disusun atau dihasilkan oleh sistem.
3. UC3: *Tanya Jawab Pengelolaan Kinerja*
Aktor yang terlibat adalah Admin dan *User*. Ini merupakan fitur *chatbot* atau kolom pencarian pintar. Pengguna bertanya terkait aturan atau data kinerja. Sistem akan meneruskan pertanyaan ini ke AI untuk "*Memberikan Jawaban*" berdasarkan dokumen SAKIP yang ada di *vector database*.
4. UC4: *Analisis Dokumen Kinerja*

Aktor yang terlibat adalah Admin dan *User*. Ini merupakan fitur untuk mengevaluasi dokumen kinerja (misalnya Rencana Strategis, Penetapan Kinerja, Laporan Kinerja, Hasil Evaluasi AKIP atau dokumen kinerja lainnya). Sistem berinteraksi dengan AI untuk "Membandingkan Konteks", seperti mengecek apakah dokumen tersebut sudah sesuai dengan standar yang berlaku atau membandingkan antar dokumen.

5. UC5: Lihat Infografis

Aktor yang terlibat adalah Admin dan *User*. Pengguna dan admin dapat melihat ringkasan hasil penilaian mandiri SAKIP dalam bentuk visual (grafik dan diagram) agar lebih mudah dipahami.

6. UC6: Evaluasi Model

Aktor yang terlibat adalah Admin. Evaluasi model adalah fitur khusus Admin untuk memantau seberapa akurat jawaban yang diberikan oleh AI. Sistem akan menggunakan AI untuk "Menghitung Hasil" metrik evaluasi (seperti *faithfulness*, *answer relevancy*, dan *context precision*, *context recall*, dan *answer relevancy* pada model RAG).

4.3 Activity Diagram

Berdasarkan *Use Case Diagram* di atas, telah dirancang juga *Activity Diagram* yang memetakan alur aktivitas untuk masing-masing aktor (Pengelola Kinerja dan Administrator), serta bagaimana sistem berinteraksi dengan model AI di latar belakang. Berikut adalah representasi visual *Activity Diagram* dapat dilihat pada Lampiran 7. Adapun penjelasan *Activity Diagram* adalah sebagai berikut:

1. Titik Awal (*Start*): Proses selalu dimulai saat pengguna (baik Pengelola maupun Admin) masuk ke dalam sistem melalui proses Login.
2. Pencabangan Hak Akses (*Decision*): Sistem mengevaluasi *role* pengguna.
 - a. Jika yang masuk adalah Pengelola Kinerja, sistem hanya membuka empat cabang aktivitas (Lihat Laporan, Lihat Infografis, Tanya Jawab, Analisis Dokumen).
 - b. Jika yang masuk adalah Administrator, sistem membuka seluruh aktivitas Pengelola Kinerja ditambah dua aktivitas khusus pengelola sistem (Manajemen *Database* dan Evaluasi Model).
3. Interaksi Langsung dengan AI (*Backend Processing*):
 - a. Saat masuk ke Tanya Jawab, sistem meneruskan kueri ke AI untuk mengeksekusi aksi "Memberikan Jawaban".
 - b. Saat masuk ke Analisis Dokumen, pengguna memilih dokumen, lalu AI melakukan tugas komputasi untuk "Membandingkan Konteks".
 - c. Pada menu Admin *Manajemen Database*, dokumen yang diunggah diproses oleh AI untuk Membuat *Knowledge* (menjadi *vector embeddings*).
 - d. Pada menu Admin Evaluasi Model, AI bekerja di latar belakang untuk "Menghitung Hasil" metrik (seperti akurasi pencarian atau kualitas jawaban dari model RAG).
4. Titik Akhir (*End*): Setelah sistem menyajikan hasil (menampilkan laporan, infografis, jawaban AI, hasil analisis, pembaruan basis data, atau metrik evaluasi), alur aktivitas tersebut selesai.

4.4 Sequence Diagram

Diagram ini menunjukkan interaksi antar komponen sistem dalam urutan waktu tertentu, selengkapnya dapat dilihat pada Lampiran 8. Diagram urutan (*Sequence Diagram*) di atas dibagi menjadi empat tahapan utama, meliputi:

1. Tahap 1: Manajemen *Database* (Langkah 1 - 6)
Ini adalah fase persiapan data atau *ingestion* yang dilakukan oleh Administrator.
 - a. Langkah 1: Admin mengunggah dokumen kinerja (seperti Rencana Strategis, Penetapan Kinerja, Laporan Kinerja, Hasil Evaluasi AKIP atau dokumen kinerja organisasi lainnya) ke dalam Aplikasi SAKIP.
 - b. Langkah 2-3: Aplikasi mengirimkan teks dokumen tersebut ke Model Gemini 2.5 Pro untuk diubah menjadi *Embedding* (representasi angka/vektor dari teks). Inilah proses AI "Membuat *Knowledge*".
 - c. Langkah 4-6: Aplikasi menyimpan nilai vektor tersebut ke dalam *vector database*. Setelah tersimpan, Admin menerima notifikasi bahwa dokumen berhasil dijadikan basis pengetahuan.
2. Tahap 2: Tanya Jawab dan Analisis Dokumen (Langkah 7 - 14)
Ini adalah fase *retrieval and generation* yang merupakan inti dari penggunaan RAG oleh *User*.
 - a. Langkah 7: *User* mengetik pertanyaan di *chatbot* (UC3) atau mengunggah draf laporan untuk dievaluasi kesesuaiannya (UC4).
 - b. Langkah 8-9: Sama seperti di Tahap 1, pertanyaan *user* dikirim ke AI untuk diubah menjadi bentuk vektor agar bisa dibandingkan.
 - c. Langkah 10-11: Aplikasi melakukan pencarian ke *vector database* untuk mencari dokumen yang makna dan konteksnya paling mirip dengan pertanyaan *User* (*similarity search*). *Vector database* mengembalikan potongan dokumen yang relevan.
 - d. Langkah 12: Aplikasi menggabungkan pertanyaan *user* dan potongan dokumen asli, lalu mengirimkannya sebagai *prompt* (perintah) akhir ke Gemini 2.5 Pro.
 - e. Langkah 13-14: AI menganalisis konteks tersebut untuk "Memberikan Jawaban" yang akurat, tidak mengarang, dan merujuk pada dokumen yang tepat. Hasilnya kemudian ditampilkan di layar *User*.
3. Tahap 3: Lihat Laporan dan Infografis (Langkah 15 - 17)
Ini adalah fase visualisasi dan rekapitulasi.
 - a. Langkah 15: *User* atau Admin mengakses menu laporan (UC2) atau *dashboard* infografis (UC5).
 - b. Langkah 16-17: Aplikasi memproses kumpulan data kinerja yang telah disimpan (baik hasil olahan AI maupun data tabular) dan menampilkannya dalam format visual yang mudah dipahami. Proses ini terjadi di internal sistem aplikasi tanpa memanggil AI generasi.
4. Tahap 4: Evaluasi Model AI (Langkah 18 - 22)
Ini adalah fase monitoring dan pemeliharaan (*maintenance*) untuk memastikan sistem pintar tidak mengalami penurunan kualitas.
 - a. Langkah 18: Admin dapat menjalankan menu Evaluasi Model (UC6) secara berkala.
 - b. Langkah 19-20: Aplikasi mengambil *log* riwayat tanya-jawab pengguna, lalu memintakan ke AI untuk "Menghitung Hasil".

- c. Langkah 21-22: AI mengukur metrik otomatis (misalnya mengukur apakah jawaban yang diberikan kemarin sesuai dengan dokumen dasar). Hasil skor RAGAS dikirim ke aplikasi untuk ditampilkan.

4.5 Arsitektur RAG

Berikut adalah desain arsitektur sistem RAG untuk SAKIP, yang dirancang selaras dengan *Use Case Diagram*, *Activity Diagram*, dan *Sequence Diagram* yang telah dibahas sebelumnya. Arsitektur ini menggunakan Gemini 2.5 Pro sebagai inti pemrosesan bahasa, baik untuk ekstraksi metadata maupun penalaran akhir, seperti terlihat pada Lampiran 9.

Diagram dibagi menjadi empat node server utama yang disusun secara logis dari kiri ke kanan (alir data operasional) dan atas ke bawah (layanan luar dan evaluasi):

1. **Server Aplikasi Operasional (*Operational App Server*)**: Node kuning di sebelah kiri merupakan "pusat otak" sistem dan bertindak sebagai orkestrator yang mengatur alur ingesti data, kompilasi prompt dan antar muka pengguna. Di dalamnya terdapat artefak perangkat lunak yang menampung komponen-komponen *pipeline* ingesti (*loader*, *chunker*, *embedder*), *retrieval controller* sebagai pencari konteks, *generation controller* sebagai pembuat jawaban, dan layanan ekspor dataset. Hal ini sejalan dengan studi Gao *et al.* (2024) yang membahas arsitektur standar implementasi RAG, di mana diperlukan sebuah sistem terpusat (aplikasi) yang mengatur alur *chunking* dokumen, pemanggilan model *embedding*, dan penggabungan konteks sebelum dikirim ke LLM.
2. **Server Basis Data Vektor (*Vector Database Server*)**: Node kuning di tengah adalah node penyimpanan khusus yang menyimpan representasi vektor dari dokumen dan menjalankan komputasi mesin pencarian kesamaan (*similarity search*), serta melayani ingesti data dan pencarian konteks. Hal ini sejalan dengan studi Lewis *et al.* (2021) yang menekankan pentingnya pemisahan indeks dokumen berbasis vektor (seperti FAISS atau basis data vektor modern) yang bertindak sebagai *retriever* non-parametrik yang terpisah dari generator utama.
3. **Layanan LLM Eksternal (*External LLM Service*)**: Node abu-abu di bagian atas mewakili layanan luar, yaitu model Gemini 2.5 Pro, berada di luar infrastruktur lokal dan diakses melalui pemanggilan jaringan (*API Call*) oleh klien LLM di Server Aplikasi. Hal ini sejalan dengan studi Wei *et al.* (2023) yang menggunakan layanan eksternal memungkinkan aplikasi berfokus pada teknik rekayasa prompt (seperti *Chain-of-Thought*) tanpa harus membebani server lokal dengan komputasi model bahasa yang sangat besar. Node terpisah ini memvalidasi konsep *Language-Model-as-a-Service* (LMaaS).
4. **Server Evaluasi Kinerja (*Performance Evaluation Server*)**: Node kuning di sebelah kanan ini adalah lingkungan terpisah yang didedikasikan untuk proses evaluasi dengan menghitung skor metrik AI tanpa mengganggu performa server operasional utama. Node ini menerima file hasil ekspor (*eval_dataset.json*), menjalankan kerangka RAGAS, dan menampilkan *dashboard* metrik. Hal ini sejalan dengan studi Es *et al.* (2024) yang menyarankan untuk memisahkan evaluasi ke dalam *node* tersendiri dalam

praktik rekayasa data untuk memastikan bahwa komputasi perhitungan metrik evaluasi berjalan independen dari alur interaksi *user* secara *real-time*.

4.6 Hasil Praproses Data

Hasil dari tahapan praproses data dapat dijelaskan sebagai berikut:

1. Parsing

Pustaka yang akan digunakan yaitu *pdfplumber*, yang digunakan untuk membaca dan mengekstrak dokumen *pdf* yang akan digunakan sebagai fakta tambahan dengan presisi tinggi. Secara sederhana, proses dimulai dengan membuka file PDF dan membacanya lembar demi lembar (*parsing*). Pada setiap halamannya, program akan memilih salah satu dari dua metode pembacaan. Jika menggunakan strategi "*AI Vision*", program akan mengubah halaman tersebut menjadi gambar dan meminta bantuan AI untuk mendeteksi seluruh teksnya. Namun, jika menggunakan strategi biasa, program akan langsung mengekstrak teks asli dan memisahkan tabel-tabel yang ditemukannya pada halaman tersebut. Salah satu masalah ketidakakuratan pembacaan tabel pada dokumen *pdf* adalah tantangan terbesar dalam RAG. Oleh karena itu, solusi terbaik adalah memberikan pilihan metode untuk mengurai dokumen akuntabilitas kinerja instansi pemerintah yang akan menjadi fakta tambahan dalam mendukung kinerja sistem, antara lain:

- Lattice*, untuk jenis tabel bergaris hitam tebal yang tegas.
- Stream*, untuk tabel yang hanya mengandalkan spasi.
- Semantic*, untuk merubah tabel menjadi kalimat narasi.

Cara kerjanya, program akan membaca tabel tersebut baris demi baris, lalu memasangkan setiap judul kolom dengan isi selnya menggunakan kata sambung "adalah" (contoh: "Nama adalah Budi", "Usia adalah 25"). Program ini cukup pintar untuk mengabaikan sel yang kosong atau berisi teks "N/A", kemudian ia akan menggabungkan semua informasi pada baris tersebut dengan tanda koma dan mengakhirinya dengan titik, sehingga barisan data yang kaku berubah menjadi kalimat-kalimat penjelasan utuh.

- AI Vision*, untuk mengurai halaman penuh dengan gambar yang memiliki resolusi tinggi.

Secara sederhana, prosesnya dimulai dengan menyiapkan AI (*ChatGoogleGenerativeAI*) agar menjawab seakurat mungkin tanpa menebak-nebak (nilai *temperature* diatur sebesar 0,0), lalu gambar yang diterima akan disesuaikan warnanya ke format standar dan diubah menjadi kode teks khusus (*Base64*) agar bisa dikirim lewat internet. Setelah gambar siap, program mengirimkan gambar tersebut bersama sebuah instruksi tegas kepada AI untuk bertindak sebagai ahli pembaca data: menyalin seluruh teks, angka, maupun tabel ke dalam catatan yang rapi (*Markdown*), dan memintanya menjawab "KOSONG" jika gambar tersebut tidak berisi apa-apa. Terakhir, program akan menerima teks hasil bacaan AI tersebut untuk digunakan lebih lanjut, atau mengembalikan pesan *error* jika terjadi kendala saat proses pengiriman.

2. Preprocessing

Setelah data mentah berhasil diambil, program melakukan tahap pembersihan dan *preprocessing*. Untuk tabel yang ditemukan, program akan menghapus baris kosong, menyambungkan judul kolom yang terpotong, lalu mengubah

strukturnya menjadi bentuk narasi atau tabel teks sederhana (*markdown*). Terakhir, program akan menghapus baris-baris kosong berlebih pada teks, menggabungkannya dengan tabel yang sudah dirapikan, lalu menyimpannya sebagai satu hasil bacaan utuh yang dilengkapi dengan informasi nama file dan nomor halaman. *Pseudocode* untuk penjelasan di atas dapat dilihat pada Lampiran 10.

Dokumen kinerja yang digunakan sebagai sumber data utama dapat dilihat pada Tabel 5.

Tabel 5 Dokumen Kinerja

No	Komponen SAKIP	Jumlah File (pdf) *	Ukuran File
1	Perencanaan Kinerja	11	27,5 MB
2	Pengukuran Kinerja	11	7,92 MB
3	Pelaporan Kinerja	2	59,6 MB
4	Evaluasi Kinerja	7	3,41 MB

Keterangan: Rincian nama file bisa dilihat pada Bagian 3.2.2

3. *Chunking*

Tantangan yang dihadapi pada tahap ini yaitu jika kita memotong teks dengan potongan yang terlalu kecil maka konteks akan hilang, sedangkan jika potongannya terlalu besar maka banyak informasi yang tidak perlu yang akan diambil. Alat yang akan digunakan yaitu *RecursiveCharacterTextSplitter* dari pustaka *langchain_text_splitters*, yang dirancang untuk memecah teks panjang menjadi potongan-potongan kecil (*chunks*) yang lebih terkelola, sambil berusaha menjaga struktur semantik (makna) teks tersebut tetap utuh. Untuk parameter *chunk size* (panjang *chunk* maksimum) diatur sebesar 2.000, sementara itu parameter *chunk overlap* (jumlah karakter yang tumpang tindih antara *chunk* yang berurutan) diatur sebesar 300. Hasil proses *chunking* dapat dilihat pada Tabel 6.

Tabel 6 Hasil *Chunking*

No	Komponen SAKIP	Jumlah File (pdf)*	Panjang Text yang Diekstrak	Jumlah <i>Chunk</i>	Rata-rata Panjang <i>Chunk</i>
1	Perencanaan Kinerja	11	15.091.539	8.518	1.772
2	Pengukuran Kinerja	11	142.241	100	1.422
3	Pelaporan Kinerja	2	884.437	519	1.704
4	Evaluasi Kinerja	7	134.855	105	1.284

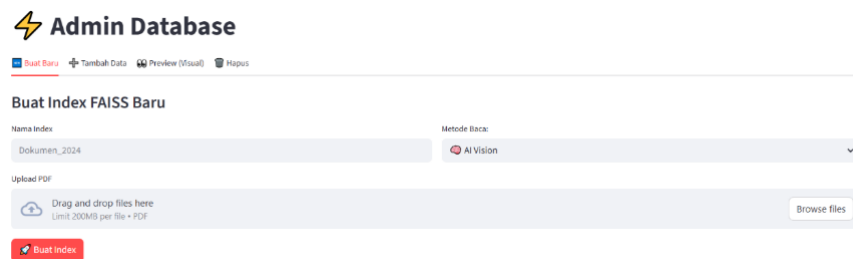
Keterangan: Rincian nama *file* bisa dilihat pada Bagian 3.2.2

4. *Vectorize*

Untuk mendukung tahapan ini, kita menggunakan *Facebook AI Similarity Search* (FAISS) untuk meningkatkan performa *indexing* dan *retrieval*,

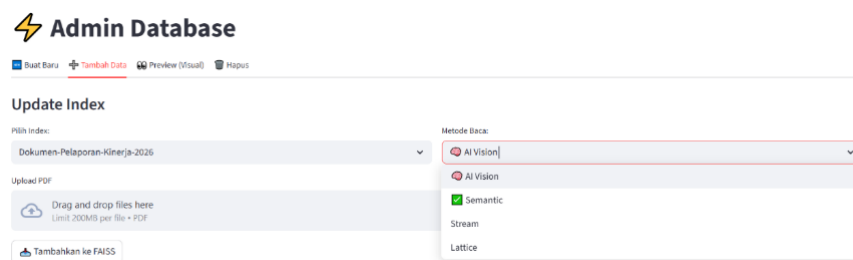
terutama jika jumlah dokumen mulai banyak. FAISS jauh lebih cepat dan ringan untuk pencarian vektor lokal. Dalam pencarian teks secara tradisional kita hanya dapat menggunakan kata kunci yang cocok, namun pada *vector database* memungkinkan kita melakukan pencarian berdasarkan makna atau konteks untuk mendapatkan hasil yang relevan meskipun tidak menggunakan kata-kata yang persis. Untuk menuju tahapan itu, akan digunakan model *embedding* dari *GoogleGenerativeAIEmbeddings* yaitu *gemini-embedding-001*, model penyematan teks canggih dari Google yang dirancang untuk memetakan teks ke dalam representasi vektor berdimensi tinggi (3.072 dimensi). Model ini dioptimalkan untuk pencarian semantik, pengambilan dokumen (RAG), dan sistem rekomendasi dengan efisiensi tinggi serta mendukung lebih dari 100 bahasa (Google Cloud Documentation 2026). Terdapat empat fitur dalam mengelola *index* yang akan digunakan, antara lain:

a. Buat Baru, fitur pembuatan *index* baru, seperti terlihat pada Gambar 12.



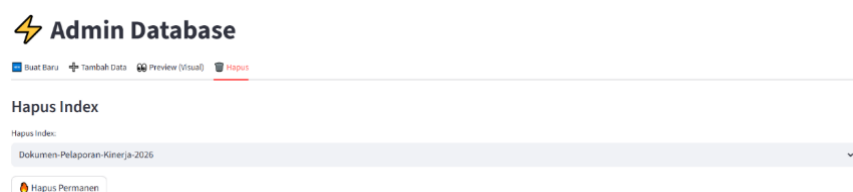
Gambar 12 Tampilan Menu Buat Baru

b. Tambah Data, fitur untuk menambahkan *index* yang tersedia, seperti terlihat pada Gambar 13.



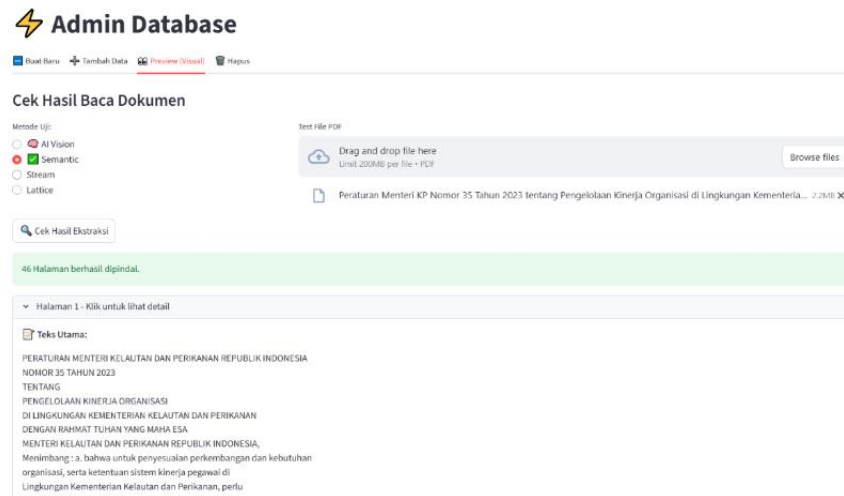
Gambar 13 Tampilan Menu Tambah Data

c. Hapus, fitur untuk menghapus *index* yang tersedia, seperti terlihat pada Gambar 14.



Gambar 14 Tampilan Menu Hapus Data

- d. *Preview (Visual)*, fitur untuk melihat hasil pembuatan *index* sebelum dilakukan penyimpanan, seperti terlihat pada Gambar 15.



Gambar 15 Tampilan Menu *Preview*

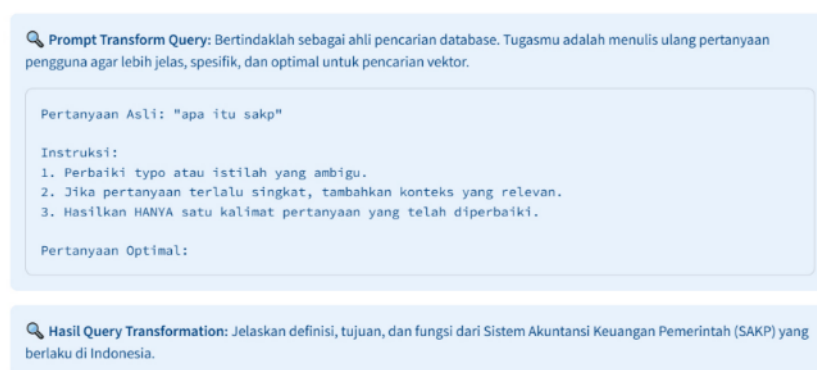
4.7 Hasil Pembangunan RAG

Tahapan Pembangunan RAG merupakan kelanjutan dari tahapan sebelumnya yang dapat dijelaskan sebagai berikut:

1. *Augmented*

Hasil augmentasi ini sangat penting, karena nantinya akan digunakan sebagai bahan evaluasi. Oleh karena itu, terdapat tiga bagian dalam tahapan ini, antara lain:

- a. *Query Transformation*, yaitu menulis ulang pertanyaan pengguna agar lebih optimal untuk pencarian semantik (*vector search*). Tahap ini menggunakan model Gemini 2.5 Pro dengan pengaturan *temperature* sebesar 0,3 dan *prompt template*, seperti terlihat pada Gambar 16.



Gambar 16 Contoh *Prompt Template* untuk *Query Transformation*

- b. *Retrieval (High Recall)*, yaitu mengambil lebih banyak dokumen dengan menggunakan *similarity_search*. Sebanyak 15 dokumen berhasil diperoleh, seperti terlihat pada Gambar 17.

Hasil Retrieval: [Document(id='00f6ee28-ebba-4861-87b6-7051735fe792', metadata={'source': 'Peraturan Menteri KP Nomor 35 Tahun 2023 tentang Pengelolaan Kinerja Organisasi di Lingkungan Kementerian Kelautan dan Perikanan.pdf', 'page': 2}, page_content='Dala adalah 11., m Peraturan Menteri ini adalah Sistem Akuntabilitas K, yang dimaksud adalah inerja Instans, d dengan: adalah i Pemerintah yang.\nm Peraturan Menteri ini adalah selanjutnya disingkat S, yang dimaksud adalah AKIP adalah r, d dengan: adalah angkaian sistematik.\nm Peraturan Menteri ini adalah dari berbagai aktivitas, a, yang dimaksud adalah lat, dan prose, d dengan: adalah dur yang dirancang.\nm Peraturan Menteri ini adalah untuk tujuan penetapan, yang dimaksud adalah n dan penguk, d dengan: adalah uran, pengumpulan.\nm Peraturan Menteri ini adalah data, pengklasifikasian., yang dimaksud adalah pengikhtisar, d dengan: adalah an, dan pelaporan.\nm Peraturan Menteri ini adalah Kinerja pada instans, yang dimaksud adalah i pemerinta, d dengan: adalah h, dalam rangka.\nm Peraturan Menteri ini adalah pertanggungjawaban d, yang dimaksud adalah an peningkata, d dengan: adalah n kinerja instansi.\nm Peraturan Menteri ini adalah pemerintah..\nDala adalah 12., m

Gambar 17 Contoh Hasil Retrieval

- c. *Reranking (High Precision)*, yaitu memfilter 15 dokumen menjadi 5 dokumen terbaik menggunakan model Gemini 2.5 Pro dengan pengaturan *temperature* 0. Tujuannya adalah untuk mengevaluasi kembali relevansi dokumen yang diperoleh. Adapun *prompt template* yang disusun, seperti terlihat pada Gambar 18.

Prompt Rerank Document: Kamu adalah Asisten Penilai Relevansi (Reranker). Tugasmu adalah memilih maksimal 5 dokumen yang PALING RELEVAN untuk menjawab pertanyaan pengguna.

PERTANYAAN: "apa itu sakp"

KANDIDAT DOKUMEN:

[ID_0]: Dala adalah 11., m Peraturan Menteri ini adalah Sistem Akuntabilitas K, yang dimaksud adalah

[ID_1]: melaksanakan program/kegiatan yang disertai dengan Indikator Kinerja. 10. Target Kinerja adalah hasil dan satuan hasil yang direncanakan akan dicapai dari setiap Indikator Kinerja. 11. Sistem Akuntabilitas Kinerja Instansi Pemerintah yang

[ID_14]: Nilai PM SAKIP adalah aspek penilaian, dihitung adalah di dala, berdasarkan Permen PAN d adalah m evaluasi atas implementas, an RB Nomor 8 adalah i SAKIP, yakni: Nilai PM SAKIP adalah 1. perencanaan, dihitung adalah kinerja. Nilai PM SAKIP adalah 2. pengukuran, dihitung adalah kinerja. Nilai PM SAKIP adalah 3. pelaporan k, dihitung adalah inerja. Nilai PM SAKIP adalah 4. evaluasi kin, dihitung adalah erja. Nilai PM SAKIP adalah Penilaian man, dihitung adalah diri SAK, berdasarkan Permen PAN d adalah IP dilaksanakan oleh Tim S, an RB Nomor 8 adalah AKIP Sekretaris, 8 Tahun adalah iat, dan. Nilai...

INSTRUKSI:

1. Evaluasi setiap dokumen: Apakah dokumen ini berisi jawaban untuk pertanyaan di atas?
2. Abaikan dokumen yang tidak relevan atau hanya berisi noise.
3. Kembalikan HANYA daftar ID dokumen terpilih dalam format JSON Array integer. Urutkan dari yang p

Contoh Output: [0, 4, 1]

OUTPUT JSON:

Hasil Reranking: [Document(id='4fe4f0ce-f6da-4615-80f0-7fe9fe825230', metadata={'source': 'Peraturan Menteri KP Nomor 35 Tahun 2023 tentang Pengelolaan Kinerja Organisasi di Lingkungan Kementerian Kelautan dan Perikanan.pdf', 'page': 2}, page_content='melaksanakan program/kegiatan yang disertai dengan\nIndikator Kinerja.\n10. Target Kinerja adalah hasil dan satuan hasil yang\nndirencanakan akan dicapai dari setiap Indikator Kinerja.\n11. Sistem Akuntabilitas Kinerja Instansi Pemerintah yang\nselanjutnya disingkat SAKIP adalah rangkaian sistematik\ndari berbagai aktivitas, alat, dan prosedur yang dirancang\nuntuk tujuan penetapan dan pengukuran, pengumpulan\ndata, pengklasifikasian, pengikhtisaran, dan pelaporan\nKinerja pada instansi pemerintah, dalam rangka\npertanggungjawaban dan peningkatan kinerja

Gambar 18 Contoh Prompt Template untuk Reranking Dokumen

2. Generate

Pada tahapan ini dilakukan penyiapan konteks akhir yang telah diverifikasi pada tahapan sebelumnya, dengan menambahkan aturan khusus dan pertanyaan dari pengguna pada *prompt* yang dibuat. LLM yang digunakan adalah model Gemini 2.5 Pro, dengan penambahan parameter *temperature* sebesar 0,1 untuk mengurangi halusinasi. Halusinasi pada LLM menghasilkan jawaban yang tampak meyakinkan namun sebenarnya keliru,

tidak akurat, atau tidak berdasar. Hal ini wajar karena LLM tidak "mengetahui" fakta, melainkan memprediksi kata berikutnya berdasarkan pola bahasa dari data pelatihannya. Adapun *prompt template* yang disusun, seperti terlihat pada Gambar 19.

Prompt Akhir: Peran: Analis Data Profesional. Instruksi: Jawab pertanyaan pengguna berdasarkan KONTEKS yang telah diverifikasi berikut.

Aturan Khusus:

1. Jika ada tabel markdown atau narasi angka, baca dengan teliti.
2. Jangan mengarang angka sendiri.
3. Jika data tidak ada dalam konteks, katakan "Tidak ada informasi dalam dokumen yang tersedia".
4. Gunakan Bahasa Indonesia yang formal.

KONTEKS TERVERIFIKASI:
melaksanakan program/kegiatan yang disertai dengan

Indikator Kinerja. 10. Target Kinerja adalah hasil dan satuan hasil yang direncanakan akan dicapai dari setiap Indikator Kinerja.
11. Sistem Akuntabilitas Kinerja Instansi Pemerintah yang selanjutnya disingkat SAKIP adalah rangkaian sistematis dari berbagai aktivitas, alat, dan prosedur yang dirancang untuk tujuan penetapan dan pengukuran, pengumpulan data, pengklasifikasian, pengikhtisaran, dan pelaporan Kinerja pada instansi pemerintah, dalam rangka pertanggungjawaban dan

Gambar 19 Contoh *Prompt Template* Akhir

Pseudocode untuk alur kerja *ingestion*, *retrieval*, *reranking*, dan *generation* dapat dilihat pada Lampiran 11-12. Respon yang dihasilkan kemudian diberikan evaluasi secara manual oleh pengguna melalui pengisian *ground truth*, untuk kemudian dievaluasi kinerjanya menggunakan *framework* RAGAS. Evaluasi *ground truth* tersebut dapat dilakukan atau dilewati tanpa disimpan untuk menjadi sebuah *dataset*, seperti terlihat pada Gambar 20.

Verifikasi Data untuk Evaluasi

Pertanyaan: bagaimana hasil evaluasi AKIP setiap tahun

Jawaban AI: Berdasarkan konteks yang tersedia, berikut adalah rincian hasil evaluasi AKIP (Sistem Akuntabilitas Kinerja Instansi Pemerintah) per tahun:

- Tahun 2020:
 - Nilai Hasil Evaluasi: 87,93
 - Predikat Penilaian: A (Memuaskan)
- Tahun 2021:
 - Nilai Hasil Evaluasi: 86,65
 - Predikat Penilaian: A (Memuaskan)
 - Catatan: Terdapat penurunan nilai dibandingkan tahun 2020. Hasil evaluasi ini menghasilkan 3 rekomendasi dari Inspektorat Jenderal (Itjen).
- Tahun 2022:
 - Tidak ada informasi mengenai nilai akhir dan predikat SAKIP dalam dokumen yang tersedia.
 - Namun, terdapat data bahwa nilai efisiensi kinerja output program/Level 1 pada tahun 2022 adalah 82,23%, menurun dibandingkan tahun 2021 yang sebesar 92,08%.
- Tahun 2023:
 - Tidak ada informasi mengenai nilai akhir dan predikat SAKIP dalam dokumen yang tersedia. Dokumen hanya menjelaskan beberapa pencapaian dan kelemahan yang perlu menjadi perhatian.

Jawaban Seharusnya (Ground Truth):

Tulis jawaban ideal di sini agar AI bisa dinilai...

☒ Simpan ke Dataset

Gambar 20 Contoh Tampilan Pengisian *Ground Truth*

Pengujian sederhana dilakukan dengan menghitung waktu respons ke dalam dua jenis yaitu *knowledge ingestion* dan *AI interaction*. Dokumen yang digunakan berkaitan dengan pelaporan kinerja (lihat bagian 3.2.b.3). Total waktu yang diperlukan untuk menyiapkan pengetahuan adalah 20,6 detik (Tabel 7), sedangkan waktu untuk percakapan adalah 47,6 detik (Tabel 8), dengan rincian sebagai berikut.

Tabel 7 Waktu Respon *Knowledge Ingestion*

No	Tahapan	Waktu Mulai	Waktu Selesai	Latensi (Detik)
1	<i>PDF Extraction</i>	08:40:58	08:41:10	11.8800
2	<i>Chunking</i>	08:41:10	08:41:10	0.0000
3	<i>Embedding</i>	08:41:10	08:41:18	7.7000
	Total	08:40:58	08:41:19	20.6000

Tabel 8 Waktu Respon Interaksi AI

No	Tahapan	Waktu Mulai	Waktu Selesai	Latensi (Detik)
1	<i>Query Transformation</i>	09:02:41	09:02:56	15.2524
2	<i>Embedding</i>	09:02:56	09:02:56	0.4756
3	<i>Retrieval (Top-15)</i>	09:02:56	09:02:56	0.0009
4	<i>LLM Reranking (Top-5)</i>	09:02:56	09:03:12	15.2084
5	<i>Generation</i>	09:03:12	09:03:28	16.6979
	Total	09:02:41	09:03:28	47.6374

4.8 Hasil Pembangunan *Dashboard* Evaluasi SAKIP

Dashboard Evaluasi SAKIP dirancang untuk mengatasi keterbatasan instrumen visualisasi dari hasil penilaian mandiri SAKIP satker lingkup BPPSDMKP, dengan contoh tampilan seperti terlihat pada Gambar 21, Gambar 22, dan Gambar 23.



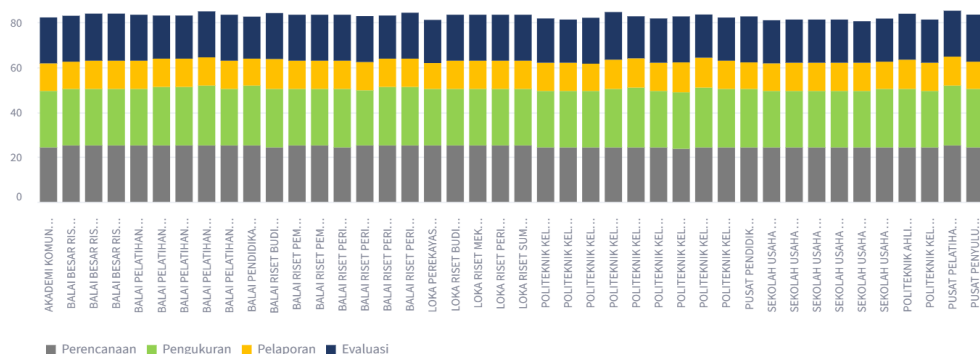
Infografis & Rekapitulasi SAKIP

Pilih Tahun Anggaran

2025

Cetak PDF

Grafik Penilaian Mandiri SAKIP Tahun 2025



Gambar 21 Tampilan Grafik Perbandingan Hasil Penilaian Mandiri SAKIP

Tabel Matriks Penilaian SAKIP (Rinci & Berobot)										
Komponen / Satker	Bobot	403837	403829	427511	403875	403830	440013	403817	403832	2392
TOTAL NILAI AKIP	100.00	80.45 A	80.45 A	80.45 A	79.55 BB	79.55 BB	79.55 BB	78.90 BB	78.00 BB	77.4 BB
A. Perencanaan Kinerja	30.00	25.20	24.60	25.20	24.60	24.60	24.60	25.20	25.20	25.2
a. Dokumen Perencanaan kinerja telah tersedia	6.00	AA	A	AA	A	A	A	AA	AA	AA
b. Dokumen Perencanaan kinerja telah memenuhi standar yang baik	9.00	BB	BB	BB	BB	BB	BB	BB	BB	BB
c. Perencanaan Kinerja telah dimanfaatkan untuk mewujudkan hasil yang berkesinambungan	15.00	BB	BB	BB	BB	BB	BB	BB	BB	BB
B. Pengukuran Kinerja	30.00	23.70	23.70	23.70	24.60	23.10	23.10	21.30	23.70	23.1
a. Pengukuran Kinerja telah dilakukan	6.00	AA	AA	AA	A	A	A	AA	AA	A
b. Pengukuran Kinerja telah menjadi kebutuhan dalam mewujudkan Kinerja	9.00	BB	BB	BB	BB	BB	BB	B	BB	BB
c. Pengukuran Kinerja telah dijadikan dasar dalam pemberian Reward dan Punishment	15.00	B	B	B	BB	B	B	CC	B	B
C. Pelaporan Kinerja	15.00	12.30	12.15	12.30	11.10	11.85	11.85	12.15	11.10	11.1
a. Terdapat Dokumen Laporan yang menggambarkan Kinerja	3.00	A	AA	A	A	A	A	AA	A	BB
b. Dokumen Laporan Kinerja telah memenuhi Standar	4.50	BB	B	BB	B	B	B	B	B	CC
c. Pelaporan Kinerja memberikan dampak dalam penyesuaian strategi	7.50	BB	BB	BB	B	BB	BB	BB	B	BB
D. Evaluasi Akuntabilitas Kinerja Internal	25.00	19.25	20.00	19.25	19.25	20.00	20.00	20.25	19.00	18.0
a. Evaluasi Akuntabilitas Kinerja Internal telah dilaksanakan	5.00	BB	BB	BB	BB	BB	BB	AA	A	BB
b. Evaluasi dilaksanakan secara berkualitas dengan Sumber Daya memadai	7.50	B	BB	B	B	BB	BB	B	BB	B
c. Implementasi SAKIP meningkat karena evaluasi	12.50	BB	BB	BB	BB	BB	BB	BB	CC	B

Gambar 22 Contoh Tampilan Perbandingan Nilai SAKIP Satker Berdasarkan Komponen dan Sub-Komponen SAKIP.

Rekapitulasi dan Pemingkatan Nilai AKIP					
No	Kode	Satuan Kerja	Nilai AKIP	Kategori	Keterangan
1	403837	POLITEKNIK KELAUTAN DAN PERIKANAN KARAWANG	80.45	A	Memuaskan
2	403829	BALAI RISET PERIKANAN BUDIDAYA AIR TAWAR DAN PENYULUHAN PERIKANAN	80.45	A	Memuaskan
3	427511	POLITEKNIK AHLI USAHA PERIKANAN (POLTEK AUP) JAKARTA	80.45	A	Memuaskan
4	403875	POLITEKNIK KELAUTAN DAN PERIKANAN DUMAI (POLTEK KP DUMAI)	79.55	BB	Sangat Baik
5	403830	BALAI RISET BUDIDAYA IKAN HIAS	79.55	BB	Sangat Baik

Gambar 23 Contoh Tampilan Perbandingan Nilai Satker (Tertinggi - Terendah)

Sedangkan untuk laporan menampilkan laporan evaluasi SAKIP per satker secara detail (skor, penjelasan, dan data dukung setiap pertanyaan SAKIP), seperti terlihat pada Gambar 24 dan Gambar 25.



Laporan Evaluasi SAKIP

Cetak PDF

TAHUN ANGGARAN 2025	SATUAN KERJA 403832 - BALAI RISET PEMULIAAN IKAN	TOTAL NILAI 83.50	PREDIKAT A (Memuaskan)
------------------------	---	----------------------	---------------------------

Rekomendasi dan Tindak Lanjut						
No	Komponen	Nilai	Hasil Evaluasi	Rekomendasi	Tindak Lanjut	Bukti
1	Perencanaan Kinerja	25.20	Apresiasi atas ketersediaan dokumen perencanaan kinerja selama 5 tahun terakhir	Menjadikan budaya kinerja yang baik dan diimplementasikan dalam perencanaan kinerja secara berkala	-	-
2	Pengukuran Kinerja	25.20	Apresiasi atas ketersediaan dokumen pengukuran kinerja selama 5 tahun terakhir	Menjadikan budaya kinerja yang baik dalam pengukuran kinerja	-	-

Gambar 24 Contoh Tampilan Rekomendasi Hasil Penilaian Mandiri SAKIP

A. Perencanaan Kinerja

Skor: 25.20

B. Pengukuran Kinerja

Skor: 25.20

C. Pelaporan Kinerja

Skor: 12.60

D. Evaluasi Akuntabilitas Kinerja Internal

Skor: 20.50

No		Indikator Penilaian	Bobot	Skor	Capaian	Nilai Sub	Nilai Komp	Penjelasan	Bukti
1	a	Evaluasi Akuntabilitas Kinerja Internal telah dilaksanakan	5.00	300.00	5.00 (100.00%)	90.00 A	4.50	Kualitas seluruh kriteria telah terpenuhi dan telah dipertahankan dalam setidaknya 2 tahun terakhir	
	1	Terdapat pedoman teknis Evaluasi Akuntabilitas Kinerja Internal		100.00				Terdapat pedoman teknis Evaluasi Akuntabilitas Kinerja Internal	
	2	Evaluasi dilaksanakan pada seluruh unit kerja level II		100.00				Evaluasi Akuntabilitas Kinerja Internal telah dilaksanakan pada seluruh unit kerja level II	

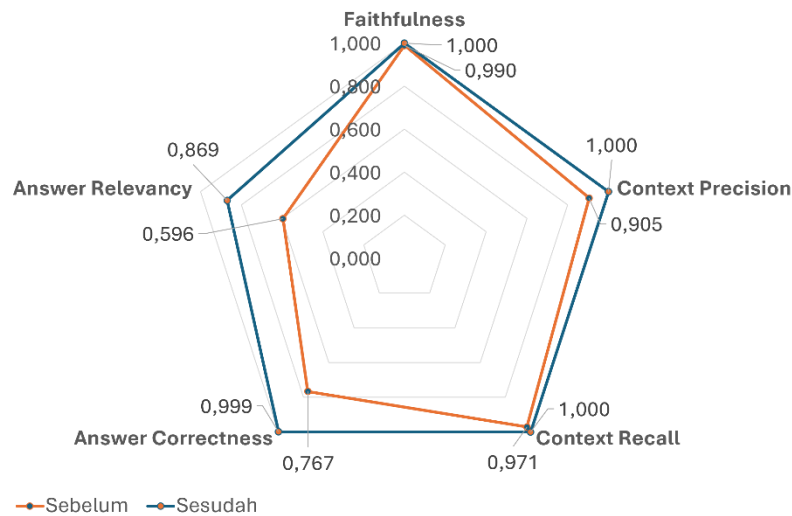
Gambar 25 Contoh Tampilan Laporan Hasil Penilaian Mandiri SAKIP

4.9 Evaluasi Model RAG

Studi Es *et al.* (2024) memperkenalkan *framework* RAGAS menggunakan tiga skor metrik yaitu *faithfulness*, *answer relevancy*, dan *context relevance*. Untuk melengkapi kebutuhan evaluasi penulis menggunakan lima skor metrik yang digunakan untuk mengevaluasi kinerja RAG menggunakan *framework* RAGAS, antara lain:

1. *Faithfulness*, yaitu mengukur sejauh mana jawaban yang dihasilkan didukung oleh konteks yang diberikan. Jawaban dipecah menjadi beberapa pernyataan (klaim). Setiap klaim diperiksa apakah ada buktinya di dalam dokumen yang di-retrieve.
2. *Answer Relevancy*, yaitu menilai apakah jawaban yang diberikan relevan dengan pertanyaan pengguna. RAGAS meminta LLM untuk membuat beberapa pertanyaan berdasarkan jawaban yang dihasilkan, lalu menghitung *cosine similarity* antara pertanyaan buatan tersebut dengan pertanyaan asli dari pengguna.
3. *Context Precision*, yaitu mengevaluasi relevansi dari konteks yang berhasil diambil (*retrieved*). Fokusnya pada kualitas pengurutan (*ranking*) dokumen.
4. *Context Recall*, mengukur kelengkapan konteks yang diambil, yaitu apakah semua informasi yang diperlukan untuk menjawab pertanyaan berhasil ditemukan.
5. *Answer Correctness*, membandingkan jawaban sistem dengan *ground truth* untuk menilai kebenaran faktual. Nilai *Answer Correctness* adalah rata-rata tertimbang (*weighted average*) dari kesamaan faktual dan kesamaan semantik antara jawaban sistem dengan jawaban referensi (*ground truth*).

Berikut adalah hasil perhitungan RAGAS dengan *pseudocode* dapat dilihat pada Lampiran 13.



Gambar 26 Hasil perhitungan model RAG (sebelum dan sesudah perbaikan)

Berdasarkan Gambar 26 di atas, pada kondisi sebelum perbaikan, sistem RAG menunjukkan performa yang sangat baik pada komponen *retrieval* (penemuan konteks) dan *faithfulness* (kesetiaan pada konteks). Namun, terdapat kelemahan signifikan pada aspek *answer relevancy* (relevansi jawaban). Hal ini mengindikasikan bahwa meskipun sistem mampu menemukan konteks yang tepat, jawaban yang dihasilkan seringkali tidak menjawab pertanyaan pengguna secara langsung, yang pada akhirnya berdampak pada skor *answer correctness* (kebenaran jawaban) yang masih moderat.

Fokus perbaikan adalah menjembatani kesenjangan antara konteks yang sudah relevan dengan jawaban akhir yang harus relevan pula. Selain perbaikan saat menginput *ground truth*, perbaikan lainnya yang dilakukan yaitu menerapkan modul *re-ranking* yang bertujuan untuk menempatkan potongan konteks yang paling relevan dengan pertanyaan di urutan paling atas, sehingga LLM akan memprioritaskannya saat menyusun jawaban dan tidak terdistraksi oleh informasi relevan lainnya.

Kondisi sesudah perbaikan, sistem RAG menunjukkan performa yang sangat baik, terutama pada komponen *retrieval* dan akurasi jawaban. Skor *context precision*, *context recall*, *faithfulness*, dan *answer correctness* yang mendekati sempurna mengindikasikan bahwa sistem mampu menemukan konteks yang relevan dan menghasilkan jawaban yang akurat. Namun, terdapat ruang perbaikan yang signifikan pada metrik *answer relevancy*, yang menunjukkan bahwa jawaban yang dihasilkan terkadang kurang fokus atau tidak sepenuhnya menjawab inti pertanyaan pengguna, meskipun informasinya benar.

Berikut adalah analisis mendalam untuk setiap metrik evaluasi:

1. *Faithfulness*

Metrik ini mengukur sejauh mana jawaban yang dihasilkan didukung sepenuhnya oleh konteks yang diberikan. Skor sebelum perbaikan sebesar 0,990, menunjukkan bahwa model generator hampir tidak pernah menghasilkan informasi di luar konteks yang diberikan. Ini adalah fondasi yang kuat untuk sistem yang dapat dipercaya. Dengan peningkatan skor

sesudah perbaikan sebesar 1,000 adalah hasil sempurna, yang berarti sistem sama sekali tidak menghasilkan informasi di luar konteks.

2. *Context Precision*

Metrik ini mengevaluasi rasio sinyal terhadap derau (*signal-to-noise ratio*) dalam konteks yang diambil. Skor sebelum perbaikan sebesar 0,905, menunjukkan bahwa komponen *retrieval* berhasil menemukan hampir semua informasi yang relevan dari basis data untuk menjawab pertanyaan. Sistem tidak melewatkan potongan informasi krusial. Dengan peningkatan skor sesudah perbaikan sebesar 1,000 menunjukkan bahwa hampir semua informasi yang diambil oleh *retriever* relevan dan digunakan untuk membentuk jawaban, dengan kata lain komponen *retriever* bekerja dengan sangat efisien.

3. *Context Recall*

Metrik ini mengukur kemampuan sistem untuk mengambil semua bagian konteks yang relevan dan diperlukan untuk menjawab pertanyaan. Skor sebelum perbaikan sebesar 0,971 menandakan bahwa sebagian besar konteks yang diambil oleh *retriever* memang relevan dengan pertanyaan. Tingkat "noise" atau informasi tidak relevan yang ikut terambil relatif rendah. Dengan peningkatan skor setelah perbaikan sebesar 1,000 menandakan tidak ada informasi penting yang terlewatkan oleh *retriever*.

4. *Answer Correctness*

Metrik ini menilai kebenaran faktual dari jawaban yang dihasilkan, berdasarkan konteks yang ada. Skor sebelum perbaikan sebesar 0,767 berada di level moderat dan perlu ditingkatkan. Rendahnya skor ini bukan disebabkan oleh halusinasi (karena *faithfulness* tinggi), melainkan karena jawaban yang tidak relevan dianggap sebagai jawaban yang salah dari perspektif pengguna. Dengan peningkatan skor setelah perbaikan sebesar 0,999 menegaskan bahwa jawaban yang diberikan akurat.

5. *Answer Relevancy*

Metrik ini mengukur seberapa relevan jawaban yang dihasilkan terhadap pertanyaan asli pengguna. Sebelum perbaikan, diperoleh skor sebesar 0,596 dan menjadi area masalah utama. Artinya, meskipun konteks yang diambil sudah benar dan jawaban yang dihasilkan faktual sesuai konteks, jawaban tersebut seringkali tidak fokus atau tidak menjawab inti dari pertanyaan pengguna. Setelah perbaikan diperoleh skor sebesar 0,869 yang mengindikasikan bahwa respons sistem kini jauh lebih fokus, padat, dan langsung menjawab esensi pertanyaan pengguna.

Tabel 9 Dataset Evaluasi

Komponen SAKIP	Jumlah Pertanyaan	Jenis Pertanyaan	Sumber Dokumen*	Panjang Jawaban (Kata)
Perencanaan Kinerja	10	Faktual, Instruksional	2.a: 1), 3), 4), 6), 9), 10)	35-209
Pengukuran Kinerja	10	Faktual, Instruksional	2.a.3), 2.b: 1), 2), 3), 4), 5), 8), 9)	41-178

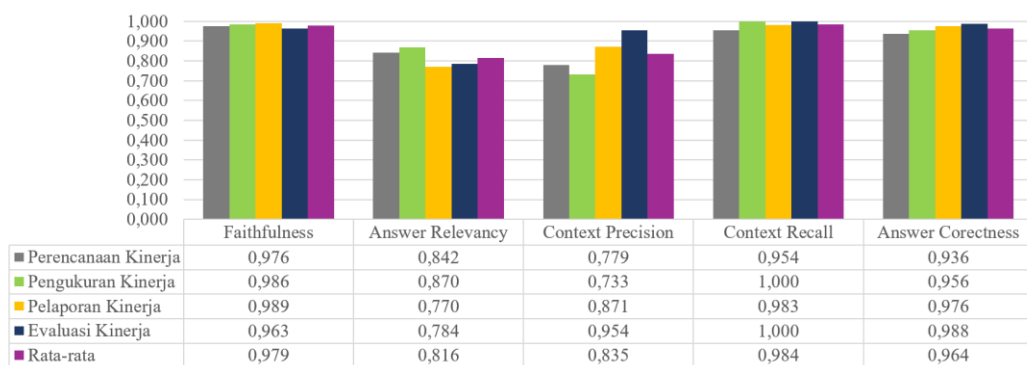
Tabel 9 Dataset Evaluasi (lanjutan)

Komponen SAKIP	Jumlah Pertanyaan	Jenis Pertanyaan	Sumber Dokumen*	Panjang Jawaban (Kata)
Pelaporan Kinerja	10	Deskriptif, Faktual, Biner	2.c: 1), 2)	39-325
Evaluasi Kinerja	10	Konseptual, Faktual, Biner	2.d: 1), 2)	46-186

* Keterangan: Sumber dokumen dapat dilihat pada Bagian 3.2

Hasil evaluasi RAGAS pada sistem sebelum perbaikan menggunakan sepuluh *dataset*, sedangkan pada sistem setelah perbaikan menggunakan tiga *dataset*. Hal ini tentu menimbulkan bias, sehingga evaluasi RAGAS dilakukan kembali pada empat komponen SAKIP dengan masing-masing dataset berisi 10 buah pertanyaan pengguna, 10 buah jawaban dari sistem, 10 buah konteks yang dihasilkan sistem, dan 10 buah jawaban sebenarnya (*ground truth*) yang dihasilkan oleh administrator sekaligus sebagai pakar pengelolaan kinerja, seperti terlihat pada Tabel 9.

Perbandingan nilai evaluasi RAGAS antar komponen SAKIP dapat dilihat pada Gambar 27.



Gambar 27 Hasil perhitungan model RAG berdasarkan komponen SAKIP

Berikut adalah analisis mendalam untuk setiap metrik evaluasi:

1. *Faithfulness*

Skor *faithfulness* berkisar antara 0,963-0,989 dengan skor rata-rata sebesar 0,979, menjadikannya salah satu metrik dengan performa tertinggi dan variabilitas paling kecil (rentang hanya 0,026) di antara kelima aspek. Nilai tertinggi diperoleh pada komponen Pelaporan Kinerja (0,989) dan terendah pada Evaluasi Kinerja (0,963), namun perbedaan ini relatif tidak signifikan secara praktis. Secara statistik, konsistensi tinggi ini menunjukkan bahwa jawaban yang dihasilkan sistem hampir selalu dapat ditelusuri langsung ke konteks atau dokumen sumber, dengan risiko sangat rendah untuk informasi yang tidak didukung sumber di seluruh komponen SAKIP.

Contoh percakapan pada komponen Perencanaan Kinerja yang memperoleh skor *faithfulness* sebesar 1,000, dapat dilihat pada Gambar 28.

Jawab dalam bentuk narasi. Apa dasar hukum pelaksanaan SAKIP?

Berdasarkan informasi yang terdapat dalam dokumen yang tersedia, dasar hukum yang digunakan untuk Penilaian Mandiri Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) adalah Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (Permen PAN dan RB) Nomor 88 Tahun 2021 tentang Pedoman Evaluasi atas Implementasi SAKIP.

Gambar 28 Contoh Percakapan (Skor *Faithfulness*)

2. Context Precision

Context Precision menunjukkan variabilitas tertinggi di antara semua metrik, dengan rentang antara 0,733-0,954 (selisih 0,221) dan skor rata-rata sebesar 0,835. komponen Evaluasi Kinerja unggul signifikan (0,954), sedangkan Pengukuran Kinerja tertinggal jauh (0,733). Kesenjangan besar ini secara statistik mengindikasikan inkonsistensi performa mekanisme *retrieval* antar komponen dokumen yang disebabkan oleh perbedaan struktur, kepadatan, atau tingkat ambiguitas dokumen pada komponen Pengukuran dan Perencanaan Kinerja (0,779), sehingga sistem lebih sering mengambil potongan konteks yang kurang relevan sebelum menghasilkan jawaban.

Contoh percakapan pada komponen Pengukuran Kinerja yang memperoleh skor *context precision* 0,999, dapat dilihat pada Gambar 29.

Kapan, dimana dan oleh siapa Keputusan Kepala BPPSDM Nomor 177 Tahun 2026 ditetapkan?

Berdasarkan dokumen yang tersedia, Keputusan Kepala BPPSDM Nomor 177 Tahun 2026 ditetapkan:

- Kapan: Pada tanggal 9 Maret 2026.
- Dimana: Di Jakarta.
- Oleh siapa: Kepala Badan Penyuluhan dan Pengembangan Sumber Daya Manusia Kelautan dan Perikanan, yaitu I Nyoman Radiarta.

Gambar 29 Contoh Percakapan (Skor *Context Precision*)

3. Context Recall

Context Recall merupakan metrik dengan performa paling stabil dan tertinggi secara keseluruhan, dengan rata-rata skor sebesar 0,984 dan rentang sempit 0,954-1,000. Komponen Pengukuran dan Evaluasi Kinerja bahkan mencapai skor sempurna (1,000). Ini menunjukkan bahwa sistem *retrieval* hampir selalu berhasil menemukan seluruh potongan informasi relevan yang dibutuhkan untuk menjawab pertanyaan, terlepas dari skor *context precision* yang bervariasi (lihat poin 3). Hal ini berarti sistem cenderung mengambil konteks secara luas atau inklusif, sehingga kadang mengorbankan presisi demi kelengkapan.

Contoh percakapan pada komponen Evaluasi Kinerja yang memperoleh skor *context recall* 1,000, dapat dilihat pada Gambar 30.

Apakah Pengukuran kinerja secara kualitas telah menjadi kebutuhan dalam mewujudkan kinerja secara efektif dan efisien?

Berdasarkan informasi dari dokumen yang tersedia, ya, pengukuran kinerja secara kualitas telah menjadi kebutuhan dalam mewujudkan kinerja yang efektif dan efisien.

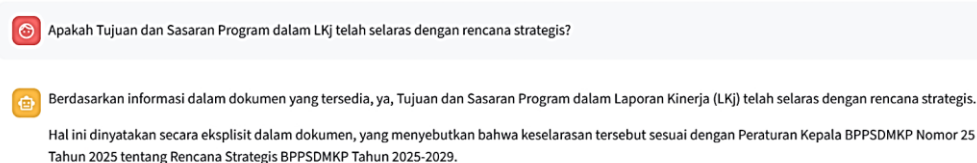
Hal ini secara eksplisit dinyatakan dalam dokumen "Hasil Evaluasi AKIP BPPSDMKP Tahun 2022" yang menyebutkan bahwa, "Pengukuran kinerja secara kualitas telah menjadi kebutuhan dalam mewujudkan kinerja secara efektif dan efisien serta telah dilakukan secara berjenjang dan berkelanjutan yang memanfaatkan Teknologi Informasi."

Gambar 30 Contoh Percakapan (Skor *Context Recall*)

4. *Answer Correctness*

Skor berkisar 0,936-0,988 dengan rata-rata skor sebesar 0,964, menempatkannya sebagai metrik dengan performa tertinggi kedua setelah *context recall*. Komponen Evaluasi Kinerja mencatat skor tertinggi (0,988) dan komponen Perencanaan Kinerja terendah (0,936), namun rentang variasi (0,052) tergolong kecil. Tingginya skor ini konsisten dengan tingginya skor *faithfulness* dan *context recall*, sekaligus mengonfirmasi bahwa jawaban akhir sistem secara umum akurat dan sesuai dengan jawaban referensi atau *ground truth*.

Contoh percakapan pada komponen Pelaporan Kinerja yang memperoleh skor *answer correctness* 1,000, dapat dilihat pada Gambar 31.

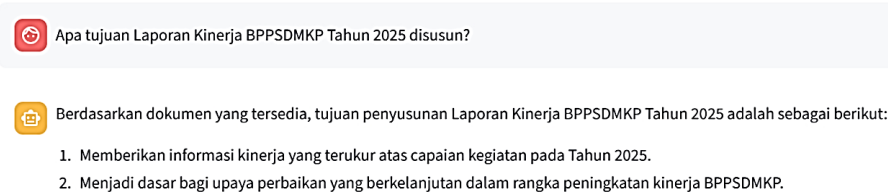


Gambar 31 Contoh Percakapan (Skor *Answer Correctness*)

5. *Answer Relevancy*

Aspek ini memiliki rata-rata terendah di antara seluruh metrik, yaitu 0,816, dengan rentang skor 0,770-0,870. Komponen Pengukuran Kinerja mencatat skor tertinggi (0,870), sementara komponen Pelaporan Kinerja mencatat skor terendah (0,770). Rentang variasi (0,100) menunjukkan ketidakkonsistenan yang lebih besar dibanding *faithfulness* dan *context recall*. Ini mengindikasikan bahwa meskipun jawaban sistem faktual dan bersumber dari dokumen yang benar, terdapat kecenderungan jawaban kurang fokus atau kurang langsung menjawab inti pertanyaan pengguna, khususnya pada komponen Pelaporan dan Evaluasi Kinerja yang kemungkinan memiliki pertanyaan lebih kompleks.

Contoh percakapan pada komponen Pelaporan Kinerja yang memperoleh skor *answer relevancy* 0,921, dapat dilihat pada Gambar 32.



Gambar 32 Contoh Percakapan (Skor *Answer Relevancy*)

4.10 Evaluasi Manfaat

Evaluasi terhadap manfaat sistem ini menggunakan metode *System Usability Scale* (SUS). Evaluasi ini bertujuan untuk mengukur tingkat kebergunaan (*usability*) dan pengalaman pengguna (*user experience*) terhadap sistem. Pengujian ini dilakukan untuk mengidentifikasi sejauh mana sistem dapat digunakan dengan mudah, efisien, dan memuaskan oleh target pengguna.

Pengujian ini dilaksanakan pada tanggal 22 Juni 2026 melalui *Google Form* dengan melibatkan 35 orang responden (laki-laki sebanyak 19 orang dan perempuan sebanyak 16 orang) yang mewakili pengguna akhir (anggota tim pengelola kinerja sekaligus operator aplikasi) pada satker lingkup BPPSDMKP.

Responden survei memiliki rentang usia antara 33 hingga 56 tahun, dengan mayoritas berada pada kelompok usia dewasa muda hingga paruh baya (33–40 tahun). Usia 35 tahun merupakan kelompok terbanyak, mencakup empat responden atau 26,67% dari total responden. Diikuti oleh usia 33 tahun dan 40 tahun yang masing-masing diisi oleh 2 responden (13,33%). Sisanya tersebar merata di usia 34, 36, 38, 39, 42, 44, dan 56 tahun, masing-masing dengan 1 responden (6,67%). Secara umum, sebagian besar responden (73,33%) berusia antara 33–40 tahun, menunjukkan bahwa mayoritas partisipan survei berada pada rentang usia produktif, dengan satu responden berusia jauh lebih tua (56 tahun) yang menjadi *outlier* dalam distribusi ini. Setiap responden diminta untuk menyelesaikan beberapa skenario tugas (*task*) yang umum dilakukan di dalam sistem, kemudian mengisi kuesioner SUS yang terdiri dari 10 pernyataan (skala Likert 1-5), seperti terlihat pada Lampiran 14.

Secara umum, responden memberikan respons positif terhadap sistem yang dievaluasi. Hal ini terlihat dari tingginya persetujuan pada item-item pernyataan positif, seperti kemudahan penggunaan sistem (65,71% Setuju), keinginan untuk menggunakan sistem ini kembali (48,57% Sangat Setuju dan 48,57% Setuju), serta kepercayaan bahwa fitur-fitur sistem berjalan dengan semestinya (74,29% Setuju). Responden juga merasa sistem ini mudah dipahami oleh orang lain (68,57% Setuju) dan tidak menemukan banyak hambatan dalam penggunaannya (68,57% Setuju).

Pada sisi pernyataan negatif, mayoritas responden juga menunjukkan sikap yang konsisten dengan penilaian positif, yaitu dengan menyatakan ketidaksetujuan terhadap pernyataan bahwa sistem ini rumit (54,29% Tidak Setuju), membingungkan (74,29% Tidak Setuju), maupun tidak konsisten (71,43% Tidak Setuju). Meski demikian, terdapat catatan pada item kebutuhan bantuan teknis, di mana 51,43% responden menyatakan tidak setuju bahwa mereka memerlukan bantuan, namun proporsi setuju masih cukup terlihat (22,86%), yang mengindikasikan sebagian kecil pengguna masih memerlukan pendampingan awal.

Berdasarkan hasil perhitungan skor SUS menggunakan metode standar Bangor *et al.* (2009), skor keseluruhan yang diperoleh berada pada kisaran 73,21, skor tersebut berada pada kategori "*Good*" (Baik) dengan tingkat penerimaan "*Acceptable*" (Dapat Diterima), karena berada di atas nilai rata-rata SUS (68). Terdapat kesamaan penggunaan SUS sebagai instrumen evaluasi pada penelitian Hermadi *et al.* (2021) yang mengembangkan dan mengevaluasi prototipe KMS berbasis partisipasi untuk petani kelapa sawit menggunakan metode *System Usability Scale* (SUS) dan *Cognitive Task Analysis* (CTA) dengan skor usability 80,625. Meskipun skor SUS yang diperoleh saat ini belum sebaik hasil penelitian tersebut, tetapi pencapaian ini menunjukkan bahwa sistem yang dievaluasi secara keseluruhan telah mudah digunakan (*usable*) dan dapat diterima dengan baik oleh pengguna, meskipun masih terdapat ruang perbaikan pada aspek kemandirian pengguna dalam mengoperasikan sistem tanpa bantuan teknis. Temuan ini sejalan dengan pandangan Dwivedi *et al.* (2021) bahwa keberhasilan adopsi kecerdasan buatan di sektor publik tidak semata bergantung pada kecanggihan model,

melainkan juga pada faktor kesiapan kelembagaan, pelatihan pengguna, dan desain antarmuka yang ramah bagi aparatur dengan latar belakang teknis yang beragam.

4.11 Evaluasi *Dashboard*

Evaluasi terhadap *dashboard* dan model RAG menggunakan metode *Wilcoxon Signed-Rank Test*. Evaluasi ini bertujuan untuk membandingkan dua sampel berpasangan, pengukuran sebelum dan sesudah intervensi pada subjek yang sama). Pengujian ini dilakukan untuk mengidentifikasi sejauh mana pengaruh sistem sebelumnya (Aplikasi e-SAKIP Reviu) dan sistem sesudahnya (Aplikasi DSMS SAKIP BPPSDMKP).

Pengujian ini dilaksanakan pada tanggal 20 Mei 2026 untuk *Pretest* dan tanggal 22 Juni 2026 untuk *Posttest* melalui *Google Form* dengan melibatkan 15 orang responden (laki-laki sebanyak 7 orang dan perempuan sebanyak 8 orang) yang mewakili pengguna akhir (anggota tim pengelola kinerja sekaligus operator aplikasi) pada satker lingkup BPPSDMKP.

Responden survei memiliki rentang usia yang cukup luas, yaitu antara 32 hingga 56 tahun. Kelompok usia yang paling banyak muncul adalah pada usia 35 dan 42 tahun, dengan dua titik usia yang masing-masing diisi oleh 4 responden (11,43%) sebagai proporsi tertinggi. Selain itu, terdapat beberapa kelompok usia 32, 33 dan 40 tahun dengan 3 responden (8,57%) dan 2 responden (5,71%) yang tersebar di berbagai titik usia seperti kelompok usia 39, 41, 43, 44, 45, dan 49 tahun. Sisanya merupakan usia-usia dengan jumlah responden lebih sedikit (1 responden atau 2,86%), dengan kelompok usia 34, 37, 47, 50, 51 dan 56 tahun. Secara umum, mayoritas responden berada pada rentang usia produktif 32–44 tahun, dengan sebagian kecil responden berusia di atas 47 tahun, menunjukkan bahwa peserta survei didominasi oleh kelompok usia dewasa hingga paruh baya dengan variasi usia yang cukup beragam.

Setiap responden diminta untuk membandingkan kedua aplikasi tersebut, kemudian mengisi kuesioner yang terdiri dari 21 pernyataan (skala Likert 1-5). Pertanyaan tersebut dibagi menjadi empat bagian (seperti terlihat pada Lampiran 15-16), meliputi:

1. Manfaat *dashboard* perbandingan sebanyak tujuh pertanyaan, selaras dengan studi Yigitbasioglu dan Velcu (2012) yang mengkaji bagaimana fitur visualisasi (*dashboard*) dalam manajemen kinerja pemerintah/perusahaan dapat meningkatkan kecepatan pemrosesan informasi pengguna dibandingkan membaca dokumen terpisah secara manual.
2. Pencarian informasi kinerja sebanyak enam pertanyaan, dengan mengukur kelancaran alur tanya jawab, rasa percaya pengguna pada sistem, serta efisiensi fungsi *query* Borsci *et al.* (2023). Hal ini juga selaras dengan studi Ren *et al.* (2022) tentang indikator efektivitas dan efisiensi sistem *Natural Language Processing* (NLP) dalam menjawab pertanyaan spesifik, selaras dengan metode kerja RAG yang digunakan.
3. Perbandingan regulasi atau dokumen kinerja sebanyak enam pertanyaan, yang mengevaluasi dokumen SAKIP, dengan menampilkan metode analisis dokumen kinerja yang dilakukan secara manual dan membandingkannya dengan AI. Selaras dengan studi Zhou dan Hu (2025) yang membandingkan efisiensi waktu pemrosesan 91 artikel oleh manusia dan dua model LLM (ChatGPT-4.0 dan DeepSeek R1). Rata-rata waktu pemrosesan per artikel

oleh LLM hanya 33,09 detik, sedangkan manusia membutuhkan 1.582,50 detik. Hal ini menunjukkan peningkatan efisiensi sebesar 47,80 kali lipat.

4. Manfaat sistem secara umum sebanyak dua pertanyaan, yang diadopsi dari Bangor *et al.* (2009) sebagai landasan utama dalam memvalidasi keandalan instrumen evaluasi sistem.

Pengujian hipotesa menggunakan *Uji Wilcoxon Signed-Rank*. Secara sederhana, kita dapat membandingkan nilai W_{Hitung} menggunakan persamaan 15 berikut.

$$W_{Hitung} = \min(TotalR_+, TotalR_-) \quad (15)$$

Kemudian membandingkannya dengan *Table of Critical Values for the Wilcoxon Test* (Sani dan Todman 2006). Jika $W_{Hitung} > W_{Kritis}$, maka H_0 diterima, H_1 ditolak, tetapi jika $W_{Hitung} \leq W_{Kritis}$, maka H_1 diterima, H_0 ditolak.

Wilcoxon Signed Rank Test dilakukan untuk mengevaluasi perbedaan antara *Pretest* menggunakan Aplikasi e-SAKIP Reviu dan *Posttest* menggunakan Aplikasi DSMS SAKIP BPPSDMKP (rincian data dapat dilihat pada Lampiran 17-18), seperti terlihat pada Tabel 10 dan Tabel 11.

Tabel 10 Hasil Survey *Pretest* dan *Posttest*

Responden	Pre	Post	Post-Pre	Post-Pre	Rank (+)	Rank (-)
1	3,571	4,810	1,238	1,238	9,5	
2	3,524	3,905	0,381	0,381	3	
3	4,286	4,000	-0,286	0,286		2
4	3,857	4,429	0,571	0,571	5	
5	3,762	4,524	0,762	0,762	6	
6	2,619	3,952	1,333	1,333	11	
7	3,429	5,000	1,571	1,571	12,5	
8	4,476	5,000	0,524	0,524	4	
9	2,238	4,333	2,095	2,095	15	
10	3,048	4,619	1,571	1,571	12,5	
11	2,905	4,714	1,810	1,810	14	
12	3,762	4,619	0,857	0,857	7	
13	4,048	4,238	0,190	0,190	1	
14	2,762	3,952	1,190	1,190	8	
15	3,571	4,810	1,238	1,238	9,5	

Sumber: Data primer diolah, 2026.

Hipotesis Pengujian sebagai berikut:

H_0 : Tidak ada perbedaan signifikan sebelum dan sesudah penggunaan *Dashboard* dan Model RAG pada Aplikasi DSMS SAKIP BPPSDMKP

H_1 : Terdapat perbedaan signifikan (*Dashboard* dan Model RAG pada Aplikasi DSMS SAKIP BPPSDMKP efektif digunakan)

Tabel 11 Hasil Uji *Wilcoxon Signed-Rank*

Indicator (Post - Pre)	N	Mean Rank	Sum of Ranks
Negative Ranks	1	2	2
Positive Ranks	14	8.249	118
Ties	0	-	-

Sumber: Data primer diolah, 2026.

Berdasarkan tabel di atas, nilai W_{Hitung} sebesar 2, yang dihitung berdasarkan nilai minimum dari total peringkat negatif dan peringkat positif. Merujuk pada tabel nilai kritis uji *Wilcoxon Sani* dan *Todman* (2006), untuk total responden sebanyak 15 orang, uji dua arah dengan tingkat signifikansi α sebesar 0,05 menghasilkan nilai W_{Kritis} sebesar 25. Berdasarkan kondisi tersebut, maka diketahui $W_{\text{Hitung}} \leq W_{\text{Kritis}}$, sehingga dapat disimpulkan bahwa H_1 diterima dan H_0 ditolak.

Kita dapat mengambil pendekatan lain dengan mengubah nilai statistik *Wilcoxon* menjadi nilai Z terstandarisasi yang dinormalisasi, kemudian menghitung probabilitas kumulatif dari nilai Z tersebut. Dengan nilai probabilitas lebih kecil dari tingkat signifikansi 0,05, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara *pretest* dan *posttest*.

Hasil uji *Wilcoxon Signed-Rank* yang menunjukkan perbedaan sangat signifikan ($Z = -3,294$; $p < 0,05$) memperkuat temuan sebelumnya mengenai efektivitas Model RAG dalam tugas-tugas pengetahuan-intensif (Lewis *et al.* 2021; Fan *et al.* 2024). Secara praktis, hal ini membuktikan bahwa pemanfaatan *Dashboard* dan Model RAG pada Aplikasi DSMS SAKIP BPPSDMKP ini memiliki dampak yang efektif terhadap perubahan nilai dari kondisi awal ke kondisi akhir (H_1 diterima, H_0 ditolak).



V SIMPULAN DAN SARAN

5.1 Simpulan

Model ini telah menggabungkan dua fungsi utama yaitu evaluasi SAKIP (analisis data penilaian kinerja, matriks nilai, dan visualisasi infografis) dan penerapan *Retrieval-Augmented Generation* (RAG) yang dapat menjawab pertanyaan berdasarkan dokumen kinerja internal menggunakan Model Gemini 2.5 Pro. Penerapan AI ini mengubah cara evaluasi dilakukan menjadi sistem pendukung keputusan yang mengurangi tugas-tugas rutin bagi asesor, sehingga mereka dapat lebih fokus pada analisis konteks strategis yang berorientasi pada hasil. Meskipun demikian, kecerdasan buatan dalam sistem ini berfungsi sebatas sebagai alat bantu, sedangkan keputusan akhir tetap sepenuhnya berada pada pengguna.

Kemampuan sistem dalam mencari dan menyajikan dokumen referensi mengalami peningkatan luar biasa. Skor *Context Precision* mencapai 0,835, yang berarti sistem kini dapat mengambil *chunks* dokumen yang relevan. Hal ini didukung oleh skor *Context Recall* yang mencapai 0,984, yang berarti sistem dipastikan mampu mengumpulkan seluruh informasi krusial yang dibutuhkan untuk menjawab pertanyaan. Pada komponen LLM, skor *Faithfulness* menyentuh angka 0,979, yang menandakan bahwa jawaban sistem sepenuhnya patuh pada dokumen acuan. Perbaikan paling krusial terjadi pada *Answer Relevancy* yang memperoleh skor 0,816 yang menunjukkan bahwa respons sistem kini jauh lebih fokus, padat, dan langsung menjawab esensi pertanyaan pengguna. Sebagai dampak linear dari optimalnya proses pencarian konteks dan penalaran LLM, *Answer Correctness* memperoleh skor 0,979, yang berarti secara faktual dan semantik, jawaban yang dihasilkan sistem kini hampir selaras sepenuhnya dengan dokumen kebenaran (*ground truth*).

Skor *System Usability Scale* (SUS) yang diperoleh sebesar 73,21 menunjukkan bahwa sistem berada pada kategori "Good" dengan tingkat penerimaan "Acceptable", menandakan bahwa model ini secara umum siap digunakan secara fungsional namun masih memiliki ruang untuk optimasi minor. Hasil uji *Wilcoxon Signed-Rank* yang menunjukkan perbedaan sangat signifikan ($Z = -3,294$; $p < 0,05$), sebelum dan sesudah menggunakan *dashboard* dan model RAG dalam implementasi SAKIP.

5.2 Saran

Sistem ini akan terus berkembang seiring bertambahnya kebutuhan informasi, sehingga dalam proses pengembangannya, disarankan agar tim pengelola kinerja yang ada di seluruh satker BPPSDMKP dapat memperbaiki standarisasi data kinerja. Karena metode RAG memerlukan *file* PDF berupa teks yang bisa dicari, bukan gambar hasil *scan*, maka standarisasi menjadi syarat yang wajib dan tidak bisa dihindari sehingga sistem evaluasi birokrasi yang dibuat bisa lebih baik.

Skor *Answer Relevancy* dan *Context Precision* belum menyentuh angka lebih dari 0,900, sehingga disarankan kepada pengguna sistem agar melakukan *prompting engineering*, misalnya dengan menambahkan instruksi batasan panjang kalimat (*output length constraint*) atau pendekatan *few-shot prompting* sehingga LLM memberikan jawaban yang lebih taktis dan bebas dari kalimat pengantar yang redundan. Selain itu kepada pengembang selanjutnya disarankan untuk merancang

algoritma *re-ranking* yang lebih baik, sehingga LLM mendapatkan konteks yang lebih bersih dan fokus. Melakukan uji usabilitas lanjutan secara berkala setelah perbaikan antarmuka (UI/UX) diimplementasikan, untuk memvalidasi apakah perubahan tersebut secara signifikan menaikkan tingkat kepuasan dan efisiensi pengguna.

@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR PUSTAKA

- Afzal A, Vladika J, Fazlija G, Staradubets A, Matthes F. 2025. Towards optimizing a retrieval augmented generation using large language model on academic data. Di dalam: *Proceedings of the 2024 8th International Conference on Natural Language Processing and Information Retrieval*. New York (NY): Association for Computing Machinery. hlm 250–257.
- Altarazi F, Santos D, Tomchesson A. 2023. Framework for developing interactive dashboards and evaluating their usage effectiveness: a case study. Di dalam: *Proceedings of the 8th North American International Conference on Industrial Engineering and Operations Management*. Houston (TX): IEOM Society International.
- Anggraini FA, Akbar R. 2022. Penerapan sistem akuntabilitas kinerja instansi pemerintah (studi pada Pemerintah Kota Yogyakarta). *Account Bus Inf Syst J*. 10(4).
- Bach B, Freeman E, Abdul-Rahman A, Turkay C, Khan S, Fan Y, Chen M. 2022. Dashboard design patterns. *IEEE Trans Vis Comput Graph*. 2022 Sep. doi:10.1109/TVCG.2022.3209448.
- Bangor A, Kortum P, Miller J. 2009. Determining what individual SUS scores mean: adding an adjective rating. *J Usability Stud*. 4(3):114–123.
- Bangor A, Kortum PT, Miller JT. 2008. An empirical evaluation of the System Usability Scale. *Int J Human-Computer Interact*. 24(6):574–594. doi:10.1080/10447310802205776.
- Bayer H, Aracı FGİ, Gürkan G. 2024. ChatGPT-4o, ChatGPT-4 and Google Gemini are compared with students: a study in higher education. *Int J Technol Educ Sci*. 8(4):627–644. doi:10.46328/ijtes.585.
- Borsci S, Schmettow M, Malizia A, Chamberlain A, van der Velde F. 2023. A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. *Pers Ubiquitous Comput*. 27(2):317–330. doi:10.1007/s00779-022-01690-0.
- Brooke J. 1996. SUS - a quick and dirty usability scale. *Usability Eval Ind*. 189(3):189–194.
- Burbano J, Landeta-López P, Guevara-Vega C, Quiña-Mera A. 2026. Design and evaluation of a retrieval-augmented generation LLM chatbot with structured database access. *Appl Sci*. 16(7):3147. doi:10.3390/app16073147.
- Comanici G, Bieber E, Schaeckermann M, Pasupat I, Sachdeva N, Dhillon I, Blistein M, Ram O, Zhang D, Rosen E, et al. 2025. *Gemini 2.5: pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities*. Mountain View (CA): Google.
- Douze M, Guzhva A, Deng C, Johnson J, Szilvasy G, Mazaré PE, Lomeli M, Hosseini L, Jégou H. 2025. The Faiss library. *IEEE Trans Big Data*. 12:346–361. doi:10.1109/TBDDATA.2025.3618474.
- Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Duan Y, Dwivedi R, Edwards J, Eirug A, et al. 2021. Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int J Inf Manage*. 57:101994. doi:10.1016/j.ijinfomgt.2019.08.002.

- Es S, James J, Espinosa-Anke L, Schockaert S. 2024. Ragas: automated evaluation of retrieval augmented generation. *EACL 2024 - 18th Conf Eur Chapter Assoc Comput Linguist Proc Syst Demonstr.*, siap terbit.
- Fan W, Ding Y, Ning L, Wang S, Li H, Yin D, Chua TS, Li Q. 2024. A survey on RAG meeting LLMs: towards retrieval-augmented large language models. [diakses 8 Jul 2026]. <https://arxiv.org/abs/2405.06211>.
- Fanelli F, Saleh M, Troiano G. 2025. Development and comparative evaluation of a reinstructed GPT-4o model specialized in periodontology. *J Clin Periodontol*. 52:707–716. doi:10.1111/jcpe.14101.
- Few S. 2006. *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol (CA): O'Reilly.
- Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, Dai Y, Sun J, Wang M, Wang H. 2024. Retrieval-augmented generation for large language models: a survey. *Proc - 2024 Conf AI, Sci Eng Technol AIXSET 2024.*, siap terbit.
- [Google] Google Cloud. 2026. Get text embeddings. [diakses 19 Jul 2026]. <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/embeddings/get-text-embeddings>.
- Hendrik L, Asri S, Santosa A. 2022. Akuntabilitas kinerja instansi pemerintah pada Pemerintah Kabupaten Paser. *J Publicuho*. 5(4):1367–1376.
- Hermadi I, Kesuma INR, Nurhadryani Y, Asfarian A. 2021. Usability evaluation of the participatory-based KMS Sawit mobile application. Di dalam: *2021 2nd International Conference on ICT for Rural Development (IC-ICTRuDev)*. hlm 1–6.
- [Permen KP] Peraturan Menteri Kelautan dan Perikanan Republik Indonesia Nomor 35 Tahun 2023 tentang Pengelolaan Kinerja Organisasi di Lingkungan Kementerian Kelautan dan Perikanan. 2023. [diakses 22 Jun 2025]. <https://jdih.kkp.go.id/Homedev/DetailPeraturan/5088>.
- Jana S, Biswas R, Pal K, Biswas S, Roy K. 2024. The evolution and impact of large language model systems: a comprehensive analysis. *Alochana J*. 13(3):65–77.
- Karakurt E, Akbulut A. 2026. Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) for enterprise knowledge management and document automation: a systematic literature review. *Appl Sci*. 16(1):368. doi:10.3390/app1601368.
- Kobi J. 2024. Developing dashboard analytics and visualization tools for effective performance management and continuous process improvement. *Int J Innov Sci Res Technol*. 2024 Mei 9. doi:10.38124/ijisrt/IJISRT24MAY1147.
- Lestari S, Salomo RV. 2022. Implementasi sistem akuntabilitas kinerja pada pemerintah daerah di Indonesia: sebuah tinjauan literatur. *J Moderat*. 8(1):798–808.
- Leutz-Schmidt P, Palm V, Mathy RM, Grözinger M, Kauczor HU, Jang H, Sedaghat S. 2025. Performance of large language models ChatGPT and Gemini on workplace management questions in radiology. *Diagnostics*. 15(4):497. doi:10.3390/diagnostics15040497.
- Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W, Rocktäschel T, *et al.* 2021. Retrieval-augmented generation for knowledge-intensive NLP tasks. [diakses 31 Jul 2026]. <https://arxiv.org/abs/2005.11401>.

- Marjan NA, Supianto AA, Fanani L. 2021. Perancangan user experience aplikasi E-SAKIP (Elektronik Sistem Akuntabilitas Kinerja Instansi Pemerintah) menggunakan metode Human-Centered Design (studi kasus: Kabupaten Kepulauan Selayar). *J Pengemb Teknol Inf dan Ilmu Komput.* 5(10):4175–4181.
- Matheus R, Janssen M, Maheshwari D. 2020. Data science empowering the public: data-driven dashboards for transparent and accountable decision-making in smart cities. *Gov Inf Q.* 37(3):101284. doi:10.1016/j.giq.2018.01.006.
- Jensen MM, Danielsen MB, Riis J, Kristjansen KA, Andersen S, Okubo Y, Jørgensen MG. 2025. ChatGPT-4o can serve as the second rater for data extraction in systematic reviews. *PLoS One.* 20(1):1–12. doi:10.1371/journal.pone.0313401.
- Neha F, Bhati D, Shukla DK. 2025. Retrieval-Augmented Generation (RAG) in healthcare: a comprehensive review. *AI.* 6(9):226. doi:10.3390/ai6090226.
- Pradana AAA, Maghfiroh LR. 2024. Kajian peningkatan implementasi sistem akuntabilitas kinerja instansi pemerintah melalui pembangunan aplikasi evaluasi akuntabilitas kinerja instansi pemerintah. Di dalam: *Seminar Nasional Official Statistics.* hlm 819–830.
- Pratama AY. 2024. Efektivitas tata kelola Sistem Akuntabilitas Kinerja Instansi Pemerintah (SAKIP) pada Kementerian Kesehatan Republik Indonesia [skripsi]. Bogor: Institut Pertanian Bogor.
- Raharja EL. 2018. Strategi peningkatan Sistem Akuntabilitas Kinerja Instansi Pemerintah Kota Bogor [tesis]. Bogor: Institut Pertanian Bogor.
- Raharja EL, Baga LM, Falatehan AF. 2018. Strategi peningkatan akuntabilitas kinerja instansi pemerintah Kota Bogor. *J Manaj Pembang Drh.* 10(2). doi:10.29244/jurnal_mpd.v10i2.27787.
- Rane N, Choudhary S, Rane J. 2024. Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation. *J Appl Artif Intell.* 5(1):69–93. doi:10.48185/jaai.v5i1.1052.
- Rasmussen NH, Bansal M, Chen CY. 2009. *Business Dashboards: A Visual Catalog for Design and Deployment.* Hoboken (NJ): John Wiley & Sons.
- Ren R, Zapata-Rodríguez M, Castro J, Dieste O, Acuña S. 2022. Experimentation for chatbot usability evaluation: a secondary study. *IEEE Access.* 10:12430–12464. doi:10.1109/ACCESS.2022.3145323.
- Sadewa IGBB, Divayana DGH, Pradnyana IMA. 2020. Pengujian usability pada aplikasi E-SAKIP Kabupaten Buleleng menggunakan metode usability testing. *Inf Syst Emerg Technol J.* 1(2):76–87.
- Salomo RV, Rahmayanti KP. 2023. Progress and institutional challenges on local governments performance accountability system reform in Indonesia. *SAGE Open.* 2023 Des. doi:10.1177/21582440231196659.
- Sani F, Todman J. 2006. Appendix 1: Statistical tables. Di dalam: *Experimental Design and Statistics for Psychology.* Chichester: John Wiley & Sons. hlm 183–196.
- Saputra H, Pradana F, Priyambadha B. 2017. Pengembangan elektronik-sistem akuntabilitas kinerja instansi pemerintah untuk Dinas Perindustrian Kota Malang. *J Pengemb Teknol Inf dan Ilmu Komput.* 1(10):1167–1175.

- Sihombing WW, Aryadita H, Rusdianto DS. 2019. Perancangan dashboard untuk monitoring dan evaluasi (studi kasus: FILKOM UB). *J Pengemb Teknol Inf dan Ilmu Komput*. 3(1):434–441.
- Team L, Modi A, Veerubhotla A, Rysbek A, Huber A, Anand A, Bhoopchand A, Wiltshire B, Gillick D, Kasenberg D, *et al*. 2025. *Evaluating Gemini in an arena for learning*. Mountain View (CA): Google.
- Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le Q, Zhou D. 2023. Chain-of-thought prompting elicits reasoning in large language models. [diakses 19 Agu 2026]. <https://arxiv.org/abs/2201.11903>.
- Widhani A, Maria S, Chairani AP, Visco NYA, Nurhadi MFA, Harjoprawito LA, Elvira D, Tedja I, Yuniza Y, Fetarayani D, *et al*. 2026. Comparing the accuracy of ChatGPT-4o, DeepSeek-V3, and Gemini 2.5 Flash in answering frequently asked questions about Systemic Lupus Erythematosus: quantitative study. *JMIR Form Res*. 10:e89156. doi:10.2196/89156.
- Wilcoxon F. 1945. Individual comparisons by ranking methods. *Biometrics Bull*. 1(6):80–83. doi:10.2307/3001968.
- Woolson RF. 2008. Wilcoxon signed-rank test. Di dalam: *Wiley Encyclopedia of Clinical Trials*. Hoboken (NJ): John Wiley & Sons. hlm 1–3.
- Yassir M, Agunawan, Gafur, Rahman A, Fitriati I. 2024. Sistem informasi laporan akuntabilitas kinerja instansi pemerintah. *J Inform Sains dan Teknol*. 9(1):156–165.
- Yigitbasioglu OM, Velcu O. 2012. A review of dashboards in performance management: implications for design and research. *Int J Account Inf Syst*. 13(1):41–59. doi:10.1016/j.accinf.2011.08.002.
- Zhou Y, Hu H. 2025. Can AI assess literature like experts? An entropy-based comparison of ChatGPT-4o, DeepSeek R1, and human ratings. *Front Res Metrics Anal*. 10:1684137. doi:10.3389/frma.2025.1684137.

