PENGEMBANGAN CHATBOT UNTUK LAYANAN SATUAN PENJAMINAN MUTU (SPM) MENGGUNAKAN BERT (BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS)

1st Zumar Nur Firdaus
Program Studi Teknologi Informasi
Universitas Telkom
Surabaya, Indonesia
zmrnrf@student.telkomuniversity.ac.id

2nd Yohanes Setiawan
Program Studi Teknologi Informasi
Universitas Telkom
Surabaya, Indonesia
yohanessetiawan@telkomuniversity.ac.id

3rd Mastuty Ayu Ningtyas
Program Studi Teknologi Informasi
Universitas Telkom
Surabaya, Indonesia
mastutyayu@telkomuniversity.ac.id

***Abstrak* — Kendala Satuan Penjaminan Mutu (SPM) dalam penyampaian informasi yang kurang efisien, menyebabkan pelayanan Satuan Penjaminan Mutu (SPM) kurang responsif. Sehingga dibutuhkannya chatbot yang merespon secara cepat dan tepat. Penelitian ini berfokus pada pengembangan chatbot berbasis deep learning dengan arsitektur BERT untuk layanan Satuan Penjaminan Mutu (SPM) sebagai platform pembantu. Tujuannya adalah untuk memaksimalkan pencarian data dan penyampaian informasi yang efisien kepada pengguna. Metodologi yang digunakan meliputi pengumpulan data mentah dari SPM (Penelitian, Pengabdian, Prestasi, Mahasiswa, dan Dosen dari 2019-2024), pra-pengolahan data (tokenisasi, padding, label encoding), dan pengembangan model BERT yang diintegrasikan dengan platform Telegram. Evaluasi model melalui 15 percobaan menunjukkan performa optimal pada konfigurasi tertentu dengan akurasi tinggi dan tingkat kesalahan rendah, terbukti dari F1-score kelas positif mencapai 0.9831. Meskipun terdapat sedikit ketidaksesuaian pada satu skenario pengujian black-box dan adanya cold start pada respons awal chatbot, hasil kuesioner pengguna secara keseluruhan menunjukkan kepuasan tinggi terhadap kemudahan penggunaan dan kecepatan respons. Implementasi chatbot ini diharapkan meningkatkan efisiensi layanan SPM, meskipun perlu penambahan variasi data pelatihan dan deployment ke server untuk optimalisasi lebih lanjut.**

Kata kunci*— *Chatbot, Deep Learning, BERT (Bidirectional Encoder Representations from Transformers), Natural Language Processing, Blackbox Testing, Telegram.

I. PENDAHULUAN

Pada sebuah instansi perguruan tinggi, Satuan Penjaminan Mutu (SPM) memegang peran sentral dalam mengelola data akreditasi yang krusial. Namun, tantangan besar muncul karena instansi dengan data berjumlah besar memerlukan sistem pencarian yang cepat dan akurat untuk pemanfaatan data yang relevan dan efektif [1]. Saat ini, proses penyampaian informasi dari SPM kepada pengguna belum berjalan efisien, yang mengakibatkan keterlambatan, terutama dalam situasi mendesak atau ketika terdapat banyak permintaan dari berbagai unit. Inefisiensi ini menjadi permasalahan utama yang menghambat kelancaran proses akreditasi dan pemenuhan kebutuhan informasi internal.

Sebagai solusi, *chatbot* diusulkan karena sangat berguna untuk meningkatkan efisiensi di sektor pendidikan, terutama pada perguruan tinggi yang mengelola ribuan data [2]. *Chatbot* adalah program komputer yang dirancang untuk mensimulasikan percakapan manusia [3] dan dapat diintegrasikan ke berbagai platform seperti web dan aplikasi seluler [4]. Teknologi ini didukung oleh komponen utama seperti *Natural Language Processing* (NLP) dan *Deep Learning*. Model *Deep Learning* sering kali menunjukkan kinerja yang lebih superior dibandingkan pendekatan tradisional [5] dan dianggap sebagai solusi efektif untuk menganalisis data kompleks berdimensi tinggi di lingkungan universitas [6].

Beberapa arsitektur *Deep Learning* yang canggih dapat diterapkan dalam pembuatan *chatbot*. Model Transformer, misalnya, diperkenalkan untuk mengatasi kelemahan RNN/LSTM dalam memahami konteks jangka panjang [7]. Ada pula Retrieval-Augmented Generation (RAG), sebuah teknik yang mengekstrak data dari sumber pengetahuan eksternal untuk meningkatkan relevansi dan keakuratan jawaban [8]. Selain itu, arsitektur populer seperti Recurrent Neural Network (RNN) dengan Long Short-Term Memory (LSTM) digunakan agar mesin dapat memahami bahasa dan sintaksis manusia [9]. Model lain yang sangat berpengaruh adalah BERT, yang unggul karena kemampuannya melakukan pra-pelatihan representasi dua arah dari teks [10]

Penelitian ini berfokus untuk meningkatkan kualitas layanan Satuan Penjaminan Mutu dalam penyampaian informasi. Tujuan utamanya adalah memberikan pengalaman pengguna yang lebih baik dengan mengatasi masalah efisiensi dalam penyampaian data akreditasi. Melalui solusi yang diusulkan, diharapkan instansi perguruan tinggi dapat memperoleh informasi secara lebih efisien dan efektif. Pada akhirnya, hal ini akan mendukung peningkatan kualitas layanan secara keseluruhan dan membantu proses pengambilan keputusan yang lebih baik di masa mendatang.

II. KAJIAN TEORI

A. Chatbot

Chatbot merupakan sebuah program komputer yang dirancang untuk berkomunikasi dan berinteraksi dengan pengguna melalui dialog berbasis teks maupun suara untuk memenuhi kebutuhan pengguna. Dalam pengembangannya, *chatbot* memanfaatkan beberapa komponen utama seperti pemrosesan bahasa alami (NLP), pemahaman bahasa alami (NLU), dan generasi bahasa alami (NLG). Seiring perkembangannya, fungsi *chatbot* telah meluas dari sekadar hiburan ke berbagai bidang penting seperti pendidikan, bisnis, dan pengambilan informasi [11].

B. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang dari kecerdasan buatan yang menjadi fondasi utama bagi *chatbot* untuk dapat memahami dan berinteraksi dengan bahasa manusia. NLP membantu komputer memproses teks masukan, mengidentifikasi maksud pengguna, dan menyimpulkan makna dari kalimat yang kompleks. Tanpa NLP, *chatbot* akan terbatas pada respons yang kaku, sehingga NLP menjadi bagian yang sangat penting dari pembuatan *chatbot* yang cerdas dan adaptif [12], [13].

C. Deep Learning

Deep learning adalah cabang dari *machine learning* dan kecerdasan buatan yang menggunakan *neural network* berlapis (Deep Neural Network) untuk menganalisis data. Teknologi ini sangat efektif dalam memahami bahasa alami karena kemampuannya mempelajari data dalam skala besar dan menghasilkan representasi data tingkat tinggi. Kinerja *deep learning* telah meningkat secara signifikan dan dianggap efektif untuk menganalisis arsitektur dengan data berskala tinggi, seperti yang umum ditemukan di lingkungan universitas [6].

D. BERT (Bidirectional Encoder Representation from Transformers)

BERT adalah model pemrosesan bahasa alami (NLP) yang dikembangkan oleh Google berdasarkan arsitektur Transformer. Keunggulan utama model ini adalah kemampuannya untuk memahami konteks sebuah kata secara dua arah (bidirectional), yaitu dari kiri dan kanan secara bersamaan di semua lapisan. Arsitektur BERT didasarkan pada *multilayer bidirectional transformer* [14]. Untuk tugas seperti *Question Answering*, BERT memproses input yang terdiri dari pertanyaan dan paragraf konteks, kemudian menangkap hubungan kontekstual antar token untuk memprediksi posisi awal dan akhir jawaban di dalam paragraf tersebut [15].

III. METODE

Penelitian ini dilaksanakan melalui beberapa tahapan sistematis yang mencakup pengumpulan data, pra-pengolahan data, pengembangan model, dan evaluasi sistem secara menyeluruh. Alur perancangan ini dimulai dari akuisisi data mentah hingga pengujian fungsionalitas dan kinerja chatbot yang terintegrasi.

A. Sumber dan Pra-Pengolahan Data

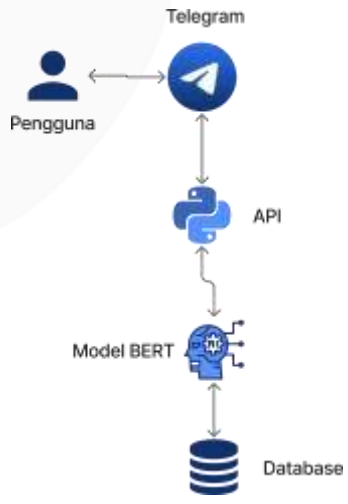
Sumber data utama dalam penelitian ini adalah data mentah yang diperoleh dari Satuan Penjaminan Mutu (SPM) Universitas Telkom Kampus Surabaya, mencakup data Penelitian, Pengabdian, Prestasi, Mahasiswa, dan Dosen dari tahun 2019 hingga 2024. Total data yang dikumpulkan terdiri dari 230 data penelitian, 4.514 data mahasiswa, 147 data pengabdian, 150 data dosen, dan 118 data prestasi. Seluruh data mentah ini kemudian dikonversi ke dalam format JSON, dan dari situ dibangkitkan 83.125 pasang pertanyaan-jawaban yang dibagi menjadi 70% data latih dan 30% data uji.

Data yang terkumpul kemudian melalui tahap pra-pengolahan yang terdiri dari tiga tahapan:

1. Tokenisasi yang merupakan proses memecah teks pertanyaan dan konteks menjadi unit-unit kecil yang disebut token.
2. *Padding/truncation* merupakan teknik untuk menyeragamkan panjang urutan token agar memiliki panjang yang konsisten untuk input model BERT.
3. *Label Encoding* merupakan proses mengubah label atau kategori menjadi format numerik, khususnya untuk menentukan posisi awal dan akhir jawaban dalam bentuk indeks token.

B. Pengembangan dan Arsitektur Sistem Model

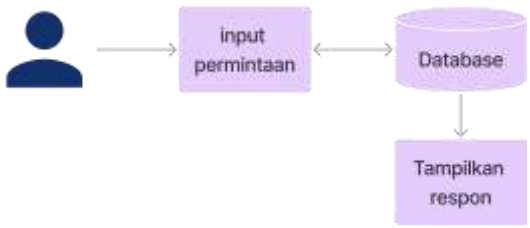
chatbot dikembangkan menggunakan arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) dengan memanfaatkan *library* SimpleTransformers dan model dasar *Rifky/Indobert-QA*.



GAMBAR 1
(ARSITEKTUR SISTEM)

Pada gambar 1 menjelaskan sistem ini diimplementasikan pada platform Telegram. Arsitektur sistem dirancang agar pengguna dapat berinteraksi melalui Telegram, yang

kemudian pesannya diteruskan oleh API ke model BERT untuk diproses seperti yang terlihat pada gambar 1 yang menjelaskan bagaimana alur sistem bekerja. Kemudian model akan menganalisis pertanyaan, mengambil data pendukung dari *database* jika diperlukan, dan mengembalikan jawaban kepada pengguna.



GAMBAR 2
(ALUR KERJA SISTEM)

Untuk alur nya seperti pada gambar 2 yang menjelaskan alur penggunaan sistem dari pengguna menginput permintaan atau menanyakan pertanyaan hingga menghasilkan respon yang sesuai dari keinginan pengguna.

C. Metode Evaluasi Sistem

Evaluasi sistem dilakukan dengan beberapa metode untuk mengukur aspek yang berbeda, yaitu kinerja model, fungsionalitas, dan pengalaman pengguna.

1. Confusion Matrix

Kinerja model dalam melakukan klasifikasi dievaluasi menggunakan *Confusion Matrix*. Metode ini menyajikan tabel yang menggambarkan jumlah prediksi yang benar dan salah yang dikelompokkan berdasarkan label sebenarnya. Dari matriks ini, dihitung beberapa metrik utama untuk mengukur performa model [15], yaitu:

a) Akurasi, mengukur rasio total prediksi yang benar dari keseluruhan data. Untuk menghitung akurasi dapat dilihat pada persamaan dibawah ini:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

b) Presisi, mengukur rasio prediksi positif yang benar dibandingkan dengan total prediksi positif.

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

c) Recall, mengukur rasio prediksi positif yang benar dibandingkan dengan total data aktual yang positif.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

d) F1- Score, menghitung rata-rata harmonik dari presisi dan *recall* untuk memberikan ukuran keseimbangan antara keduanya.

$$F1\ Score = 2 \times \frac{Presisi\ .Recall}{Presisi + Recall} \quad (4)$$

2. Blackbox Testing

Pengujian fungsionalitas sistem dilakukan menggunakan metode *Blackbox Testing*. Metode ini berfokus pada pengujian spesifikasi fungsional dari sistem tanpa memperhatikan struktur internal atau kode programnya. Pengujian dilakukan dengan memberikan berbagai skenario input, seperti pertanyaan relevan, tidak relevan, dan ambigu, untuk memastikan *chatbot* dapat memberikan respons yang tepat dan menjaga alur komunikasi sesuai dengan yang diharapkan pengguna [16], [17].

3. Pengujian Waktu Respons dan Kepuasan Pengguna

Kemudian pada evaluasi sistem dilakukan juga pengukuran waktu respons dari *chatbot* untuk setiap permintaan pengguna dan penyebaran kuesioner untuk mengukur tingkat kepuasan pengguna terhadap kemudahan penggunaan, relevansi jawaban, dan kecepatan sistem.

IV. HASIL DAN PEMBAHASAN

A. Verifikasi dan Validasi

1. Evaluasi Parameter Pelatihan Model BERT

Untuk menentukan konfigurasi model BERT yang paling optimal, dilakukan serangkaian 15 percobaan dengan memvariasikan tiga parameter utama yaitu *max_seq_length* , *n_best_size*, dan *learning_rate*. Tujuan dari optimasi ini adalah untuk menemukan keseimbangan terbaik antara akurasi tinggi dan stabilitas model.

TABEL 1
(PARAMETER YANG DIUJI)

Percobaan	Max_seq_length	n_best_size	Learning_rate
1	224	30	3e-6
2	384	30	3e-6
3	512	30	3e-6
4	150	30	3e-6
5	384	20	3e-6
6	384	40	3e-6
7	384	50	3e-6
8	384	30	3e-7
9	384	30	3e-5
10	384	30	3e-4
11	224	20	3e-7
12	512	40	3e-5
13	150	50	3e-4
14	512	50	3e-7
15	512	20	3e-6

Hasil dari beberapa percobaan yang paling representatif dirangkum dalam Tabel 1 untuk mencari hasil yang terbaik dari 3 kombinasi parameter tersebut untuk pelatihan model dengan total 15 kombinasi yang akan dicari hasil *Confussion Matrix*, *F1- Score, Recall*, dan Presisi.

2. Hasil Pelatihan Model BERT

TABEL 2
(HASIL PELATIHAN BERT)

No	Correct	Similar	Incorrect	Eval loss
1	24063	242	409	-9.844
2	24071	142	501	-10.271
3	24073	74	567	-10.134
4	23642	863	209	-9.524
5	24076	158	480	-10.181
6	24058	268	388	-10.205
7	24071	104	539	-10.061
8	24013	134	567	-10.041
9	24077	37	600	-10.479
10	5376	18776	562	-4.623
11	24034	111	569	-9.777
12	24062	76	576	-10.830
13	4921	12682	7111	-4.504
14	24068	86	560	-10.571
15	24069	90	553	-10.444

Tabel 2 memperlihatkan bahwa performa model tidak dapat dinilai hanya dari satu metrik, melainkan perlu mempertimbangkan keseimbangan antara akurasi dan tingkat kesalahan. Meskipun Percobaan ke-12 mencatat nilai *Evaluation loss* paling rendah (-10.830), konfigurasi pada Percobaan ke-5 justru dipilih sebagai yang paling optimal. Hal ini karena Percobaan ke-5 menunjukkan kombinasi terbaik antara jumlah prediksi yang benar (24.076), serupa (158), dan salah (480), yang mencerminkan stabilitas serta konsistensi pemahaman model secara keseluruhan.

Sebaliknya, konfigurasi pada Percobaan ke-10 dan ke-13 menghasilkan jumlah prediksi benar yang sangat rendah serta prediksi serupa yang tinggi, yang mengindikasikan kemungkinan *overfitting* atau kesalahan dalam pemilihan nilai *learning rate*. Sementara itu, meskipun Percobaan ke-6 dan ke-14 memiliki nilai Eval loss yang baik, tingkat kesalahan (*Incorrect*) yang dicatat masih lebih tinggi dibandingkan Percobaan ke-5. Oleh karena itu, dengan mempertimbangkan seluruh metrik secara menyeluruh, konfigurasi pada Percobaan ke-5 dinilai paling layak untuk digunakan sebagai model akhir dalam implementasi *chatbot*.

3. Evaluasi Hasil Pelatihan

Pada Hasil pelatihan menggunakan *confussion matrix* dalam mengukur seberapa akurat model BERT dalam melatih data untuk *chatbot* yang akan digunakan. Dalam pertimbangan dengan nilai *Precision, Recall,* dan *F1-Score* pada *confussion matrix* tersebut. Dilakukan 15 kali percobaan untuk melihat seberapa konsisten BERT dalam melatih data tersebut dengan parameter yang berbeda. Setelah melewati 15 kali percobaan tersebut, model yang lebih baik akan terpakai dalam pembuatan *chatbot*.

a) Confussion Matrix

TABEL 3
(HASIL CONFUSSION MATRIX, F1-SCORE, PRESISI, DAN RECALL)

Percobaan				
Confussion Matrix				
1		True		False
	Negatif	5568		651
	Positif	18495		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8953	1	0.9448
	1	1	0.9660	0.9827
Confussion Matrix				
2		True		False
	Negatif	5568		643
	Positif	18503		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8965	1	0.9454
	1	1	0.9664	0.9829
Confussion Matrix				
3		True		False
	Negatif	5568		641
	Positif	18505		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8968	1	0.9456
	1	1	0.9665	0.9830
Confussion Matrix				
4		True		False
	Negatif	5568		1061
	Positif	18085		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8399	1	0.9130
	1	1	0.9446	0.9715
Confussion Matrix				
5		True		False
	Negatif	5568		638
	Positif	18508		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8972	1	0.9458
	1	1	0.9667	0.9831
Confussion Matrix				
6		True		False
	Negatif	5568		656
	Positif	18490		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8946	1	0.9444
	1	1	0.9657	0.9826
Confussion Matrix				
7		True		False
	Negatif	5568		642
	Positif	18504		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8966	1	0.9455
	1	1	0.9665	0.9829
Confussion Matrix				
8		True		False
	Negatif	5568		698
	Positif	18448		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8886	1	0.9410
	1	1	0.9635	0.9814
Confussion Matrix				
9		True		False
	Negatif	5568		638
	Positif	18508		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8972	1	0.9458
	1	1	0.9667	0.9831
Confussion Matrix				
10		True		False
	Negatif	5568		19146
	Positif	0		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.2253	1	0.3677
	1	0	0	0
Confussion Matrix				
11		True		False
	Negatif	5568		679
	Positif	18467		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8913	1	0.9425
	1	1	0.9645	0.9819
Confussion Matrix				
12		True		False
	Negatif	5568		652
	Positif	18494		0
	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8952	1	0.9447
	1	1	0.9659	0.9827
Confussion Matrix				
13		True		False
	Negatif	5568		19141
	Positif	5		0
	F1- Score, Recall, dan Presisi			

Percobaan				
		Presisi	Recall	F1-Score
	0	0.2253	1	0.3678
	1	1	0.0003	0.0005
Confussion Matrix				
		True	False	
	Negatif	5568	691	
	Positif	18455	0	
14	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8896	1	0.9416
	1	1	0.9639	0.9816
Confussion Matrix				
		True	False	
	Negatif	5568	646	
	Positif	18500	0	
15	F1- Score, Recall, dan Presisi			
		Presisi	Recall	F1-Score
	0	0.8960	1	0.9452
	1	1	0.9663	0.9828

Menurut tabel 3, metrik dari *Confusion Matrix*, dengan F1-Score untuk kelas positif (pertanyaan yang dapat dijawab) sebagai indikator utama. Hasil evaluasi menunjukkan bahwa sebagian besar konfigurasi parameter memberikan performa yang sangat baik. Secara khusus, Percobaan ke-5 dan ke-9 berhasil mencapai performa puncak dengan F1-Score sebesar 0.9831. Pencapaian ini menunjukkan bahwa model yang dihasilkan sangat akurat dalam memprediksi jawaban yang benar dari konteks yang diberikan.

Sebaliknya, beberapa konfigurasi terbukti tidak efektif, Percobaan ke-10 dan ke-13 menunjukkan performa yang sangat rendah dengan F1-Score di bawah 0.01, yang mengindikasikan kegagalan model akibat pemilihan parameter yang tidak tepat, seperti *learning_rate* yang terlalu besar. Meskipun Percobaan ke-5 dan ke-9 memiliki metrik F1-Score yang identik, Percobaan ke-5 dipilih sebagai konfigurasi terbaik secara keseluruhan. Keputusan ini didasarkan pada analisis lebih lanjut yang menunjukkan bahwa Percobaan ke-5 menghasilkan jumlah prediksi salah (*incorrect*) yang lebih sedikit dibandingkan Percobaan ke-9, sehingga dianggap lebih stabil dan dapat diandalkan untuk implementasi akhir *chatbot*.

b) Waktu Respon

GAMBAR 3
(RESPON TIME)

B Berdasarkan hasil pengujian dan log waktu pada Gambar 3, respons *chatbot* pada pertanyaan pertama cenderung lebih lambat, dengan jeda sekitar 3 detik, dibandingkan pertanyaan-pertanyaan berikutnya yang hanya membutuhkan waktu kurang dari 1 detik. Hal ini disebabkan oleh proses inisialisasi awal *(cold start*) pada pemanggilan pertama model BERT dan SentenceTransformer, di mana komponen seperti *tokenizer*, *encoder*, dan pipeline prediksi masih dalam tahap pemuatan. Setelah inisialisasi selesai, sistem berada dalam kondisi siap (*warm state*), sehingga respons pada pertanyaan selanjutnya menjadi lebih cepat karena semua komponen telah termuat di memori. Selain itu, kestabilan jaringan juga memengaruhi kecepatan respons, di mana koneksi yang lancar akan mempercepat proses pengambilan data dan pengiriman jawaban ke pengguna.

4. Evaluasi Sistem

a) Blackbox Testing

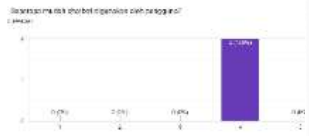
TABEL 4
(BLACKBOX TESTING)

Skenario Pengerjaan	Input	Hasil yang diharapkan	Hasil Pengujian	Status
Input tanpa kode khusus	/start skripsi133	Kode akses salah.	Kode akses salah.	Berhasil
Input dengan kode khusus	/start skripsi123	Akses diberikan! Silakan pilih kategori:	Akses diberikan! Silakan pilih kategori:	Berhasil
Pemberian bantuan /help	/help	Cara Penggunaan Chatbot	Cara Penggunaan Chatbot	Berhasil
Menanyakan data penelitian	ketua penelitian "antenna mimo 5g-nr band n77"	muhsin	muhsin	Berhasil
Menanyakan data pengabdian	detail pengabdian "implementasi 60 unit plts pada rsud oksibil"	kegiatan ini merupakan bagian dari skema pengmas kerjasama yang diselenggarakan pada tahun 2023 - 2024. berdasarkan surat keputusan (sk) nomor rek.259_abdi1_rek_xii_2027 dan tercatat pada nomor kontrak belum diketahui, kegiatan ini mendapatkan dukungan pendanaan sebesar 400000000 yang bersumber dari eksternal	kegiatan ini merupakan bagian dari skema pengmas kerjasama yang diselenggarakan pada tahun 2023 - 2024. berdasarkan surat keputusan (sk) nomor rek.259_abdi1_rek_xii_2027 dan tercatat pada nomor kontrak belum diketahui, kegiatan ini mendapatkan dukungan pendanaan sebesar 400000000 yang bersumber dari eksternal	Berhasil

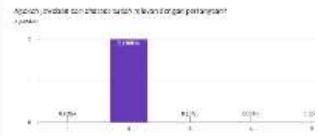
Menanyakan data prestasi	juara logo design competition satu data unair pada 2023	mahasiswa prodi teknologi informasi	mahasiswa prodi teknologi informasi	Berhasil
Menanyakan data dosen	nip mastuty ayu ningtyas	22970009	20950049	Belum Sesuai
Menanyakan data mahasiswa	program studi m.baihaqi ilmi	s1 teknologi informasi - kampus surabaya	s1 teknologi informasi - kampus surabaya	Berhasil

Pada tabel 4 *Blackbox* dilakukan untuk mengevaluasi sistem dari perspektif pengguna tanpa melihat struktur kode internal. Pengujian mencakup berbagai skenario seperti validasi akses, permintaan bantuan, dan kueri data spesifik untuk semua kategori. Hasil pengujian menunjukkan bahwa sebagian besar skenario yang diuji berstatus "Berhasil", yang mengindikasikan sistem berfungsi sesuai spesifikasi. Namun, ditemukan satu kegagalan signifikan pada skenario permintaan data dosen untuk "nip mastuty ayu ningtyas", di mana sistem memberikan hasil "20950049" padahal yang diharapkan adalah "22970009". Kegagalan ini disebabkan oleh akurasi model yang belum stabil untuk pertanyaan spesifik tersebut, yang menunjukkan perlunya penambahan variasi data pelatihan.

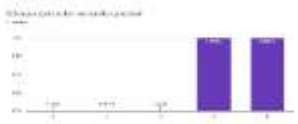
b) Hasil Kuesioner



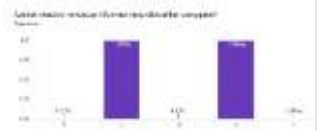
GAMBAR 4
(PERTANYAAN KUESIONER 1)



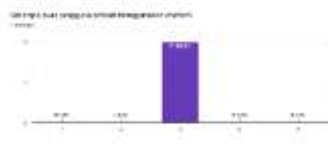
GAMBAR 5
(PERTANYAAN KUESIONER 2)



GAMBAR 6
(PERTANYAAN KUESIONER 3)



GAMBAR 7
(PERTANYAAN KUISIONER 4)



GAMBAR 7
(PERTANYAAN KUESIONER 5)

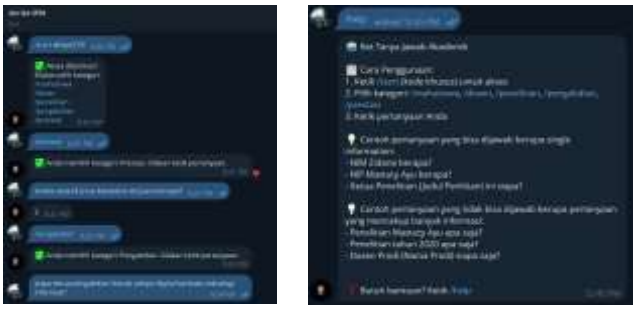
Berdasarkan hasil kuesioner evaluasi pengujian sistem yang diisi oleh pihak SPM pada Gambar 4 sampai 8, dapat dianalisis beberapa aspek penting terkait kinerja *chatbot*. Pada pertanyaan mengenai cakupan informasi yang dibutuhkan pengguna, sebanyak 50% responden memberikan nilai 2 dan 50% memberikan nilai 4. Distribusi jawaban ini mengindikasikan adanya variasi dalam persepsi pengguna terhadap kelengkapan informasi yang disediakan oleh chatbot, di mana sebagian merasa informasi cukup memadai sementara sebagian lain melihat adanya ruang untuk peningkatan. Selanjutnya, tingkat kepuasan pengguna setelah menggunakan *chatbot* menunjukkan hasil yang sangat positif, dengan 100% responden menyatakan puas (nilai 3), menunjukkan bahwa secara keseluruhan pengalaman pengguna berada pada tingkat yang diharapkan atau bahkan melampaui ekspektasi.

Hasil evaluasi menunjukkan bahwa *chatbot* dinilai sangat mudah digunakan, dengan 100% responden memberikan nilai 4 pada aspek kemudahan penggunaan. Untuk aspek relevansi jawaban, seluruh responden memberikan nilai 2, mengindikasikan bahwa jawaban yang diberikan cukup relevan, meskipun masih memerlukan peningkatan agar lebih akurat dan lengkap. Sementara itu, kecepatan respons *chatbot* juga mendapat penilaian positif, dengan 50% responden memberikan nilai 4 dan 50% lainnya nilai 5, menunjukkan bahwa waktu respons dianggap cepat dan efisien. Secara keseluruhan, *chatbot* menunjukkan performa yang baik dari segi kemudahan penggunaan dan kecepatan, namun masih perlu penyempurnaan pada aspek isi jawaban agar lebih memenuhi harapan pengguna.

5. Hasil

Dalam penerapan *chatbot* tersebut dalam Telegram yang melalui beberapa proses pengolahan data dan pelatihan model, *chatbot* tersebut membagi beberapa kategori untuk pertanyaan yang lebih spesifik agar mempermudah *chatbot* memahami pertanyaan dari pengguna dan menerapkan kode khusus untuk penggunaan *chatbot* dalam Telegram agar tidak bisa diakses oleh semua orang, Berikut pada Gambar 9 sampai 10 adalah hasil dari beberapa pertanyaan untuk uji coba.

GAMBAR 8
(HASIL CHATBOT 1)



GAMBAR 9
(HASIL CHATBOT 2)

Pada Gambar 9 sampai 10, terdapat percobaan untuk menanyakan beberapa hal pada chatbot. Langkah awal untuk menanyakan *chatbot* adalah mengetikkan /*start* dan kode khusus yang diberikan nanti agar bisa diakses oleh tertentu. Pada tahap selanjutnya pengguna akan memilih kategori pertanyaan apa yang ditanyakan, hal ini dapat mempermudah *chatbot* karena sudah terkategori pertanyaan tersebut kearah mana. Namun *chatbot* hanya bisa menanyakan informasi individu saja dan beberapa pertanyaan. Dalam hasil tersebut *chatbot* mampu menjawab pertanyaan oleh pengguna dengan singkat. Pada Gambar 10 terdapat panduan chatbot yang bisa dilakukan pada chatbot tersebut terdapat batasan dalam chatbot beritu pertanyaan yang bisa ditanyakan tersebut.

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan teknologi *chatbot* berbasis deep learning dengan menggunakan arsitektur BERT untuk meningkatkan efisiensi layanan Satuan Penjaminan Mutu (SPM) di lingkungan perguruan tinggi. Sistem yang dibangun mampu memahami konteks pertanyaan dalam bahasa alami dan memberikan respons yang relevan melalui integrasi dengan platform Telegram. Melalui proses pelatihan model dan evaluasi mendalam menggunakan metrik seperti *confusion matrix*, *precision*, *recall*, dan F1-score, diperoleh hasil bahwa model memiliki akurasi tinggi dan tingkat kesalahan yang cukup rendah. Penambahan struktur embedding berdasarkan kategori pertanyaan seperti mahasiswa, dosen, penelitian, pengabdian, dan prestasi turut meningkatkan relevansi jawaban yang diberikan *chatbot*. Hal ini membuktikan bahwa penerapan NLP dan model BERT dalam layanan informasi akademik memberikan kontribusi signifikan terhadap kemudahan akses data dan kualitas pelayanan digital kampus.

Secara keseluruhan, penelitian ini berhasil menjawab rumusan masalah terkait bagaimana mengimplementasikan *chatbot* berbasis *deep learning* untuk layanan SPM. Hasil pengujian menunjukkan bahwa sistem memiliki performa yang stabil dan akurat dalam menjawab berbagai jenis pertanyaan pengguna. *Chatbot* yang dikembangkan mampu beradaptasi dengan pertanyaan yang cukup bervariasi dan memberikan jawaban berdasarkan konteks yang relevan secara cepat. Namun terdapat beberapa pertanyaan yang multi informasi tidak dapat terjawab dikarenakan terbatasnya variasi pertanyaan yang terlatih dalam model BERT. Dengan demikian, solusi ini tidak hanya memberikan efisiensi dalam pencarian data, tetapi juga menjadi inovasi layanan akademik yang dapat diterapkan lebih luas ke unit dan program studi lainnya. Ke depannya, pengembangan lebih lanjut dapat difokuskan pada peningkatan variasi data pelatihan serta penerapan teknologi ini dalam sistem informasi kampus yang terintegrasi secara menyeluruh.

REFERENSI

[1] J. Byun, B. Kim, K. A. Cha, and E. Lee, “Design and Implementation of an Interactive Question-Answering System with Retrieval-Augmented Generation for Personalized Databases,” *Applied Sciences (Switzerland)*, vol. 14, no. 17, 2024, doi: 10.3390/app14177995.

[2] C. V. Misischia, F. Poecze, and C. Strauss, “Chatbots in customer service: Their relevance and impact on service quality,” *Procedia Comput Sci*, vol. 201, no. C, pp. 421–428, 2022, doi: 10.1016/j.procs.2022.03.055.

[3] M. P. Naresh, S. Venkataramana, N. Pavan, A. R. Mohammed, M. Tharun, and N. Chanti, “Implementing a Flask-based Chatbot for College Enquiries using Spacy and TensorFlow,” vol. 14, no. 05, pp. 77–82, 2023.

[4] S. Pokhrel, “A Systematic Review of Chatbot Model and chatbot technologies,” *Αγαη*, vol. 15, no. 1, pp. 37–48, 2024, [Online]. Available: https://www.ieeesem.com/researchpaper/A_Systematic_Literature_Review_and_Existing_Challenges_of_Chatbot_Models.pdf

[5] S. Sharma and P. Chaudhary, “Machine learning and deep learning,” *Quantum Computing and Artificial Intelligence: Training Machine and Deep Learning Algorithms on Quantum Computers*, pp. 71–84, 2023, doi: 10.1515/9783110791402-004.

[6] I. H. Sarker, “Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions,” *SN Comput Sci*, vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/s42979-021-00815-1.

[7] D. Dharrao and S. Gite, “TherapyBot: a chatbot for mental well-being using transformers,” *International Journal of Advances in Applied Sciences*, vol. 13, no. 1, pp. 1–12, 2024, doi: 10.11591/ijaas.v13.i1.pp1-12.

[8] Q. Zhou *et al.*, “GastroBot: a Chinese gastrointestinal disease chatbot based on the retrieval-augmented generation,” *Front Med (Lausanne)*, vol. 11, no. May, 2024, doi: 10.3389/fmed.2024.1392555.

[9] T. A. Zuraiyah, D. K. Utami, and D. Herlambang, “Implementasi Chatbot Pada Pendaftaran Mahasiswa Baru Menggunakan Recurrent Neural Network,” *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 24, no. 2, pp. 91–101, 2019, doi: 10.35760/tr.2019.v24i2.2388.

[10] A. Babu and S. B. Boddu, “BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding,” *Exploratory Research in Clinical and Social Pharmacy*, vol. 13, no. January, p. 100419, 2024, doi: 10.1016/j.rcsop.2024.100419.

[11] E. Adamopoulou and L. Moussiades, *An Overview of Chatbot Technology*, vol. 584 IFIP, no. June. Springer International Publishing, 2020. doi: 10.1007/978-3-030-49186-4_31.

[12] S. Ayanouz, B. A. Abdelhakim, and M. Benhmed, "A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance," *ACM International Conference Proceeding Series*, no. April, 2020, doi: 10.1145/3386723.3387897.

[13] C. S. Patil, "NLP Assisted Text Annotation," *Interantional Journal of Scientific Research in Engineering and Management*, vol. 06, no. 06, pp. 1–10, 2022, doi: 10.55041/ijsrem14651.

[14] J. A. Alzubi, R. Jain, A. Singh, P. Parwekar, and M. Gupta, "COBERT: COVID-19 Question Answering System Using BERT," *Arab J Sci Eng*, vol. 48, no. 8, pp. 11003–11013, 2023, doi: 10.1007/s13369-021-05810-5.

[15] M. V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021, [Online]. Available: http://arxiv.org/abs/2103.11943

[16] Supriyono, "Software Testing with the approach of Blackbox Testing on the Academic Information System," *International Journal of Information System & Technology*, vol. 3, no. 36, pp. 227–235, 2020.

[17] J. Shadiq, A. Safei, and R. W. R. Loly, "Pengujian Aplikasi Peminjaman Kendaraan Operasional Kantor Menggunakan BlackBox Testing," *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS : Journal of Information Management*, vol. 5, no. 2, p. 97, 2021, doi: 10.51211/imbi.v5i2.1561.