



**IMPLEMENTASI INDOBERT DAN DIALOGUE STATE TRACKING  
PADA CHATBOT LAYANAN INFORMASI PENYEDIA KURSUS**

**LAPORAN TUGAS AKHIR**

**DHIAZ RUSYDA NAFSYI  
41521010163**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2025**



**IMPLEMENTASI INDOBERT DAN DIALOGUE STATE TRACKING  
PADA CHATBOT LAYANAN INFORMASI PENYEDIA KURSUS**

**DHIAZ RUSYDA NAFSYI  
41521010163**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2025**

## **HALAMAN PERNYATAAN KARYA SENDIRI**

### **HALAMAN PERNYATAAN KARYA SENDIRI**

Saya yang bertanda tangan dibawah ini :

Nama : Dhiaz Rusyda Nafsyi  
NIM : 41521010163  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam laporan skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 11 Juli 2025



Dhiaz Rusyda Nafsyi

## HALAMAN PENGESAHAN

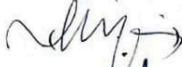
### HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Dhiaza Rusyda Nafsyi  
NIM : 41521010163  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Implementasi IndoBERT dan Dialogue State Tracking Pada Chatbot Layanan Informasi Penyedia Kursus

Telah berhasil dipertahankan pada sidang di hadapan Dewan Pengaji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

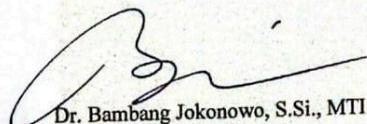
Disahkan oleh:

Pembimbing : Wawan Gunawan, S.Kom., M.T.,  
M.Kom.  
()  
NIDN : 0424108104  
Ketua Pengaji : Dr. Afiyati, S.Si., M.T.,  
()  
NIDN : 0316106908  
Pengaji 1 : Dr. Ir. Eliyani.  
()  
NIDN : 0321026901  
Pengaji 2 : Dr. Harwikarya, M.T  
()  
NIDK : 8969090024

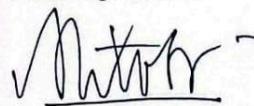
Jakarta, 21 Juli 2025

Mengetahui,

Dekan

  
Dr. Bambang Jokonowo, S.Si., MTI  
NIDN : 0320037002

Ketua Program Studi

  
Dr. Hadi Santoso, S.Kom., M.Kom  
NIDN : 0225067701

## KATA PENGANTAR

Puji syukur kehadirat Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan tugas akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
4. Bapak Wawan Gunawan, S.Kom., MT., M.Kom. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mensuport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana
6. Kepada Kakak saya, yang selalu memberikan semangat, dan nasihat di setiap langkah perjalanan saya menyelesaikan skripsi ini. Terima kasih telah menjadi tempat berbagi cerita, dan teman diskusi yang tak pernah lelah mendengarkan keluh kesah saya.
7. Semua teman kuliah yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.
8. Saya juga ingin menyampaikan terima kasih yang sebesar-besarnya kepada teman-teman online saya, Toru, Ian, dan Cece. Terima kasih telah menjadi teman berbagi cerita, semangat, serta selalu mendukung dan

menemani saya di saat suka maupun duka selama proses penyusunan skripsi ini.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencerahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 11 Juli 2025



Dhiaz Rusyda Nafsyi

## **HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

### **HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai Sivitas akademik Universitas Mercu Buana, Saya yang bertanda tangan dibawah ini :

Nama : Dhiaz Rusyda Nafsyi  
NIM : 41521010163  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus

Demi Pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian Pernyataan ini saya buat dengan sebenarnya.

Jakarta, 11 Juli 2025

Yang Menyatakan,



Dhiaz Rusyda Nafsyi

## ABSTRAK

Nama	:	Dhiaz Rusyda Nafsyi
NIM	:	41521010163
Program Studi	:	Teknik Informatika
Judul Laporan Skripsi	:	Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus
Dosen Pembimbing	:	Wawan Gunawan., S. Kom, MT., M. Kom

Sistem layanan informasi merupakan aspek penting bagi semua Instansi, terutama instansi pendidikan seperti lembaga penyedia kursus anak-anak hingga remaja untuk penyebaran informasi mereka. Di beberapa tempat penyedia kursus, Sistem layanan informasi masih dilakukan secara manual menggunakan WhatsApp atau akun Sosial Media dengan cara membalas komentar maupun pesan yang masuk. Tantangan yang harus dihadapi adalah respon lamban dan waktu yang terikat dengan jam kerja. Hal tersebut dapat meningkatkan rasa ketidakpuasan pengguna bila pertanyaan dan permasalahan mereka tidak terjawab. Maka dari itu, solusi untuk permasalahan tersebut adalah menyediakan dan membangun Chatbot. Chatbot adalah program komputer yang dirancang untuk membuka percakapan dengan manusia baik melalui teks maupun suara secara instan dan cepat. Penelitian ini akan membahas Implementasi IndoBERT pada chatbot layanan penyedia kursus untuk membantu menjawab pertanyaan pengguna secara umum seperti pertanyaan mengenai pendaftaran, biaya, metode pembayaran ataupun pengenalan lebih dalam terkait program kursus yang ada. Adanya Dialogue State Tracking juga dapat membantu Chatbot untuk mengingat status percakapan pengguna dalam bertanya dan menyimpan informasi dengan baik pada label ‘biaya’. Dataset sejumlah 1368 akan dilatih dengan model IndoBERT dan menghasilkan akurasi sebesar 89% untuk data latih dan data uji. Hal tersebut membuktikan bahwa Implementasi IndoBERT pada chatbot penyedia layanan kursus dapat digunakan dan Chatbot berhasil memahami konteks pertanyaan berbahasa Indonesia terkait seputar layanan kursus.

**Kata kunci:** IndoBERT, Chatbot, Dialogue State Tracking (DST), Natural Language Processing (NLP)

## ABSTRACT

Nama	:	Dhiaz Rusyda Nafsyi
NIM	:	41521010163
Program Studi	:	Teknik Informatika
Judul Laporan Skripsi	:	Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus
Dosen Pembimbing	:	Wawan Gunawan, S.Kom., MT., M. Kom

*Information service systems are an important aspect for all institutions, especially educational institutions such as providers of courses for children and teenagers for the dissemination of their information. In some course providers, information service systems are still carried out manually using WhatsApp or social media accounts by replying to comments or messages that come in. This has several drawbacks, including slow response times and increased user dissatisfaction if their questions are not answered. Therefore, the solution to this problem is to provide and develop a chatbot. A chatbot is a computer program designed to initiate conversations with humans via text or voice instantly and quickly. This study will discuss the implementation of IndoBERT in a course provider service chatbot to assist in answering general user questions, such as inquiries about registration, fees, payment methods, or further information about available course programs. The inclusion of Dialogue State Tracking can also help the chatbot remember the user's conversation status and store information effectively. A dataset of 1,368 instances will be trained using the IndoBERT model, achieving an accuracy rate of 89% for both training and testing data. This demonstrates that the implementation of IndoBERT in the course provider chatbot is feasible, and the chatbot successfully understands the context of Indonesian-language questions related to course services.*

**Kata kunci:** *IndoBERT, Chatbot, Dialogue State Tracking (DST), Natural Language Processing (NLP).*

## DAFTAR ISI

<b>HALAMAN PERNYATAAN KARYA SENDIRI</b>	<b>2</b>
<b>HALAMAN PENGESAHAN</b>	<b>3</b>
<b>KATA PENGANTAR</b>	<b>4</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS</b>	<b>5</b>
<b>ABSTRAK</b>	<b>6</b>
<b>ABSTRACT</b>	<b>7</b>
<b>DAFTAR ISI</b>	<b>8</b>
<b>DAFTAR TABEL</b>	<b>10</b>
<b>DAFTAR GAMBAR</b>	<b>11</b>
<b>BAB I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	4
<b>BAB II TINJAUAN PUSTAKA</b>	<b>5</b>
2.1 Teori Utama	5
2.1.1 NLP ( <i>Natural Language Processing</i> )	5
2.1.2 BERT ( <i>Bidirectional Encoder Representations of Transformer</i> )	5
2.1.3 IndoBERT ( <i>Indonesian Bidirectional Encoder Representations of Transformer</i> )	8
2.1.4 DST ( <i>Dialogue State Tracking</i> )	9
2.2 Penelitian Terdahulu	10
2.3 Teori Pendukung	18
2.3.1 Chatbot	18
2.3.2 Bot Telegram (Bot Father)	19
<b>BAB III METODE PENELITIAN</b>	<b>20</b>
3.1 Jenis Penelitian	20
3.2 Tahapan Penelitian	20
3.3 Instrumen Penelitian	24
3.4 Subjek Penelitian	25

<b>BAB IV PEMBAHASAN</b>	<b>26</b>
4.1    Persiapan Data	26
4.2    Pre-Processing Data	29
4.3    Split Data	30
4.4    Proses Training, Fine Tuning dan Testing Model IndoBERT	31
4.5    Proses Persiapan Deployment	38
4.6    Dialogue State Tracking	39
4.7    Deployment Bot Telegram	40
4.8    Testing Chatbot di Telegram Bot	41
4.9    Kuisisioner	48
<b>BAB V KESIMPULAN DAN SARAN</b>	<b>62</b>
5.1    Kesimpulan	62
5.2    Saran	62
<b>DAFTAR PUSTAKA</b>	<b>63</b>
<b>LAMPIRAN</b>	<b>65</b>

## **DAFTAR TABEL**

Tabel 2. 1 Penelitian Terdahulu	10
Tabel 2. 2 Gap Penelitian	19
Tabel 3. 1 Instrumen Penelitian Perangkat Lunak (Software)	22
Tabel 3. 2 Instrumen Penelitian Perangkat Keras (Hardware)	22
Tabel 4. 1 Keterangan WordCount	28
Tabel 4. 2 Sebagian contoh Dataset yang digunakan	29
Tabel 4. 3 Keterangan Intent	30
Tabel 4. 4 Tahap Case Folding	32
Tabel 4. 5 Tahap Penghapusan Simbol atau Tanda baca	32
Tabel 4. 6 Tahap Normalisasi	32
Tabel 4. 7 Contoh Vocab Normalisasi	33
Tabel 4. 8 Daftar Pertanyaan Kuisioner	53
Tabel 4. 9 Tabel Keterangan Skala Kuisioner	54
Tabel 4. 10 Sebaran Jawaban Responden pada Kuisioner Testing Chatbot	54
Tabel 4. 11 Tabel Keterangan Persentase Tiap Pertanyaan	56

## **DAFTAR GAMBAR**

Gambar 2. 1 Ilustrasi Proses MLM	6
Gambar 2. 2 Ilustrasi Proses NSP	7
Gambar 2. 3 Ilustrasi Proses Dialogue State Tracking	8
Gambar 3. 1 Tahapan Penelitian	20
Gambar 4. 1 Hasil Training menggunakan Epoch	35
Gambar 4. 2 Grafik pelatihan Epoch	35
Gambar 4. 3 Hasil Training	36
Gambar 4. 4 Grafik hasil training	36
Gambar 4. 5 Hasil Testing Model IndoBERT	37
Gambar 4. 6 Grafik hasil testing	38
Gambar 4. 7 Pembuatan Bot di Telegram Menggunakan Bot Father	41
Gambar 4. 8 Tombol awal untuk memulai Chat	42
Gambar 4. 9 Respon Chatbot terhadap pertanyaan pendaftaran	43
Gambar 4. 10 Respon Chatbot terhadap pertanyaan terkait Usia	44
Gambar 4. 11 Chatbot menanggapi pertanyaan terkait biaya secara umum	44
Gambar 4. 12 Chatbot merespon pertanyaan terkait biaya coding dengan bubblechat terpisah.	45
Gambar 4. 13 Chatbot dapat menjawab pertanyaan diluar konteks kursus	46
Gambar 4. 14 Chatbot salah menanggapi Pertanyaan	46
Gambar 4. 15 Chatbot merespon pertanyaan diluar konteks dengan benar.	48
Gambar 4. 16 Hasil Rangkuman Grafik Persentase Kuisioner	52

## **DAFTAR LAMPIRAN**

Lampiran 1 Detail Soal dan Jawaban Responden pada Kuisioner Testing Chatbot	
59	
Lampiran 2 Kartu Asistensi	69
Lampiran 3 Curiculum Vitae	70
Lampiran 4 Surat Pernyataan HAKI	71
Lampiran 5 Sertifikat BNSP	73
Lampiran 6 Surat Ijin Riset Perusahaan	74
Lampiran 7 Hasil Cek Turnitin	75

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Sistem layanan informasi merupakan salah satu aspek penting dalam operasional sebuah institusi, termasuk institusi pendidikan dalam menyampaikan informasi akademik pada pengguna seperti orang tua, wali murid, siswa maupun calon siswa (Patmawati, 2021). Sistem layanan informasi yang sebelumnya dilakukan secara manual seperti mendatangi institusi untuk meraih informasi, kini telah berubah sesuai dengan perkembangan teknologi. Pengguna dapat mengakses kontak maupun akun media sosial atau *website* resmi lembaga penyedia kursus secara fleksibel, dimanapun dan kapanpun. Hal tersebut dapat memudahkan pengguna untuk mengakses informasi dengan cepat (Zuniananta, 2021).

Kualitas pada sistem layanan informasi digital juga perlu dipertimbangkan dalam menciptakan pengalaman yang baik bagi pengguna. Layanan Informasi harus memiliki sistem yang mendukung keandalan, yakni memperlihatkan sejauh mana sistem dapat berjalan konsisten dan teratur tanpa gangguan. Keandalan pada layanan informasi cukup krusial dan penting karena dapat membangun kepercayaan dan memenuhi harapan pengguna, dengan demikian dapat meningkatkan loyalitas atau kesetiaan pengguna terhadap institusi terkait (Indrayani, 2024). Keandalan sistem juga mencakup kemampuan untuk memahami permasalahan yang dihadapi pengguna serta memberikan solusi nyata secara responsif. Responsif dalam artian adalah dapat menanggapi permintaan pengguna dengan cepat dan efektif (Indrayani, 2024). Namun, dalam praktiknya, masih ada beberapa penyedia kursus yang memiliki layanan informasi terbatas dan berpaku pada jam kerja (Patmawati, 2021) .

Akademi adalah tempat yang menyediakan Program kursus untuk anak-anak hingga remaja. Kursus yang disediakan antara lain adalah Kursus Coding, Kursus Design dan Kursus Inggris – Matematika. Namun, Operasional Akademi tersebut masih menggunakan sistem layanan secara manual lewat aplikasi WhatsApp dan Sosial Media, serta memiliki keterbatasan waktu yakni kurang dari 24 jam atau menyesuaikan pada jam kerja. Sistem layanan informasi digital yang bergerak secara manual maupun konvensional yang tidak memenuhi kebutuhan pengguna dan tidak memenuhi kualitas, menyebabkan respons yang

lambat dan tidak menjawab pertanyaan maupun kesulitan dari pengguna. Hal tersebut dapat meningkatkan kesalahpahaman informasi yang dibutuhkan, serta ketidakpuasan pelanggan dalam menggunakan layanan karena ketersediaan informasi yang terbatas. (Erlina et al., 2023).

Kecerdasan buatan atau *Artificial Intelligent* (AI), merupakan salah satu bukti bahwa teknologi telah berkembang cukup pesat di era modern ini. Adanya AI atau kecerdasan buatan memungkinkan mesin berpikir selayaknya seperti manusia dan membuat keputusan secara mandiri, salah satunya adalah teknologi *chatbot*. *Chatbot* merupakan aplikasi percakapan yang dirancang untuk mengatasi dan menerima masukan dari pengguna serta dapat memberikan balasan atau respon yang sesuai berdasarkan masukan tersebut. Hasil dari respon yang diberikan *Chatbot* akan dikirim kembali pada Pengguna, sehingga Pengguna mendapatkan tanggapan dan informasi yang dibutuhkan (Dwiyono et al., 2024).

Berdasarkan Penelitian terdahulu, yakni (Wiharto, 2024), dengan skripsi berjudul “Sistem Chatbot Sebagai Layanan Tugas Akhir Program Studi Teknik Informatika Unissula Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT)” menggunakan model BERT dalam pengembangan Chatbot nya untuk membantu Mahasiswa dalam menangani pertanyaan mengenai Tugas Akhir. Output yang dihasilkan adalah akurasi *chatbot* sebesar 99% serta melakukan *deployment* ke *website* lokal. Hasil menyimpulkan bahwa Chatbot dengan BERT meraih hasil maksimal dan dapat mengerti bahasa dengan baik.

Menurut Penelitian yang dilakukan oleh (Maulana et al., 2025), dengan judul “Pembuatan Chatbot pada Website Persatuan Gizi-Kabupaten Karawang menggunakan IndoBERT” juga meraih hasil yang maksimal dan memuaskan. Model IndoBERT yang diterapkan pada Chatbot, bertujuan untuk memudahkan penyebaran informasi dengan akurasi sebesar 87,67%. Hasil ini membuktikan bahwa metode IndoBERT cukup efektif dalam membangun chatbot untuk memenuhi kebutuhan informasi.

Sehingga, Solusi yang dapat digunakan dalam permasalahan layanan informasi pada Akademi penyedia program kursus adalah *Chatbot*. Chatbot juga merupakan salah satu implementasi teknologi NLP (*Natural Language*

*Processing*) yang dapat digunakan untuk membantu instansi dalam menjaga kualitas layanan terhadap pelanggan. *Chatbot* dengan teknologi NLP dirancang untuk menangani dan menjawab pertanyaan pelanggan secara instan. Sebagai contoh, penggunaan chatbot berbasis NLP adalah ketika pengguna bertanya mengenai sistem kursus di Akademi. Pertanyaan seperti “Kursus apa saja yang ada di Akademi?”, *chatbot* akan memahami bahwa pengguna ingin mengetahui program apa saja yang ada dan disediakan oleh Akademi (Tarumingkeng, 2024).

## **1.2 Perumusan Masalah**

1. Bagaimana IndoBERT dapat melakukan klasifikasi intent dan memberikan respon yang tepat?
2. Bagaimana IndoBERT membuat batasan pada pertanyaan *out of scope*?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah mengembangkan chatbot dengan implementasi IndoBERT sebagai sarana atau media dalam menyampaikan informasi kursus di Akademi, serta mendukung penyebaran informasi yang lebih mudah, cepat dan fleksibel.

## **1.4 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan solusi dan kemudahan bagi para pengguna, terutama orangtua dan wali murid serta calon siswa yang akan mendaftar dalam meraih informasi terkait kursus yang disediakan oleh Akademi. Sistem chatbot yang dibuat juga dapat membantu meringankan tugas Admin dalam menjawab pertanyaan berulang secara efisien dan meningkatkan kualitas pelayanan informasi secara keseluruhan.

## **1.5 Batasan Masalah**

1. Dataset Pertanyaan serta jawaban yang tersedia adalah menggunakan Bahasa Indonesia.
2. *Chatbot* yang dibuat, tidak menerima input berupa angka. Sehingga pengguna menulis pertanyaan sendiri.
3. *Chatbot* dirancang untuk menerima percakapan seputar informasi umum terkait kursus di Akademi, seperti aplikasi yang digunakan selama kursus, biaya kursus, kontak admin dan lokasi, informasi teori terkait coding dan desain, informasi kursus, pendaftaran, metode pembayaran, pembelajaran dan spesifikasi laptop yang digunakan.
4. Peneliti berasumsi bahwa seluruh pengguna memiliki akun Telegram untuk mengakses sumber informasi lewat *Chatbot* Akademik di Telegram.
5. *Chatbot* dirancang dengan rule-based chatbot, bukan generative (membuat kalimat jawaban sendiri).

## BAB II TINJAUAN PUSTAKA

### 2.1 Teori Utama

#### 2.1.1 NLP (*Natural Language Processing*)

*Natural Language Processing* (NLP) adalah salah satu cabang dari Kecerdasan Buatan (*Artificial Intelligent*), yang berfokus pada proses komputer untuk memahami dan berinteraksi dengan bahasa manusia. NLP sangat memiliki peran penting karena proses tersebut dapat membuat komputer menganalisis dan menghasilkan bahasa alami manusia baik dalam bentuk suara maupun teks. Teknologi NLP bisa menjadi dasar dari beragam aplikasi yang sering kali dipakai dalam kehidupan sehari-hari, sebagai contoh adalah asisten virtual (*Siri, Google Assistant*), *chatbot* layanan, penerjemah bahasa otomatis dan penggunaan untuk analisis sentimen guna mengetahui komentar negatif dan positif di media sosial.

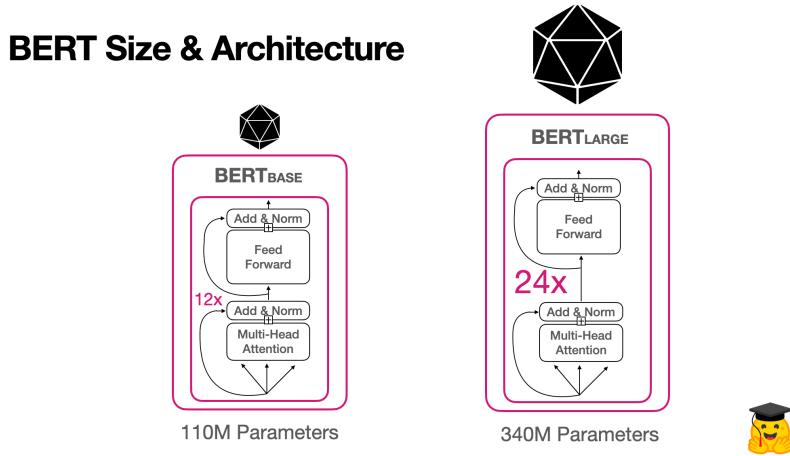
Kemajuan teknologi NLP didukung oleh hadirnya arsitektur dari Transformer, yang digunakan pada model BERT (*Bidirectional Encoder Representations From Transformers*) dan GPT (*Generative Pre-Trained Transformer*). Model seperti BERT dan GPT sangat memungkinkan komputer untuk memahami bahasa manusia yang lebih kompleks dan luas, menggunakan teknik contextual embeddings. Adanya model dan teknik yang dilakukan akan membantu komputer untuk memahami makna kata per kata dengan kalimat yang lebih besar untuk meningkatkan akurasi (Tarumingkeng, 2024).

#### 2.1.2 BERT (*Bidirectional Encoder Representations of Transformer*)

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model yang digunakan untuk pemrosesan bahasa alami berbasis arsitektur Transformers. BERT memiliki dua tahap dalam pembelajaran bahasanya, yaitu *pre-training* dan *fine-tuning*. *Pre-training* adalah model BERT yang digunakan untuk memahami bahasa secara umum dan belum diberi label. Sedangkan *fine-tuning* adalah model BERT yang dilatih ulang dari *pre-training* lalu dilatih dengan data yang sudah diberi label untuk tugas tertentu, seperti klasifikasi kalimat atau sentimen, dan *questions answering* (menjawab pertanyaan) (Das & Pedersen, 2024). BERT menggunakan *Transfer Learning*, yaitu melatih sebuah model secara umum agar mengetahui dan mengenali struktur maupun pola bahasa secara luas (*pre-training*) lalu pembelajaran selanjutnya menggunakan model yang

sudah dilatih untuk tugas spesifik (*fine-tuning*), sehingga disimpulkan bahwa *transfer learning* merupakan perpindahan pembelajaran model dari *pre-training* (pembelajaran secara umum) menuju *fine-tuning* (pembelajaran lebih lanjut untuk tugas spesifik) (Sahmoud & Mikki, 2022).

BERT terbagi menjadi beberapa jenis sesuai dengan ukurannya, yaitu BERT-base dan BERT-large. Hal yang membedakan diantara keduanya adalah jumlah layer encoder yang berbeda. BERT-base memiliki 12 layer encoder, dan BERT-large memiliki 24 layer encoder. Arsitektur BERT dapat digambarkan seperti gambar dibawah ini.



Gambar 2. 1 Arsitektur BERT

(Sumber : Muller, 2022)

Proses BERT dimulai dengan input user berupa teks, kemudian di tokenisasi menggunakan wordpiece tokenizer seperti kata “makanannya” akan dipecah beberapa bagian menjadi sub-kata seperti “makanan” dan “##nya”. Setelah tokenisasi berhasil, input yang sudah berubah menjadi token akan diolah menjadi vektor embedding. Proses selanjutnya adalah input akan masuk ke tiga jenis layer berbeda pada step Multi-Head Attention, yaitu layer Q (Query), K (Key), dan V (Value). Menurut Penelitian (Inuwa, 2025), Perhitungan pada Multi-head Attention dapat menggunakan rumus pada persamaan 1, sebagai berikut :

$$\text{Multihead}(Q, K, V) = \text{Concat}\left(\text{Head}_1, \dots, \text{Head}_h\right)W^O \quad (1)$$

$$\text{Where } \text{Head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Keterangan pada persamaan (1) :

$Q$  : Query ()

$K$  : Key ()

$V$  : Value ()

$QW_i^Q$  : Matriks Bobot Query pada Head ke - i

$KW_i^K$  : Matriks bobot key pada head ke - i

$VW_i^V$  : Matriks bobot value pada head ke - i

$W^O$  : Matriks bobot output final setelah Concat

$Head_i$  : Attention Output pada head ke - i

Sedangkan, menurut Penelitian (Wijaya, 2021), rumus untuk Self Attention adalah sebagai berikut :

$$Attention(Q, K, V) = \text{Softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (2)$$

Keterangan pada Persamaan (2) adalah :

$d_k$  : Dimensi dari Vektor Key

$QK^T$  : Dot Product

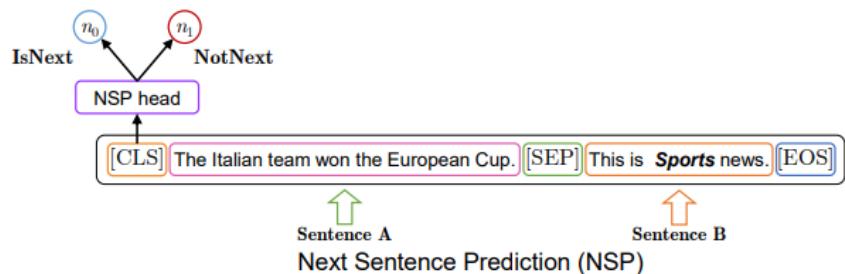
BERT dilatih menggunakan dua jenis tugas utama, yakni *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). MLM ditugaskan untuk menebak kata yang hilang atau tersembunyi dalam sebuah kalimat, seperti contoh kalimat tersembunyi akan dilabeli dengan [MASK], “Aku bermain [MASK] di siang hari” dan BERT pada MLM akan menebak kata [MASK] menjadi “Aku bermain bola di siang hari” berdasarkan dataset yang sudah dilatih secara umum. Berikut contoh ilustrasi MLM pada gambar 2.2 dibawah ini.



Gambar 2. 2 Ilustrasi Proses MLM

(Sumber : Sun et al, 2022)

Sedangkan untuk NSP, Model digunakan untuk menganalisis apakah dua kalimat yang diberikan memiliki keterkaitan yang logis (seperti berurutan) atau tidak. Sebagai contoh, ketika model diberikan pasangan kalimat “Aku makan nasi” dan diikuti dengan “Lalu aku mencuci piring”, BERT akan menyimpulkan bahwa kalimat tersebut saling berhubungan. Jika kalimat kedua diganti menjadi “Penguin berada di Antartika.” Maka BERT menyimpulkan bahwa kalimat tersebut tidak berhubungan dengan kalimat pertama (Sun et al., 2022).



Gambar 2. 3 Ilustrasi Proses NSP

Sumber : (Sun et al, 2022)

### 2.1.3 IndoBERT (*Indonesian Bidirectional Encoder Representations of Transformer*)

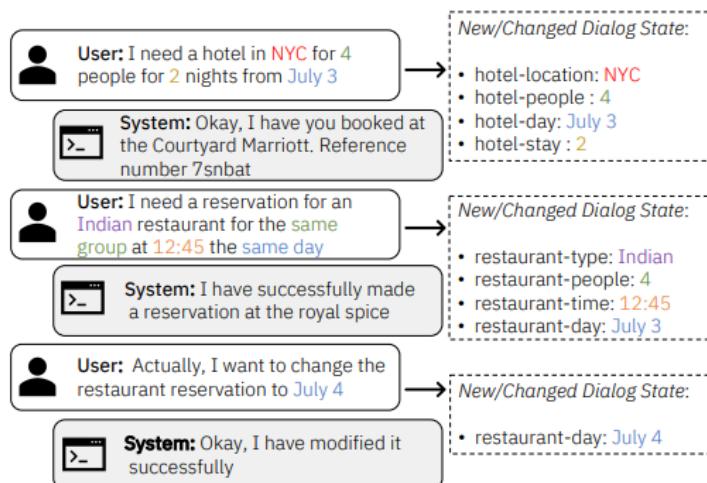
IndoBERT merupakan model dengan arsitektur Transformer yang sama dan terinspirasi dari BERT. Bila BERT adalah model dengan bahasa Inggris, maka IndoBERT dilatih menggunakan korpus atau kumpulan teks dengan ukuran besar sebagai data yang menggunakan bahasa indonesia. Model IndoBERT menggabungkan kosakata yang terkumpul lebih dari 220 juta kata, yang bersumber pada korpus web indonesia. Model IndoBERT juga dilatih dengan MLM dan NSP. IndoBERT pun dibagi menjadi tiga bagian, yakni IndoBERT-lite,

IndoBERT-base dan IndoBERT-large. Sesuai dengan namanya, ketiganya berbeda pada ukuran parameter (Koto et al., 2020).

#### 2.1.4 DST (*Dialogue State Tracking*)

*Dialogue State Tracking* berfungsi untuk mencatat informasi penting dari percakapan antara pengguna dan sistem, supaya sistem bisa mengerti kebutuhan pengguna dan membantu mencapai tujuannya. Hal ini membantu pengguna untuk mencapai tujuan mereka yang dianggap spesifik dan tertentu seperti memesan tiket, mencari tempat wisata maupun restoran serta menjadwalkan janji temu. Selama prosesnya, sistem perlu fokus dan memahami lebih dulu maksud dan keinginan pengguna, melacak perubahan tujuan maupun kebutuhan pengguna serta dan membuat langkah baru setelahnya (Xu et al., 2024).

*Dialogue State Tracking* membantu untuk merekam dan memperbarui status percakapan, sehingga penting bagi sistem untuk menyimpan informasi penting bagi pengguna. Kesimpulannya, DST membantu sistem untuk menyimpan memori yang dibutuhkan dalam isi percakapan, agar sistem mengetahui apa saja yang sudah disebutkan, apa yang masih perlu di diskusikan dan bagaimana cara merespon pengguna dengan tepat (Alhussayni et al., 2020)



Gambar 2. 4 Ilustrasi Proses Dialogue State Tracking

Sumber : (Venkateswaran et al., 2023)

## 2.2 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

No	Topik	Metode	Hasil	Ref
1.	Amal, M. I., Rahmasita, E. S., Suryaputra, E., & Rakhmawati, N. A. (2022). Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter. <i>Jurnal Teknik Informatika Dan Sistem Informasi</i> , 8(3). <a href="https://doi.org/10.28932/jutisi.v8i3.5483">https://doi.org/10.28932/jutisi.v8i3.5483</a>	Analisis Klasifikasi Sentimen pada twitter menggunakan pendekatan Machine learning dengan 4 metode yaitu Random Forest, Logistic Regression, SVM dan model Indobert.	Hasil yang diperoleh dari keempat metode diatas adalah RF sebesar 0.78, Indobert dengan akurasi 0.76 dan LR dengan akurasi 0.74. Performa IndoBERT yang rendah disebabkan ketidakseimbangan kelas serta keterbatasan jumlah data latih.	[1]
2.	Dwiyono, A., Fachrurrozi, M., Palembang-Prabu mulih, J., Ogan Ilir, K., & Selatan, S.	Klasifikasi Intent yang bertujuan untuk memahami maksud dari pengguna menggunakan tiga model transformer, yaitu BERT,	Model BERT mencapai akurasi 0.89, model roBERTa mencapai akurasi 0.84 sedangkan IndoBERT	[2]

	(2024). Analisis Perbandingan Klasifikasi Intent Chatbot Menggunakan Deep Learning BERT, RoBERTa, dan IndoBERT. <i>Journal of Information System Research</i> , 6(1), 595–606. <a href="https://doi.org/10.47065/josh.v6i1.6051">https://doi.org/10.47065/josh.v6i1.6051</a>	RoBERTa, dan IndoBERT. Penelitian menggunakan dataset University dengan 38 Intent yang sudah diterjemahkan kedalam Bahasa Indonesia.	mencapai akurasi 0.94. Performa IndoBERT lebih baik dibandingkan model lainnya dikarenakan pelatihan dataset yang lebih spesifik untuk Bahasa Indonesia	
3.	Hidayat, M. N., & Pramudita, R. (2023). Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT. <i>Information Management for Educators and Professionals</i> , 8(2), 161–170.	Model IndoBERT dimanfaatkan untuk meraih hasil dari sentimen dan menggunakan Confusion Matrix sebagai metode pengujian.	Hasil pengujian yang dilakukan oleh model IndoBERT mendapatkan akurasi sebesar 87% dengan nilai precision sebesar 87%, recall sebesar 91%, dan F-1 Score sebesar 89%.	[3]

4.	Fatonah, F. R., Maylawati, D. S., & Nurlatifah, E. (2024). Chatbot Edukasi Pra-Nikah berbasis Telegram Menggunakan Bidirectional Encoder Representations From Transformers (BERT). <i>Jurnal Algoritma</i> , 21(2), 29–40. <a href="https://doi.org/10.3364/algoritma.v.21-2.1657">https://doi.org/10.3364/algoritma.v.21-2.1657</a>	Dengan menggunakan Natural Language Processing, Penelitian menggunakan model BERT dengan konteks yang berasal dari website Kementerian Agama dan Buku Edukasi Pernikahan, lalu diimplementasikan ke Telegram. Pengujian menggunakan Non-Respon-Rate dan Metriks dari BERTScore.	Hasil pengujian menggunakan non-respon-rate menunjukkan hasil akurasi sebesar 76,92% dengan akurasi tertinggi 92%. Pengujian menggunakan BERTScore memberitahu bahwa chatbot mencapai hasil precision 86%, recall 83%, F-1 Score 84%.	[4]
5.	Ramadhan, T. I., Supriyatman, A., & Kurniawan, T. R. (2024). Evaluasi dan Implementasi Indobert Question Answering (QA) pada Domain Spesifik Menggunakan Mean Reciprocal Rank. <i>Jurnal Algoritma</i> , 21(1), 180–188. <a href="https://doi.org/10.3364/algoritma.v.21-1.1658">https://doi.org/10.3364/algoritma.v.21-1.1658</a>	Evaluasi untuk memberikan informasi spesifik model IndoBERT-QA dengan metode evaluasi Mean Reciprocal Rank (MRR).	Evaluasi untuk memberikan informasi spesifik model IndoBERT-QA dengan metode evaluasi Mean Reciprocal Rank (MRR).	[5]

	3364/algoritma/v.2 1-1.1542			
6.	<p>Widiansyah, M., Frazna Az-Zahra, F., &amp; Pambudi, A. (2024). Fine-Tuning Model Indobert (Indonesian Bidirectional Encoder Representations from Transformers) untuk Analisis Sentimen Berbasis Aspek pada Aplikasi M-Paspor. In Journal of Informatic Engineering (JOUTICA). <a href="https://jurnalteknik.uniska.ac.id/index.php/informatika">https://jurnalteknik.uniska.ac.id/index.php/informatika</a></p>	<p>Analisis sentiment terhadap ulasan Aplikasi M-Paspor menggunakan model IndoBERT.</p>	<p>Hasil dari model menunjukkan nilai akurasi tertinggi pada 94%, dan hasil evaluasi kinerja model menggunakan ROC mendapatkan nilai 94%.</p>	[6]
7.	<p>Afifa, N., Saputra, R. E., &amp; Nugrahaeni, R. A. (2023). Implementasi NLP Pada Chatbot Layanan Akademik</p>	<p>Untuk membantu petugas pelayanan akademik dalam menjawab pertanyaan, dikembangkan Chatbot dengan metode NLP (Natural Language Processing) dan Algoritma BERT.</p>	<p>Sistem menggunakan jenis pengujian Non-Respons_Rate dengan hasil akurasi sebesar 85% sehingga dapat disimpulkan bahwa sistem dapat</p>	[7]

	Dengan Algoritma Bert Implementation Of NLP On Academic Service Chatbot With Bertalgorithm.		berjalan sesuai dengan tujuannya.	
8.	Yusfida A'la, F. (2025). Edumatic: Jurnal Pendidikan Informatika Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas. 9(1), 256–265. <a href="https://doi.org/10.29408/edumatic.v9i1.29666">https://doi.org/10.29408/edumatic.v9i1.29666</a>	Penelitian bertujuan untuk mengelompokkan sentimen positif, negative dan netral pada ulasan Game menggunakan IndoBERT dan Teknik SMOTE.	Evaluasi menunjukkan model memiliki performa bervariasi, dengan akurasi awal sebesar 69%. Teknik SMOTE berhasil meningkatkan keseimbangan data dan meningkatkan akurasi menjadi 72,5%.	[8]
9.	Rayyan, F. (2022). Pengembangan Chatbot untuk Aplikasi Online Chat Telegram Dengan Pendekatan	Untuk mengatasi kesalahpahaman yang muncul pada pesan berbasis teks dalam berkomunikasi online-chat, dikembangkan sebuah chatbot yang dapat mendeteksi emosi dengan	Hasil pengujian mendapatkan akurasi, F1 Score, Recall, dan Precision pada training set sejumlah : 89%, 89%, 89% dan 90%. Sedangkan	[9]

	Klasifikasi Emosi pada Teks Menggunakan Metode IndoBERT-Lite.	model IndoBERT-lite. Dataset berisi kumpulan twit dengan total 4.403 data, dengan label 5 kelas yaitu : love, happy, anger, sadness, dan fear.	Validation didapat 70%, 71%, 70%, dan 72%.	
10.	Wildan Amru Hidayat, & Nastiti, V. R. S. (2024). PERBANDINGAN KINERJA PRE-TRAINED INDOBERT-BASE DAN INDOBERT-LITE PADA KLASIFIKASI SENTIMEN ULASAN TIKTOK TOKOPEDIA SELLER CENTER DENGAN MODEL INDOBERT. JSII (Jurnal Sistem Informasi), 11(2), 13–20. <a href="https://doi.org/10.30656/jsii.v11i2.9168">https://doi.org/10.30656/jsii.v11i2.9168</a>	Klasifikasi sentiment ulasan tiktok Tokopedia seller center dengan Model IndoBERT-lite dan IndoBERT-base. Dataset Training sebesar 64% (2012 data), Data validation 16% (504 data) dan Data testing sebanyak 20% (629 data).	IndoBERT-base-p2 menunjukkan performa lebih baik dibandingkan IndoBERT-lite-base-p2 dengan akurasi, presisi, recall, dan f1-score masing-masing sebesar 97%, sedangkan IndoBERT-lite-base-p2 hanya mencapai 94% pada keempat metrik tersebut.	[10]

11.	Wiharto, M. L. (2024). SISTEM CHATBOT SEBAGAI LAYANAN TUGAS AKHIR PROGRAM STUDI TEKNIK INFORMATIKA UNISSULA MENGGUNAKA N METODE Bidirectional Encoder Representations from Transformers (BERT).	Model IndoBERT digunakan untuk membantu Mahasiswa dalam prosedur administratif pada saat Tugas Akhir.	Evaluasi menunjukkan bahwa model IndoBERT memiliki akurasi tinggi dan baik yaitu 99,22%, F1 Score sebesar 99,3%, Presisi 99,4% dan recall 99,3%.	[11]
12.	Anugerah Simanjuntak, Rosni Lumbantoruan, Kartika Sianipar, Rut Gultom, Mario Simaremare, Samuel Situmeang, & Erwin Panggabean. (2024). Studi dan Analisis Hyperparameter Tuning IndoBERT Dalam	Penelitian ini menggunakan IndoBERT-base-p1 untuk deteksi berita palsu dan mengoptimalkan kinerjanya melalui penyetelan hyperparameter dengan tiga metode: Bayesian optimization, grid search, dan random search.	Dari ketiga metode tuning, Bayesian optimization terbukti paling efektif dengan presisi 88,79%, recall 94,5%, dan F1-score 91,56% untuk label "fake".	[12]

	Pendeteksian Berita Palsu. Jurnal Nasional Teknik Elektro Dan Teknologi Informasi, 13(1), 60–67. <a href="https://doi.org/10.22146/jnteti.v13i1.8532">https://doi.org/10.22146/jnteti.v13i1.8532</a>			
13.	Hafidz, I., Mukti, B. S., Zahra, Q., Naseela, I., Yudistira, D., Pratama, P. A., Purnama, M., Fajrin Ariyani, N., Astuti, M., & Tjahyanto, A. (2024). Halal Research 4 (2) (2024) 117-131 Chatbot Model Development Using BERT for West Sumatra Halal Tourism Information.	Pengembangan chatbot untuk pariwisata halal di Sumatera mengumpulkan data sebanyak 1125 pertanyaan dengan data split sebesar 80:20. Metode yang digunakan adalah IndoBERT-base multilingual.	Chatbot memiliki akurasi tinggi (0,96), namun performa terendah ada pada label “belanja” (recall 0,64) dan “budaya” serta “kuliner” (presisi 0,89).	[13]
14.	Pandhare, A. (2024). Using BERT to Build Chatbots for	Pengembangan chatbot untuk e-commerce menggunakan BERT dan diintegrasikan ke Website.	Aplikasi chatbot berhasil dijalankan dan memberikan respon yang sesuai	[14]

	E-Commerce. <a href="https://doi.org/10.13140/RG.2.2.2840.0.88321">https://doi.org/10.13140/RG.2.2.2840.0.88321</a>		dan meningkatkan peluang untuk membuat aplikasi yang lebih inovatif.	
15.	Maulana, A. E., Rosita, Y. D., & Sari, D. K. (2025). Pembuatan Chatbot pada Website Persatuan Gizi Kabupaten Karawang menggunakan IndoBERT.	Pengembangan Chatbot menggunakan IndoBERT untuk Informasi terkait Gizi pada Website Persatuan Gizi Kabupaten Karawang dan Persatuan Gizi Kabupaten Karawang menggunakan IndoBERT.	Akurasi chatbot didapatkan sebesar 87,67%, loss sebesar 0,2%, presisi sebesar 85,35%, recall sebesar Menggunakan Similarity untuk mendekripsi pertanyaan yang mirip.	[15]

## 2.3 Teori Pendukung

### 2.3.1 Chatbot

*Chatbot* adalah sebuah program komputer yang dirancang untuk mereplikasi percakapan dengan Pengguna, baik melalui suara maupun teks (Wahyudi Harahap & Fitria, 2020). Chatbot merupakan sebuah program komputer yang dibuat untuk berinteraksi dengan pengguna melalui teks atau pesan tertulis (Diantoni et al., 2024). Chatbot bertujuan untuk menjadi asisten virtual yang dapat memberikan solusi dari permasalahan yang dihadapi pengguna saat menggunakan aplikasi. Chatbot beroperasi tanpa batasan waktu, sehingga dapat memberikan layanan kepada pengguna kapan saja mereka membutuhkannya. Dengan demikian, penggunaan chatbot dapat berkontribusi pada peningkatan efisiensi operasional (Cahya Hutama & Titi Komalasari, 2021). Dalam pengembangannya, chatbot dapat menggunakan Artificial Intelligence (AI), Machine Learning (ML), Deep Learning dan Natural Language Processing (NLP). Dasar dari cara kerja chatbot adalah dengan mengidentifikasi kata kunci dalam data yang diterima dan memberikan balasan berdasarkan kata kunci yang paling relevan, atau pola kata

yang paling mirip dalam basis data tekstual. Dengan demikian, ketika pengguna mengirimkan suatu permintaan, chatbot akan merespons dengan jawaban yang spesifik sesuai dengan kata kunci yang dikirimkan (Wahyudi Harahap & Fitria, 2020).

### 2.3.2 Bot Telegram (Bot Father)

Bot Telegram merupakan salah satu fitur yang disediakan oleh platform Telegram untuk memungkinkan pengembangan layanan otomatis yang dapat berinteraksi langsung dengan pengguna (Nurizna, 2021). Dalam proses pembuatan bot, Telegram menyediakan sebuah bot khusus bernama BotFather, yang berfungsi sebagai pusat pengaturan dan pembuatan bot baru. Melalui BotFather, pengembang dapat membuat bot baru dengan cara memberikan nama, username, serta mendapatkan token API yang digunakan sebagai kunci autentikasi agar bot dapat terhubung dengan sistem backend yang dikembangkan. Token ini bersifat unik dan wajib dijaga kerahasiaannya, karena digunakan untuk mengontrol serta mengelola komunikasi antara server dan aplikasi Telegram. Dengan adanya BotFather, pembuatan bot di Telegram menjadi lebih mudah, terstruktur, dan aman. Bot Telegram yang dibuat dapat digunakan untuk berbagai keperluan, seperti chatbot informasi akademik, layanan customer service otomatis, hingga sistem notifikasi.

## 2.4 Gap Penelitian

Tabel 2. 2 Gap Penelitian

No	Judul Penelitian	Hasil	Gap Penelitian
1.	Maulana, A. E., Rosita, Y. D., & Sari, D. K. (2025). Pembuatan Chatbot pada Website Persatuan Gizi Kabupaten Karawang menggunakan IndoBERT.	Akurasi chatbot didapatkan sebesar 87,67%, loss sebesar 0,2%, presisi sebesar 85,35%, recall sebesar 86,30%,	Ditemukan tidak adanya proses Normalisasi, sehingga data yang digunakan

		<p>dan F1-score sebesar 84,64%. Chatbot dapat menjawab pertanyaan dengan akurat.</p>	<p>sudah bersih. Serta belum menggunakan Metode DST (Dialogue State Tracking)</p>
2	<p>Wiharto, M. L. (2024). SISTEM CHATBOT SEBAGAI LAYANAN TUGAS AKHIR PROGRAM STUDI TEKNIK INFORMATIKA UNISSULA MENGGUNAKAN METODE Bidirectional Encoder Representations from Transformers (BERT).</p>	<p>Evaluasi menunjukkan bahwa model IndoBERT memiliki akurasi tinggi dan baik yaitu 99,22%, F1 Score sebesar 99,3%, Presisi 99,4% dan recall 99,3%.</p>	<p>Ditemukan bahwa tidak menggunakan Early Stop Callback untuk mencegah Overfitting, belum menggunakan Normalisasi dan belum menggunakan metode Dialogue State Tracking.</p>
3	<p>Fatonah, F. R., Maylawati, D. S., &amp; Nurlatifah, E. (2024). Chatbot Edukasi Pra-Nikah berbasis Telegram Menggunakan Bidirectional Encoder Representations From Transformers (BERT). Jurnal Algoritma, 21(2), 29–40.</p>	<p>Dengan menggunakan Natural Language Processing, Penelitian menggunakan model BERT dengan konteks</p>	<p>Ditemukan bahwa tidak menggunakan Normalisasi</p>

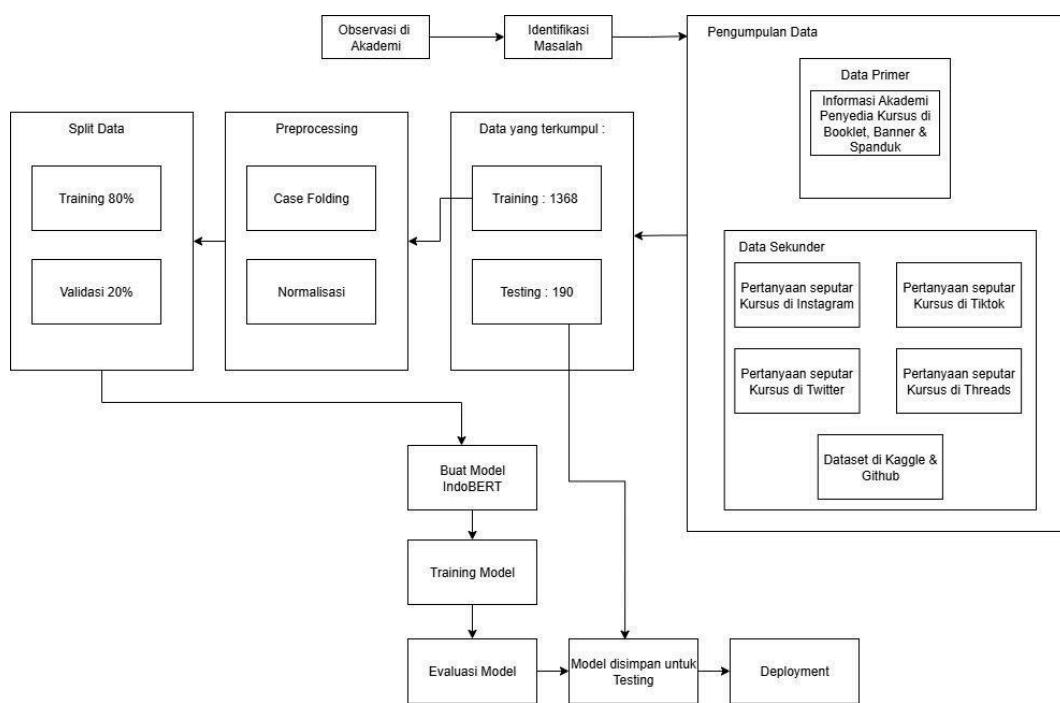
	<p><a href="https://doi.org/10.33364/algoritma/v.21-2.1657">https://doi.org/10.33364/algoritma/v.21-2.1657</a></p>	<p>yang berasal dari website Kementrian Agama dan Buku Edukasi Pernikahan, lalu diimplementasikan ke Telegram. Pengujian menggunakan Non-Respon-Rate dan Metriks dari BERTScore.</p>	
--	--	--	--

## BAB III METODE PENELITIAN

### 3.1 Jenis Penelitian

Metode penelitian yang digunakan dalam riset ini adalah metode kuantitatif. Menurut Creswell, penelitian kuantitatif merupakan metode yang digunakan untuk menguji teori melalui analisis hubungan antar variabel yang dapat diukur secara numerik (Siroj et al., 2024). Dalam Penelitian ini, metode kuantitatif digunakan untuk menguji performa model IndoBERT dalam memprediksi dan mengklasifikasi intent dari pertanyaan pengguna.

### 3.2 Tahapan Penelitian



Gambar 3. 1 Tahapan Penelitian

Tahapan penelitian pada Implementasi IndoBERT dan DST pada Chatbot Layanan Informasi Penyedia Kursus adalah sebagai berikut :

## 1. Observasi di Akademi Penyedia Kursus

Langkah pertama dalam penelitian ini adalah mengamati dan melakukan observasi di lingkungan Akademi. Observasi atau pengamatan ini bertujuan untuk memahami permasalahan yang ada di lapangan, baik dari sisi siswa, guru, admin hingga sistem yang ada pada Akademi penyedia kursus. Hasil observasi akan menjadi dasar untuk menemukan celah yang dapat diatasi dengan pendekatan teknologi.

## 2. Identifikasi Masalah

Berdasarkan hasil observasi, Peneliti melakukan identifikasi masalah secara lebih luas dan mendalam. Masalah yang ditemukan kembali dianalisis untuk menemukan permasalahan utama yang akan dijadikan fokus di penelitian hingga pencarian solusi lewat pendekatan teknologi. Permasalahan yang ditemukan adalah akademi penyedia kursus masih menggunakan sistem layanan informasi secara manual, yang mana terdapat pertanyaan pengguna sempat tidak terjawab dan membutuhkan waktu cukup lama untuk merespon satu pertanyaan, juga terbatas dengan adanya jam kerja. Hal tersebut menghambat pengguna dalam meraih informasi yang dibutuhkan. Sehingga solusi yang akan digunakan adalah pembuatan Chatbot Akademik.

## 3. Pengumpulan Data

Tahapan selanjutnya adalah pengumpulan data. Data yang relevan dengan permasalahan yang telah diidentifikasi, dikumpulkan menjadi satu berupa teks pertanyaan wali murid, orang tua maupun calon siswa yang hendak mendaftarkan diri. Data yang didapatkan pada penelitian ini terbagi menjadi dua macam, yaitu data primer dan data sekunder. Dikutip dari (Widianto & Nasution, 2022), Menurut Sunyoto, Data primer merupakan data yang diperoleh secara langsung dan dikumpulkan sendiri oleh peneliti secara khusus. Sedangkan Data sekunder adalah data yang diperoleh dari hasil pengumpulan oleh pihak lain dan berasal dari berbagai tempat.

Data primer yang dikumpulkan oleh peneliti menggunakan metode observasi adalah Informasi terkait operasional akademi seperti biaya kursus, informasi pendaftaran, pembayaran dan program lainnya yang berasal dari booklet dan brosur. Sedangkan Teknik Pengumpulan data sekunder dalam penelitian ini adalah meraih data pertanyaan di sosial media penyedia lembaga kursus anak-anak dan remaja. Data sekunder yang dikumpulkan (berupa pertanyaan seputar kursus) berasal dari sosial media seperti Instagram, Threads, Twitter, dan Tiktok, juga Website kumpulan data yakni Kaggle.

#### 4. Pembersihan Data (Preprocessing)

Data yang terkumpul akan di pre-processing terlebih dahulu, seperti pembersihan simbol dan tanda baca, case folding (mengubah teks menjadi huruf kecil), serta normalisasi teks yang berguna untuk menyempurnakan teks typo atau teks yang singkat seperti “gmn” menjadi “bagaimana”, “bljr” menjadi “belajar”. Tahapan training tidak menggunakan Stopwords maupun Stemming untuk menjaga informasi konteks (Wilie et al., 2020). Training dengan IndoBERT akan melatih model memahami struktur kalimat menggunakan bahasa indonesia.

#### 5. Pembagian Data (Spliting Data)

Setelah seluruh data telah terkumpul dengan total 1368 data pertanyaan akademik, peneliti membagi data tersebut menjadi tiga bagian, diantaranya adalah 80% untuk data training, 20% untuk data validasi, serta data testing yang terpisah sebanyak 190 data pertanyaan. Data Training akan digunakan untuk melatih model sehingga model akan dapat memahami pola dan struktur kata, data validasi digunakan untuk evaluasi performa dan mencegah overfitting, sedangkan data testing digunakan untuk menguji kemampuan generalisasi model setelah selesai dilatih.

#### 6. Membuat Model IndoBERT dan Training Model

Dataset yang bersih selanjutnya akan dilatih dengan model IndoBERT-base-p1, dan difokuskan menuju task tertentu yaitu *Classification Text*. Proses *Fine-Tuning* menggunakan data yang sudah diberi label dan membuat

model terfokus pada konteks akademik serta gaya bahasa yang digunakan oleh pengguna. Fine-tuning dengan model IndoBERT *Classification Text* akan menganalisa input pengguna dan memprediksi label apa yang sesuai dengan pertanyaan pengguna. Training model menggunakan *class* Training Arguments untuk mengatur pelatihan model yang akan berisi semua parameter yang digunakan oleh Trainer.

## 7. Evaluasi Model

Setelah model selesai dilatih dan diuji, selanjutnya adalah proses evaluasi Model terhadap hasil dari data latih dan data uji. Evaluasi menggunakan confusion matrix, akurasi, precision, recall dan f1-score. Hasil evaluasi tersebut akan menunjukkan sejauh mana model dapat melakukan klasifikasi dan dapat memahami input pengguna dengan baik, serta mengukur apakah model layak digunakan.

## 8. Testing

Model yang sudah dievaluasi selanjutnya disimpan agar dapat digunakan kembali pada tahap pengujian (testing). Penyimpanan model ini bertujuan untuk memastikan bahwa model dengan parameter terbaik yang telah diperoleh selama proses pelatihan dan evaluasi tetap terjaga dan dapat diimplementasikan tanpa perlu melakukan pelatihan ulang. Pada tahap testing, model akan diuji menggunakan 190 data uji (test data) yang belum pernah dilihat atau digunakan sebelumnya selama proses pelatihan maupun validasi. Penggunaan data uji ini bertujuan untuk mengetahui sejauh mana kemampuan model dalam mengenali dan memproses data baru secara objektif, sehingga dapat dinilai performa akhir model dalam kondisi nyata. Hasil dari tahap testing ini menjadi indikator utama dalam menilai kualitas dan keakuratan model sebelum diterapkan secara luas kepada pengguna akhir.

## 9. Deployment ke Bot Telegram

Tahap deployment ke Bot Telegram dilakukan untuk menerapkan model chatbot yang telah selesai dilatih dan diuji agar dapat digunakan secara langsung

oleh pengguna melalui aplikasi Telegram. Pada tahap ini, model diintegrasikan dengan API Telegram menggunakan token bot yang telah didaftarkan sebelumnya. Setelah proses integrasi selesai, chatbot dapat diakses oleh pengguna melalui akun bot resmi, sehingga pengguna dapat langsung berinteraksi dan mengajukan berbagai pertanyaan sesuai kebutuhan. Proses deployment ini memungkinkan chatbot memberikan jawaban secara real-time, menguji ketepatan respons di kondisi nyata, serta mempermudah distribusi layanan tanpa perlu instalasi tambahan oleh pengguna.

### 3.3 Instrumen Penelitian

Tabel 3. 1 Instrumen Penelitian Perangkat Lunak (*Software*)

NO	Perangkat Lunak ( <i>Software</i> )	Tujuan
1	<i>Microsoft Excel</i>	Mengumpulkan data pertanyaan pengguna terkait operasional Akademik dan menganalisis kategori pertanyaan
2	<i>Google Colab</i>	<i>Pre-processing</i> data pertanyaan pelanggan dengan proses <i>Natural Language Processing</i> , pembuatan model <i>chatbot</i> dengan IndoBERT serta mengevaluasi model.
3	Visual Studio Code	Penyimpanan model dan membuat Handler Bot untuk deployment ke Telegram.
4	Python Versi 3.10	Membuat Class untuk Integrasi antara kode IndoBERT dengan Bot Telegram
5	1. Library Transformers v4.52.4 2. PyTorch v2.6.0 3. numpy v2.0.2 4. scikit-learn.	Untuk memanggil model IndoBERT, Load Tokenizer hingga Evaluasi model.

Tabel 3. 2 Instrumen Penelitian Perangkat Keras (*Hardware*)

No	Perangkat Keras ( <i>Hardware</i> )	Tujuan
1	Lenovo Ideapad 3 AMD Ryzen 5000 Series RAM 16GB	Pembuatan penulisan proposal dan Pengolahan data.

### 3.4 Subjek Penelitian

Subjek penelitian dalam Implementasi IndoBERT dan DST pada chatbot layanan informasi penyedia kursus merujuk pada pengembangan model chatbot menggunakan teknologi Natural Language Processing dengan pemanfaatan model IndoBERT serta Dialogue State Tracking. Model ini akan diintegrasikan ke dalam platform Telegram untuk mengatasi keterlambatan dalam merespons pertanyaan pelanggan serta untuk mengevaluasi performa chatbot dalam menjalankan tugasnya. Performa chatbot yang dibangun dengan IndoBERT dan Dialogue State Tracking ini akan dievaluasi berdasarkan akurasi klasifikasi intent dan kemampuan chatbot untuk memberikan jawaban yang relevan dan tepat.

## BAB IV PEMBAHASAN

### 4.1 Persiapan Data

Data yang sudah terkumpul secara keseluruhan atau dataset kotor memiliki 1564 data pertanyaan terkait kursus di akademi dan belum diberikan label. Untuk mengetahui label secara otomatis dari dataset yang belum berlabel, dilakukan proses penyaringan data menggunakan kata kunci tertentu. Kata kunci ini diperoleh dengan melihat kata-kata yang paling sering muncul dalam kumpulan pertanyaan (*wordcount*).

Tabel 4. 1 Keterangan WordCount

NO	List Top WordCount	Jumlah	Keyword Label
1.	kursus	389	Kursus Coding, Kursus Design, Kursus Hisensei
2.	coding	246	Kursus Coding, Info_Coding
3.	ajar	238	Pembelajaran
4.	biaya	176	Biaya, Biaya Coding, Biaya Design, Biaya Hisensei
5.	design	117	Kursus Design, Info_Design
6.	bayar	118	Pembayaran, Biaya
7.	kelas	118	Pembelajaran, Peserta
8.	daftar	96	Pendaftaran
9.	hisensei	77	Kursus Hisensei
10.	aplikasi	67	Aplikasi
11.	harga	78	Biaya, Biaya Coding, Biaya Design, Biaya Hisensei
12.	laptop	79	Spesifikasi Laptop
13.	trial	82	Trial Class
14.	desainer	43	Info Design, Kursus Design
15.	admin	47	Hubungi Kami

Hasil dari metode ini membantu mengidentifikasi kata-kata yang mewakili topik atau maksud dari pertanyaan, sehingga bisa digunakan sebagai dasar untuk mengelompokkan data ke dalam label yang sesuai. Data yang berhasil di filter dan

digunakan untuk training sebesar 1368 data dengan 19 kategori, dan 190 data untuk data testing. Menurut penelitian (Lin et al., 2024), menggunakan dataset 100 data per-intent, membuat hasil akurasi model lebih baik. Namun dikarenakan keterbatasan data dan mengambil jumlah paling sedikit dalam dataset yakni 72 pertanyaan, Peneliti menyiapkan 1 Intent terdiri dari 72 pertanyaan untuk tetap menjaga keseimbangan atau balance pada dataset.

Tabel 4. 2 Sebagian contoh Dataset yang digunakan

No	intent	pattern
1	kursuscoding	Klo coding tuh kursusnya ngapain aja?
2	Kursusdesign	Kursus design belajar apa aja ya
3	Kursushisensei	Kursus hisensei itu apa?
4	Pendaftaran	Cara daftar gmn kakak
5	Pembayaran	Pembayarannya lewat apa? Bisa transfer atau cash?
6	Aplikasi	apk yg sering dipakai buat belajar disini apa?
7	SpekLaptop	Spesifikasi laptopnya harus yang kaya gimana buat coding?
8	Hubungikami	ini lokasiny dmn?
9	TrialClass	Trial class itu gmn maksudnya, kelas gratis?
10	Peserta	ada batesan umur ga sii ??
11	biaya_coding	Biaya kursus coding berapa?
12	biaya_design	Biaya kursus design berapa?
13	biaya_hisensei	Biaya kursus hisensei berapa?
14	Info_coding	Coding itu apa sih?
15	Info_design	Apa sih itu desain grafis?
16	Umum	halo, ada info lebih lengkap soal lesnya?
17	biaya	Biaya kursusnya berapa?
18	Out of scope	Rekomendasi online shop tempat beli laptop dong.
19	Pembelajaran	1 anak brp kalo 1 kelasnya?

Intent dibuat berdasarkan kebutuhan dan pertanyaan yang sering dilontarkan oleh pengguna di berbagai media sosial pada bidang kursus akademi, salah satunya adalah Instagram. Intent yang digunakan adalah sebagai berikut :

Tabel 4. 3 Keterangan Intent

No.	Intent / Label	Penjelasan
1.	Umum	Berisi pertanyaan mengenai informasi umum, seperti “ada berapa banyak kursus yang ada di Akademi?” atau “Mohon info lengkapnya.”
2.	Kursus Coding	Berisi pertanyaan terkait kursus coding dan pembelajaran pada kursus coding.
3.	Kursus Design	Berisi pertanyaan terkait kursus desain dan pembelajaran pada kursus desain.
4.	Kursus Bimbel	Berisi pertanyaan terkait kursus bimbel yaitu Matematika dan Bahasa Inggris untuk anak-anak.
5.	Info Coding	Berisikan pertanyaan umum terkait coding seperti “Apa itu Coding?” dan bertanya terkait kesulitan apa yang biasanya ditemui saat belajar Coding maupun profesiya.
6.	Info Design	Berisikan pertanyaan umum terkait coding seperti “Apa itu Desain?” dan bertanya terkait kesulitan apa yang biasanya ditemui saat belajar Desain maupun profesiya.
7.	Aplikasi	Berisikan pertanyaan terkait aplikasi apa yang digunakan saat pembelajaran.

8.	Pembelajaran	Berisikan pertanyaan terkait pembelajaran seperti bertanya mengenai jadwal kelas, dan sistem pembelajaran.
9.	Pendaftaran	Berisikan pertanyaan terkait pendaftaran seperti cara mendaftar kursus.
10	Pembayaran	Berisikan pertanyaan terkait metode dan proses pembayaran yang tersedia.
11	Biaya	Berisikan pertanyaan terkait biaya kursus secara umum tanpa menyebutkan jenis kursus, seperti contoh “biayanya berapa kak?”
12	Biaya Coding	Berisikan pertanyaan terkait biaya di kursus coding.
13	Biaya Design	Berisikan pertanyaan terkait biaya di kursus desain.
14	Biaya Bimbel	Berisikan pertanyaan terkait biaya di kursus bahasa inggris dan matematika.
15	Spek Laptop	Berisikan pertanyaan terkait spesifikasi laptop yang digunakan di Akademi.
16	Peserta	Berisikan pertanyaan terkait berapa usia minimal dan maksimal untuk mendaftar kursus.
17	Trial Class	Berisikan pertanyaan terkait Trial Class atau Mencoba kelas dan melakukan tes secara gratis sebelum mendaftar kursus.
18	Hubungi Kami	Berisikan pertanyaan terkait lokasi akademi, nomor admin dan jam operasional.

19	Out Of Scope	Berisikan pertanyaan yang berada diluar konteks kursus, guna menyediakan batasan percakapan Chatbot agar tidak terlalu luas.
----	--------------	--

#### 4.2 Pre-Processing Data

Pre-Processing data adalah teknik untuk membersihkan data sebelum dilatih dengan model (Agung et al., 2023). Pre-Processing data yang ada pada Implementasi IndoBERT pada Chatbot Layanan Penyedia Kursus adalah Case folding yaitu mengubah teks menjadi huruf kecil seluruhnya, pembersihan kata dari simbol maupun tanda baca dan normalisasi, bertujuan untuk menjelaskan kalimat yang memiliki kesalahan eja atau menjelaskan bahasa gaul agar dimengerti oleh komputer dan sesuai dengan ejaan Kamus Besar Bahasa Indonesia (KBBI). Berikut tabel tersedia untuk menampilkan proses Pre-Processing.

Tabel 4. 4 Tahap Case Folding

Sebelum Lowercase	Sesudah Lowercase
“Biayanya brp kak?”	“biayanya brp kak?”

Tabel 4. 5 Tahap Penghapusan Simbol atau Tanda baca

Sebelum Pembersihan	Sesudah Pembersihan
“biayanya brp kak?”	“biayanya brp kak”

Tabel 4. 6 Tahap Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
“biayanya brp kak”	“biayanya berapa kak”

Normalisasi digunakan untuk menyamakan bentuk kata agar memiliki struktur yang konsisten secara bahasa. Proses ini penting untuk meningkatkan akurasi dalam analisis teks, terutama ketika menghadapi kata-kata tidak baku, singkatan,

atau bentuk slang yang umum digunakan dalam komunikasi informal. Dalam proses ini, normalisasi dilakukan dengan menggunakan daftar kosakata dari GitHub (<https://github.com/evrintobing17/NormalisasiKata>) yang memetakan kata-kata tidak baku ke bentuk bakunya dengan total vocab 15.0007 kata.

Tabel 4. 7 Contoh Vocab Normalisasi

Slang	Formal
alooo	Halo
brp	Berapa
Cding	Coding
Makaciw	Terima Kasih
Kaka	Kakak

```
# 1. Load data
df = pd.read_excel("Training2.xlsx")
norm1 = pd.read_csv("normalisasi.csv")
norm2 = pd.read_csv("kbbaConverted.csv")

# 2. Buat dictionary normalisasi
norm_dict = dict(zip(norm1['slang'], norm1['formal']))
kbba_dict = dict(zip(norm2['slang'], norm2['formal']))

def normalize_text(text):
    text = text.lower()
    text = re.sub(r'^a-zA-Z0-9\s]', '', text)
    tokens = text.split()
    normalized = [kbba_dict.get(norm_dict.get(t, t),
norm_dict.get(t, t)) for t in tokens]
    return ' '.join(normalized)

# 3. Normalisasi kolom pattern
df["pattern_clean"] = df["pattern"].apply(normalize_text)
```

Kode diatas berfungsi untuk menggabungkan seluruh hasil dari normalisasi hingga pembersihan simbol atau tanda baca dari kolom pattern, menjadi satu kolom yaitu pattern\_clean.

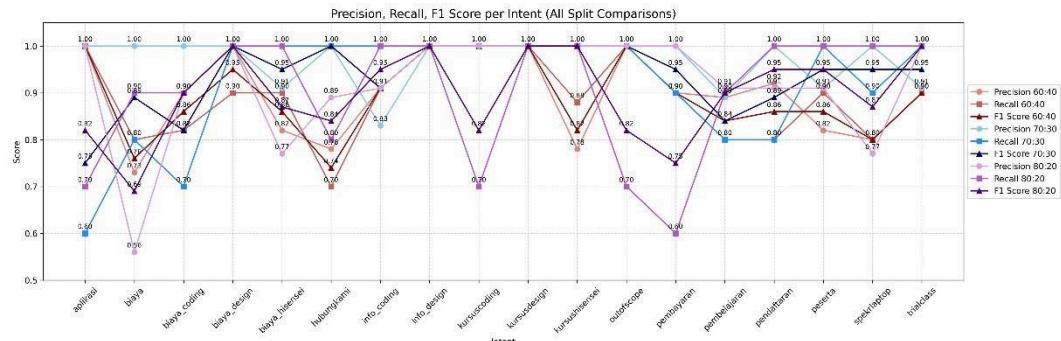
### 4.3 Split Data

Data yang sudah bersih selanjutnya dilakukan Split Data. Split Data adalah proses sebuah Dataset dibagi menjadi beberapa bagian sesuai dengan rasio yang

ditentukan. Dalam penelitian ini, peneliti mencoba menggunakan tiga skala rasio, yaitu 60:40, 70:30, dan 80:20, dengan hasil evaluasi performa model ditampilkan pada gambar di bawah ini.

label	SPLIT DATA 60:40				SPLIT DATA 70:30				SPLIT DATA 80:20			
	precision	recall	f1-score	support	precision	recall	f1-score	support	precision	recall	f1-score	support
aplikasi	1	1	1	10	1	0,6	0,75	10	1	0,7	0,82	10
biaya	0,73	0,8	0,76	10	1	0,8	0,89	10	0,56	0,9	0,69	10
biaya_coding	0,9	0,82	0,86	10	1	0,7	0,82	10	0,9	0,9	0,9	10
biaya_design	1	0,9	0,95	10	1	1	1	10	1	1	1	10
biaya_hisensei	0,82	0,9	0,86	10	0,91	1	0,95	10	0,77	1	0,87	10
hubungkami	0,78	0,7	0,74	10	1	1	1	10	0,89	0,8	0,84	10
info_coding	0,91	0,91	0,91	10	0,83	1	0,91	10	0,91	1	0,95	10
info_design	1	1	1	10	1	1	1	10	1	1	1	10
kursuscoding	1	1	1	10	1	1	1	10	1	0,7	0,82	10
kursusdesign	1	1	1	10	1	1	1	10	1	1	1	10
kursushisensei	0,78	0,88	0,82	10	1	1	1	10	1	1	1	10
outofscope	1	1	1	10	1	1	1	10	1	0,7	0,82	10
pembayaran	0,9	0,9	0,9	10	1	0,9	0,95	10	1	0,6	0,75	10
pembelajaran	0,89	0,8	0,84	10	0,89	0,8	0,84	10	0,91	0,9	0,9	10
pendaftaran	0,92	0,8	0,86	10	1	0,8	0,89	10	0,91	1	0,95	10
peserta	0,82	0,9	0,86	10	0,91	1	0,95	10	0,91	1	0,95	10
spekrlaptop	0,8	0,8	0,8	10	1	0,9	0,95	10	0,77	1	0,87	10
trialclass	0,9	0,9	0,9	10	0,91	1	0,95	10	1	1	1	10
umum	0,6	0,7	0,64	10	0,62	0,8	0,7	10	0,82	0,9	0,86	10
accuracy		0,86	0,86	190		0,91	0,9	190		0,89	0,89	190
macro avg	0,87	0,86	0,86	190	0,93	0,91	0,9	190	0,92	0,89	0,89	190
weighted avg	0,87	0,86	0,86	190	0,93	0,91	0,9	190	0,92	0,89	0,89	190

Gambar 4. 1 Tabel Keseluruhan Split Data



Gambar 4. 2 Grafik Keseluruhan Split Data

Data latih digunakan untuk melatih model agar model dapat memahami struktur kata yang ada. Sedangkan data uji digunakan untuk menguji model dan data validasi berguna untuk menjaga performa model, mencegah model dari overfitting dan underfitting, sehingga kualitas model meningkat dan terjaga kualitasnya (Maleki et al., 2020). Berdasarkan hasil pengujian, split data dengan rasio 70:30 menghasilkan akurasi dan nilai F1-score yang lebih baik pada data pengujian dibandingkan dua skala lainnya. Kombinasi ini memberikan keseimbangan yang optimal dengan jumlah data yang memadai.

#### 4.4 Proses Training, Fine Tuning dan Testing Model IndoBERT

Tahapan selanjutnya adalah memasuki proses Training dengan mempersiapkan label yang sudah di encoder menggunakan kode pada dibawah ini.

```
# Encode label
label_encoder = LabelEncoder()
train_labels_enc = label_encoder.fit_transform(train_labels)
val_labels_enc = label_encoder.transform(val_labels)

# Simpan mapping label
label2id = {label: int(idx) for idx, label in
enumerate(label_encoder.classes_)}
id2label = {int(v): k for k, v in label2id.items()}

with open("indobert_newest_label.json", "w", encoding="utf-8") as f:
    json.dump({"label2id": label2id, "id2label": id2label}, f,
indent=2, ensure_ascii=False)
```

Label Encoder digunakan untuk mengubah intent atau label yang awalnya berbentuk teks seperti “kursuscoding”, “biaya”, “pendaftaran”, menjadi representasi angka numerik. Label Encoder diperlukan karena model IndoBERT hanya bisa menganalisis data dan memproses data numerik, bukan teks. Proses label encoder ini dilakukan oleh library sklearn.preprocessing, khususnys class LabelEncoder yang mengubah label kategori dari teks menjadi angka seperti 1 hingga 10, sesuai dengan jumlah intent atau label. Fit\_transform() digunakan agar mesin dapat mempelajari semua label unik dan mengubah string menjadi angka. Transform() digunakan agar data testing tetap menggunakan urutan label yang sama saat Training. Setelah selesai diubah dari teks menjadi angka numerik, output diubah menjadi file json, yaitu “indobert\_newest\_label.json” untuk menyimpan label mapping. Berikut output dari label encoder :

```
"id2label": {
    "0": "aplikasi",
    "1": "biaya",
    "2": "biaya_coding",
    "3": "biaya_design",
    "4": "biaya_hisensei",
    "5": "hubungikami",
    "6": "info_coding",
    "7": "info_design",
    "8": "kursuscoding",
    "9": "kursusdesign",
    "10": "kursushisensei",
```

```

    "11": "outofscope",
    "12": "pembayaran",
    "13": "pembelajaran",
    "14": "pendaftaran",
    "15": "peserta",
    "16": "speklaptop",
    "17": "trialclass",
    "18": "umum"
}

```

Id2label akan menjadi representasi intent yang awalnya berupa teks, kemudian diubah menjadi representasi angka numerik. Proses ini dilakukan dengan cara memberi indeks atau label angka pada setiap intent, berdasarkan urutan abjad atau skema tertentu yang sudah ditentukan sebelumnya. Jika memiliki intent seperti "biaya", "pembayaran", dan "pendaftaran", maka masing-masing akan diubah menjadi angka 0, 1, dan 2. Tujuan dari pengubahan ini adalah agar model seperti IndoBERT, dapat memproses data dalam bentuk numerik, karena model tidak bisa langsung memproses teks sebagai input target. Konversi label ini umumnya dilakukan menggunakan fungsi LabelEncoder dari library scikit-learn. Dengan demikian, saat proses pelatihan, model akan belajar mengenali setiap intent dalam bentuk angka, dan setelah prediksi selesai, angka tersebut dapat dikonversi kembali ke bentuk teks asli dengan memanfaatkan mapping id2label.

```

from transformers import BertTokenizer

tokenizer =
BertTokenizer.from_pretrained("indobenchmark/indobert-base-p1")

def tokenize(texts):
    return tokenizer(texts, padding="max_length",
truncation=True, max_length=128)

train_encodings = tokenize(train_texts.tolist())
val_encodings = tokenize(val_texts.tolist())

```

Langkah kedua adalah memasukkan library Transformers dari Hugging Face dan memasukkan class BertTokenizer. Tokenizer berguna untuk mengubah teks mentah menjadi numerik (Token ID) yang dapat diproses oleh model. Fungsi Tokenize() menerima data teks dan menembalikan teks tersebut dalam bentuk encoding BERT, dengan ketentuan padding="max\_length" dan truncation=true, artinya teks akan dipotong jika terlalu panjang atau dipenuhi dengan token "PAD"

hingga batas panjang yang ditentukan bila terlalu pendek, dalam hal ini, panjang maksimal token encodingnya adalah max\_length = 128 token.

```
from transformers import BertForSequenceClassification,
Trainer, TrainingArguments, EarlyStoppingCallback

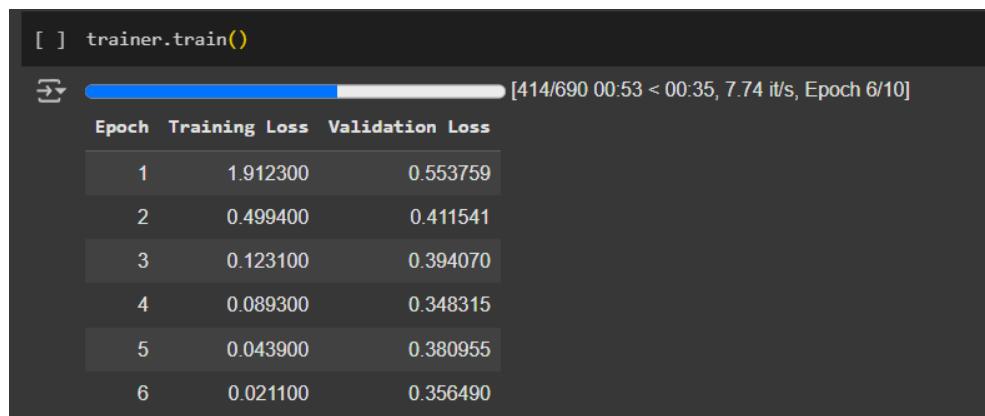
model = BertForSequenceClassification.from_pretrained(
    "indobenchmark/indobert-base-p1",
    num_labels=len(label2id),
    label2id=label2id,
    id2label=id2label
)
```

Langkah ketiga adalah memasukkan class BertForSequenceClassification, Trainer, TrainingArguments, EarlyStoppingCallback yang berguna untuk mendukung proses training dan fine-tuning klasifikasi IndoBERT. BertForSequenceClassification digunakan untuk tugas klasifikasi, trainer digunakan untuk loop training dan EarlyStoppingCallback (fitur dari HuggingFace untuk menghentikan proses training sebelum epoch selesai, jika performa model tidak membaik di data validasi), TrainingArguments adalah tempat untuk mengatur hyperparameter seperti jumlah epoch, output\_dir (), learning rate, dan eval\_strategy. Model BertForSequenceClassification memuat model IndoBERT pre-trained (indobert-base-p1) yang sudah dilatih di bahasa Indonesia, kemudian menggunakan num\_labels untuk menentukan jumlah output dari layer klasifikasi. Len(label2id) memberi arti bahwa output klasifikasi akan sesuai jumlahnya dengan panjang label yang sudah disediakan dan model juga akan membuat skor logits yang sama jumlahnya seperti label yang ada.

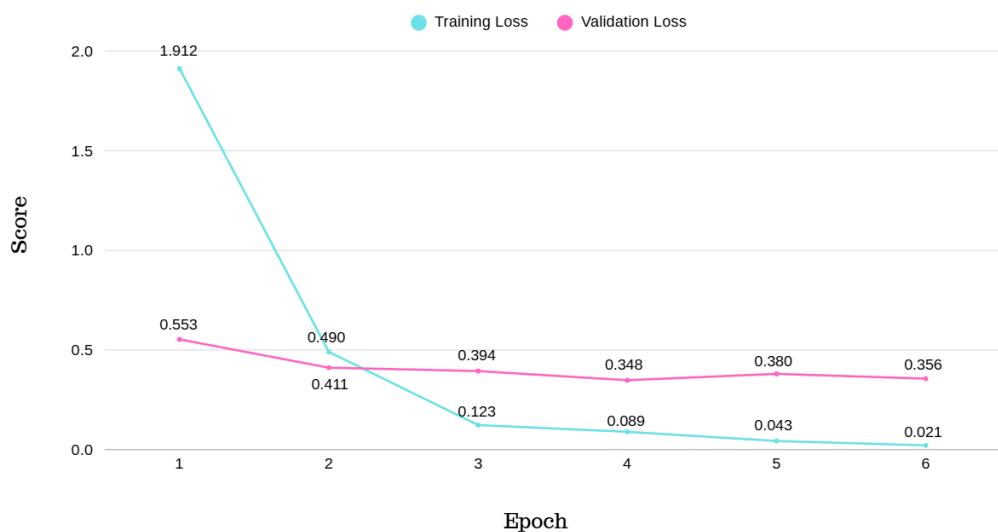
```
training_args = TrainingArguments(
    output_dir=".results",
    eval_strategy="epoch",
    save_strategy="epoch",
    save_total_limit=1,
    num_train_epochs=10,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    load_best_model_at_end=True,
    metric_for_best_model="eval_loss",
    logging_dir='./logs',
    report_to=[], # wandb OFF
```

```
    logging_steps=50,  
)
```

Training arguments diatas digunakan untuk mengatur parameter saat menggunakan Transformer. Hasil dari pelatihan akan disimpan pada folder “./results”, dan evaluasi model nantinya akan dilakukan pada setiap akhir satu epoch (eval\_strategy=”epoch”). (save\_strategy=”epoch”) berfungsi untuk menyimpan model terbaik dengan batas hanya satu model saja yang tersimpan (save\_total\_limit=1). Model selanjutnya akan dilatih selama 10 kali epoch, dengan ukuran batch sebesar 16. (load\_best\_model\_at\_end=true) artinya adalah model akan mengambil bobot terbaik saat di akhir training. Sedangkan (metric\_for\_best\_model=eval\_loss), ialah metric yang dijadikan patokan untuk memilih model terbaik yaitu eval\_loss. (logging\_dir) digunakan untuk membuat folder ./logs yang akan menyimpan catatan training.



Gambar 4. 3 Hasil Training menggunakan Epoch



Gambar 4. 4 Grafik pelatihan Epoch

Epoch yang dilakukan pada penelitian ini adalah sebanyak 10 kali epoch, namun dikarenakan adanya EarlyStoppingCallback, proses training berhenti di epoch ke-enam, menandakan bahwa selama dua epoch terakhir (Epoch 5 dan 6) tidak terjadi peningkatan performa pada model berdasarkan angka validation\_loss. Hal ini dapat disimpulkan bahwa model telah mencapai titik yang optimal (Epoch ke 4) dan epoch dihentikan lebih awal guna mencegah overfitting.

	precision	recall	f1-score	support
aplikasi	1.00	1.00	1.00	14
biaya	0.67	0.86	0.75	14
biaya_coding	0.92	0.86	0.89	14
biaya_design	1.00	1.00	1.00	14
biaya_hisensei	0.94	1.00	0.97	15
hubungikami	1.00	0.86	0.92	14
info_coding	0.93	0.87	0.90	15
info_design	1.00	0.93	0.97	15
kursuscoding	0.67	0.71	0.69	14
kursusdesign	0.93	0.87	0.90	15
kursushisensei	0.87	0.87	0.87	15
outofscope	0.93	1.00	0.97	14
pembayaran	1.00	0.67	0.80	15
pembelajaran	0.92	0.86	0.89	14
pendaftaran	0.82	0.93	0.88	15
peserta	0.81	0.93	0.87	14
speklaptop	0.88	1.00	0.93	14
trialclass	0.93	0.93	0.93	14
umum	0.93	0.87	0.90	15
accuracy			0.89	274
macro avg	0.90	0.89	0.89	274
weighted avg	0.90	0.89	0.89	274

Gambar 4. 5 Hasil Training



Gambar 4. 6 Grafik hasil training

Setelah model dilatih, didapatkan hasil performa model klasifikasi berdasarkan matriks precisions, recall, f1-score terhadap masing-masing intent. Intent yang mencapai akurasi sempurna diantaranya adalah “aplikasi”, “biaya\_design” dan “info\_design”. Sedangkan beberapa label seperti “kursuscoding”, “biaya”, masih menunjukkan nilai f1-score yang rendah, yaitu 0.69 dan 0.75, dapat disimpulkan bahwa model masih kesulitan untuk membedakan pola kalimat dari label tersebut. Nilai f1-score yang relatif rendah ini disebabkan oleh kemiripan struktur kalimat antar kelas. Namun, Secara keseluruhan, Model IndoBERT Classification memperoleh akurasi 89%. Performa yang cukup tinggi pada sebagian besar intent memberi kesimpulan bahwa model dapat mengenali konteks kalimat bahasa Indonesia dengan baik. Label dengan f1-score rendah seperti kursuscoding, dapat ditingkatkan melalui penambahan data pelatihan agar model dapat belajar membedakan pola kalimat yang mirip antar kelas.

== Classification Report ==				
	precision	recall	f1-score	support
aplikasi	1.00	0.70	0.82	10
biaya	0.56	0.90	0.69	10
biaya_coding	0.90	0.90	0.90	10
biaya_design	1.00	1.00	1.00	10
biaya_hisensei	0.77	1.00	0.87	10
hubungikami	0.89	0.80	0.84	10
info_coding	0.91	1.00	0.95	10
info_design	1.00	1.00	1.00	10
kursuscoding	1.00	0.70	0.82	10
kursusdesign	1.00	1.00	1.00	10
kursushisensei	1.00	0.90	0.95	10
outofscope	1.00	0.70	0.82	10
pembayaran	1.00	0.60	0.75	10
pembelajaran	1.00	0.90	0.95	10
pendaftaran	0.91	1.00	0.95	10
peserta	0.91	1.00	0.95	10
speklaptop	0.77	1.00	0.87	10
trialclass	1.00	1.00	1.00	10
umum	0.82	0.90	0.86	10
accuracy			0.89	190
macro avg	0.92	0.89	0.89	190
weighted avg	0.92	0.89	0.89	190

Gambar 4. 7 Hasil Testing Model IndoBERT



Gambar 4. 8 Grafik hasil testing

Setelah model selesai dilatih, langkah selanjutnya adalah melakukan testing atau pengujian menggunakan data uji sebanyak 190 data. Data uji yang digunakan berjumlah 190 data, terdiri dari 19 kelas intent dengan masing-masing 10 data per kelas. Model diuji menggunakan metrik evaluasi klasifikasi yang terdiri dari precision, recall, f1-score, dan support. Hasilnya memberitahu bahwa performa model secara umum cukup baik dan konsisten, dengan accuracy

keseluruhan sebesar 0.89 (89%) meskipun pada beberapa label seperti “biaya” dan “pembayaran” menunjukkan f1-score rendah (0.69 dan 0.75) menyimpulkan bahwa model masih sulit untuk mengenali pola pertanyaan yang ada pada label tersebut dikarenakan banyaknya kemiripan struktur kalimat dengan label lain atau kurangnya variasi dalam data uji.

#### 4.5 Proses Persiapan Deployment

Agar proses deployment model ke Telegram berjalan dengan baik dan terstruktur, terdapat beberapa file Python yang memiliki tanggung jawab spesifik sesuai dengan fungsinya. File utama yang digunakan untuk menghubungkan model klasifikasi intent dengan platform Telegram “main\_w\_dst.py.” File ini berisikan satu class yaitu IndoBERTbot dan beberapa fungsi lainnya sebagai pusat integrasi yang memanggil berbagai komponen seperti tokenizer, model IndoBERT, dan pengelolaan jawaban. Di dalamnya juga dimasukkan token bot dari BotFather, serta konfigurasi logika percakapan dengan Dialogue State Tracking agar pengguna dapat berinteraksi langsung melalui aplikasi Telegram.

#### 4.6 Dialogue State Tracking

```
class DialogManager:
    def __init__(self, answer_path):
        import json
        with open(answer_path, 'r', encoding='utf-8') as f:
            self.answers = json.load(f)["intents"]
        self.state = {}

    def update_state(self, user_id, intent):
        self.state[user_id] = {"last_intent": intent}

    def get_last_intent(self, user_id):
        return self.state.get(user_id, {}).get("last_intent", None)

    def get_response(self, intent):
        for item in self.answers:
            if item["intent"] == intent:
                return item["responses"][0]
        return "Maaf, saya tidak mengerti maksud kamu."
```

Kode pada Dialogue State Tracking (DST) diatas dibuat menggunakan Python versi 3.10 untuk persiapan deployment ke Bot Telegram. Class DialogManager merupakan implementasi dari DST, yaitu proses yang digunakan

untuk menyimpan dan melacak konteks percakapan antara pengguna dan chatbot. DST sangat penting dalam percakapan chatbot karena memungkinkan bot untuk memahami maksud pengguna tidak hanya dari satu kalimat saja, tetapi juga dari riwayat interaksi sebelumnya. Pada implementasi ini, class DialogManager membaca file answer.json yang berisi daftar intent dan respons yang relevan, lalu memuatnya ke dalam memori melalui atribut self.answers. Selanjutnya, DST menyimpan state setiap pengguna berdasarkan ID Telegram mereka. Hal ini dilakukan melalui method update\_state, yang akan mencatat intent terakhir yang dikirimkan pengguna.

Ketika pengguna mengirimkan pesan baru, method get\_last\_intent akan digunakan untuk mengambil intent terakhir tersebut agar bisa dijadikan konteks tambahan dalam proses prediksi intent berikutnya. Kemudian, untuk memberikan respons, DialogManager memiliki method get\_response, yang akan mencari respons berdasarkan intent hasil prediksi. Jika intent ditemukan di dalam daftar self.answers, maka sistem akan mengambil salah satu respons yang telah disediakan. Namun jika tidak ditemukan, sistem akan memberikan jawaban default berupa “Maaf, saya tidak mengerti maksud kamu.”

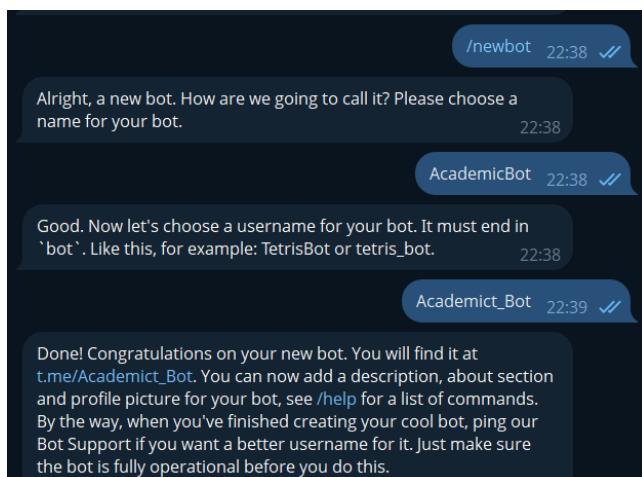
Implementasi Dialogue State Tracking ini digunakan untuk pertanyaan pengguna dengan Intent ‘biaya’. Pengguna seringkali bertanya mengenai biaya kursus namun tidak menyebutkan nama kursus secara spesifik, seperti “Berapa biaya kursusnya?”, sedangkan penyedia kursus memiliki tiga jenis kursus berbeda yaitu coding, desain dan bimbel (hisensei). Dengan Dialogue State Tracking, Chatbot akan mendeteksi intent terakhir pengguna. Bila intent terakhir pengguna di prediksi sebagai ‘Biaya’, maka Chatbot akan kembali bertanya, “Biaya kursus apa yang ingin kamu ketahui?”.

```
elif last_intent == "biaya":  
    if len(user_input_lower.split()) <= 3 and "coding" in user_input_lower:  
        intent = "biaya_coding"  
    elif len(user_input_lower.split()) <= 3 and ("desain" in user_input_lower or "design" in user_input_lower):  
        intent = "biaya_design"  
    elif len(user_input_lower.split()) <= 3 and any(k in user_input_lower for k in ["hisensei", "inggris", "mtk", "matematika", "bimbel"]):  
        intent = "biaya_hisensei"  
    else:
```

```
model_intent = self.predict_intent(user_input)
intent = resolve_intent(user_input, model_intent)
```

Kode diatas adalah sebuah percabangan if-else pada file main\_w\_dst.py, yang digunakan untuk membantu Chatbot dalam memberikan respon yang tepat setelah intent terakhir pengguna terdeteksi sebagai intent ‘biaya’. Bila ditemukan kata ‘Coding’, ‘Design’ dan ‘Hisensei’, maka chatbot akan memberikan jawaban sesuai dengan biaya yang ingin pengguna ketahui. Seperti, bila pengguna ingin mengetahui biaya coding, maka chatbot akan mendeteksi bahwa Pengguna ingin mengetahui biaya coding, lalu memberikan jawaban dengan intent ‘biaya\_coding’.

#### 4.7 Deployment Bot Telegram



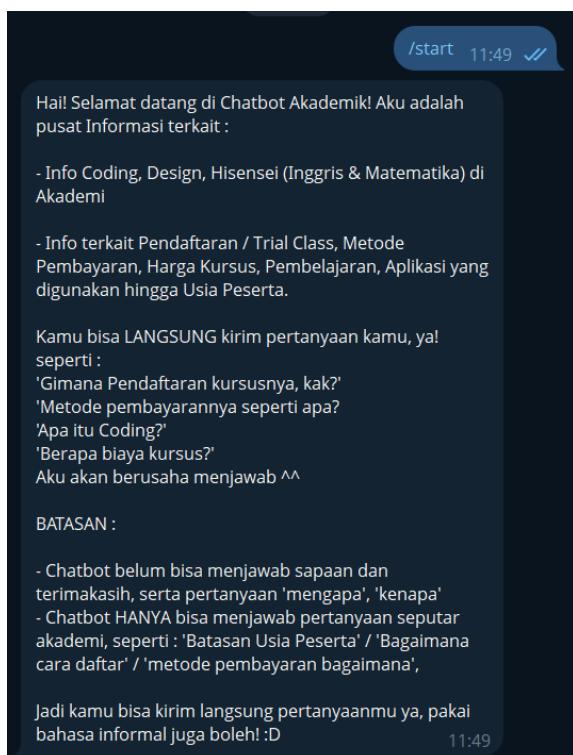
Gambar 4. 9 Pembuatan Bot di Telegram Menggunakan Bot Father

Model yang sudah dilatih kemudian akan diimplementasikan atau proses *deployment* kedalam bentuk aplikasi nyata, yaitu ChatBot Telegram dengan tujuan agar pengguna dapat langsung berinteraksi dengan model klasifikasi intent melalui Telegram. Proses *deployment* dilakukan melalui integrasi antara model IndoBERT dan Bot Telegram menggunakan API yang disediakan oleh Telegram. Sebelumnya, Bot dibuat terlebih dahulu melalui layanan resmi telegram, yaitu Bot Father. Setelah proses pembuatan bot selesai, Bot Father akan menampilkan API yang akan digunakan untuk menghubungkan kode Python dengan Telegram. Bot yang sudah dibuat dapat diakses pada link berikut ([t.me/academict\\_bot](https://t.me/academict_bot)).

#### 4.8 Testing Chatbot di Telegram Bot

Pengujian chatbot secara langsung di Telegram dilakukan untuk menilai bagaimana performa chatbot saat berinteraksi dengan pengguna sebenarnya. Pada tahap ini, chatbot yang telah dibuat dihubungkan ke bot Telegram, lalu dilakukan serangkaian percobaan dengan mengirim berbagai pertanyaan atau perintah sesuai skenario yang dirancang. Proses pengujian ini membantu melihat apakah chatbot sudah mampu memahami maksud pertanyaan (intent) pengguna dan memberikan jawaban yang sesuai serta mudah dipahami.

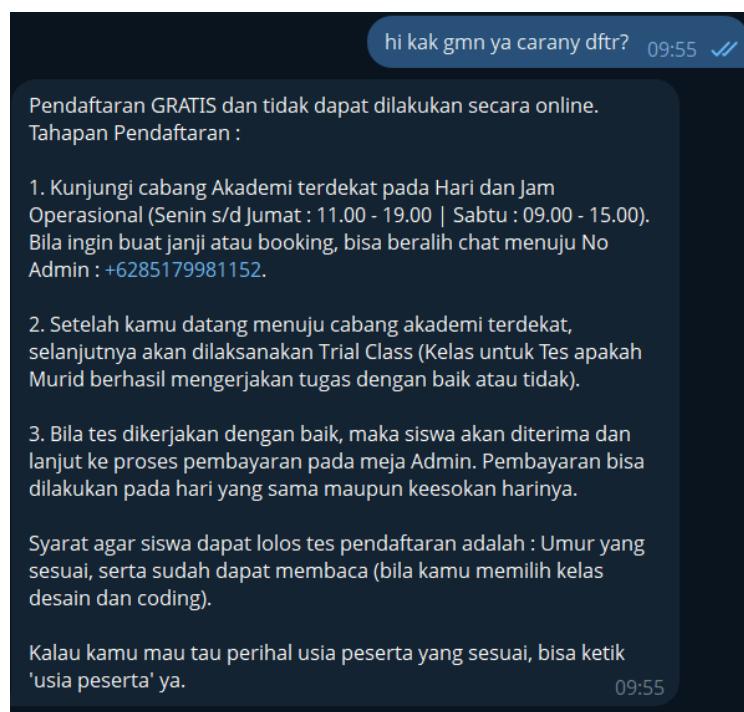
Selain itu, pengujian juga bertujuan untuk mengecek kecepatan respon, ketepatan jawaban, serta kemampuan chatbot dalam menangani variasi pertanyaan, termasuk jika terjadi kesalahan penulisan atau pertanyaan di luar skenario. Melalui uji coba langsung ini, dapat diketahui apakah chatbot sudah siap digunakan oleh pengguna akhir dan apakah masih diperlukan perbaikan lebih lanjut.



Gambar 4. 10 Tombol awal untuk memulai Chat

Gambar di atas menunjukkan tampilan awal percakapan saat pengguna memulai interaksi dengan chatbot melalui perintah /start di Telegram. Pada pesan sambutan tersebut, chatbot memperkenalkan dirinya sebagai pusat informasi mengenai berbagai layanan di Akademi, seperti coding, design, Hisensei (bahasa Inggris dan Matematika), serta informasi lain yang berkaitan dengan pendaftaran, trial class, metode pembayaran, harga kursus, hingga usia peserta. Chatbot juga memberikan contoh pertanyaan yang bisa diajukan langsung oleh pengguna, misalnya “Gimana pendaftaran kursusnya, kak?”, “Metode pembayarannya seperti apa?”, “Apa itu coding?”, atau “Berapa biaya kursus?”.

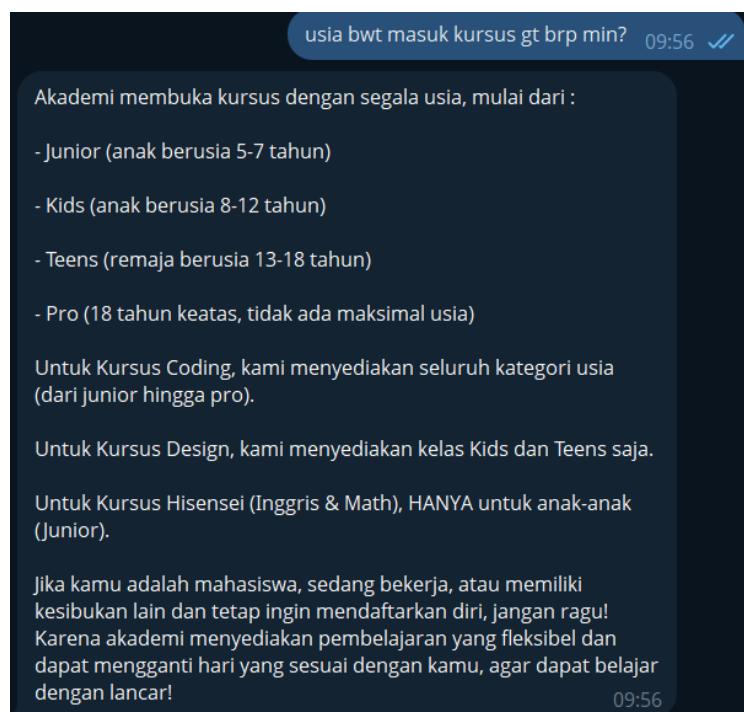
Selain itu, dijelaskan juga batasan kemampuan chatbot, yaitu chatbot belum dapat merespons sapaan, ucapan terima kasih, atau pertanyaan yang mengandung kata tanya “mengapa” dan “kenapa”. Chatbot hanya dirancang untuk menjawab pertanyaan seputar informasi akademi, seperti batas usia peserta, cara pendaftaran, dan metode pembayaran. Pada akhir pesan, chatbot juga menekankan bahwa pengguna dapat mengajukan pertanyaan dengan bahasa santai atau informal. Penjelasan ini bertujuan untuk membantu pengguna memahami jenis pertanyaan yang bisa dijawab chatbot, sehingga interaksi dapat berjalan lebih efektif dan sesuai dengan fungsinya.



Gambar 4. 11 Respon Chatbot terhadap pertanyaan pendaftaran

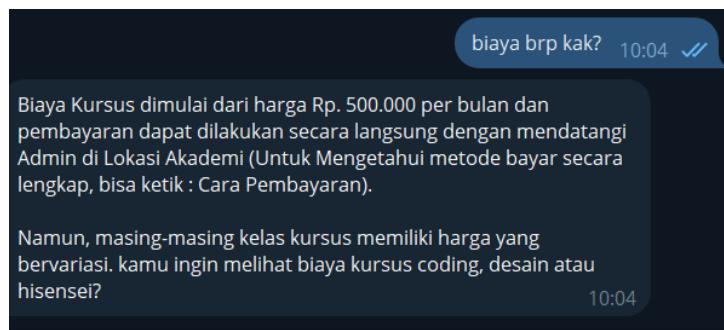
Gambar 4.9 menunjukkan contoh percakapan ketika pengguna mengirimkan pertanyaan singkat dan informal, yaitu "hi kak gmn ya carany dftr?". Meskipun pertanyaan ditulis dengan gaya santai, singkatan, dan tidak menggunakan tata bahasa formal, chatbot tetap dapat memahami maksud pengguna dan memberikan jawaban yang lengkap serta jelas. Hal ini menunjukkan bahwa chatbot mampu mendeteksi inti pertanyaan (intent) meskipun format kalimat tidak baku. Pada balasan yang diberikan, chatbot menjelaskan prosedur pendaftaran secara rinci, mulai dari kunjungan ke cabang akademi terdekat, pelaksanaan trial class, hingga proses pembayaran. Selain itu, chatbot juga memberikan informasi tambahan mengenai syarat pendaftaran dan kontak admin jika ingin melakukan booking. Respon ini memperlihatkan bahwa chatbot dirancang untuk menangani variasi gaya bahasa pengguna dan tetap dapat menyampaikan informasi secara tepat, sehingga memudahkan pengguna dalam memahami alur pendaftaran tanpa harus menyusun pertanyaan dengan format yang kaku.

Chatbot juga menyarankan arahan pada pengguna seperti “bila ingin mengetahui batasan usia, bisa mengetik ‘usia peserta’, ya.” Dengan tujuan agar Pengguna dapat terbantu dalam bertanya.



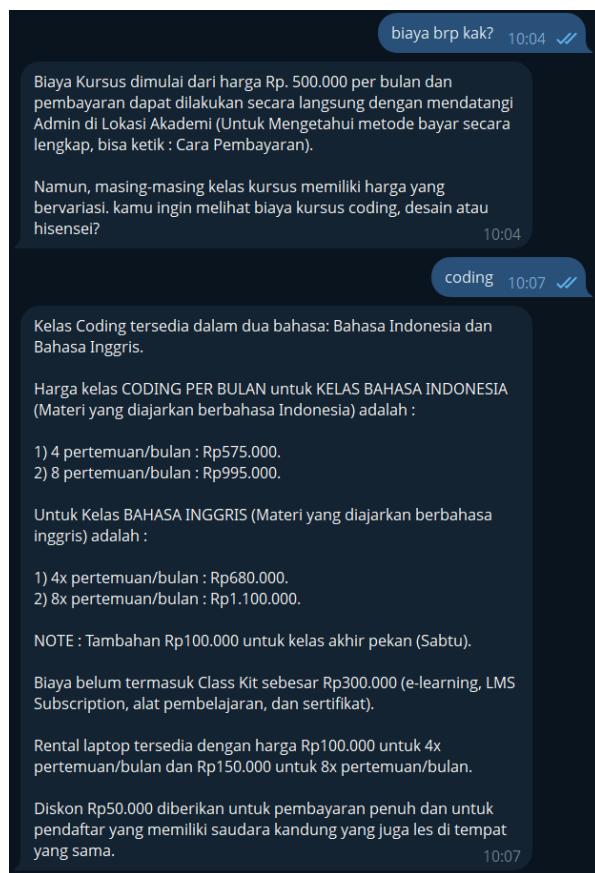
Gambar 4. 12 Respon Chatbot terhadap pertanyaan terkait Usia

Pada gambar 4.10, chatbot juga dapat merespon pertanyaan terkait usia dengan baik dan menyarankan pengguna agar tidak ragu untuk mendaftar bila mereka adalah mahasiswa ataupun yang sedang bekerja karena akademi menyediakan jam belajar yang fleksibel.



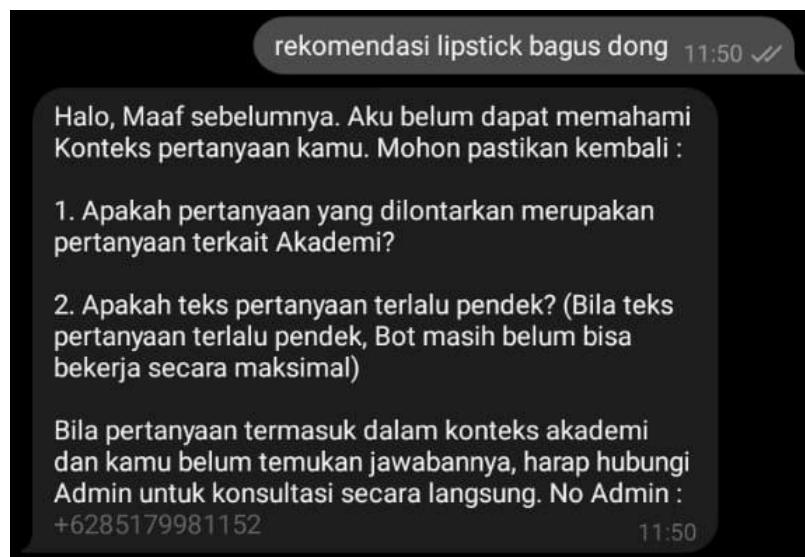
Gambar 4. 13 Chatbot menanggapi pertanyaan terkait biaya secara umum

Pada gambar 4.11, chatbot menerima pertanyaan terkait biaya secara umum tanpa menyebutkan spesifikasi kursusnya, sehingga chatbot akan mengembalikan jawaban biaya kursus secara umum yaitu “biaya kursus dimulai dari harga Rp. 500.000/bulan”. Chatbot juga bertanya kembali pada pengguna agar pengguna dapat menyebutkan biaya secara spesifik seperti biaya kursus coding, desain atau inggris dan matematika.



Gambar 4. 14 Chatbot merespon pertanyaan terkait biaya coding dengan bubblechat terpisah.

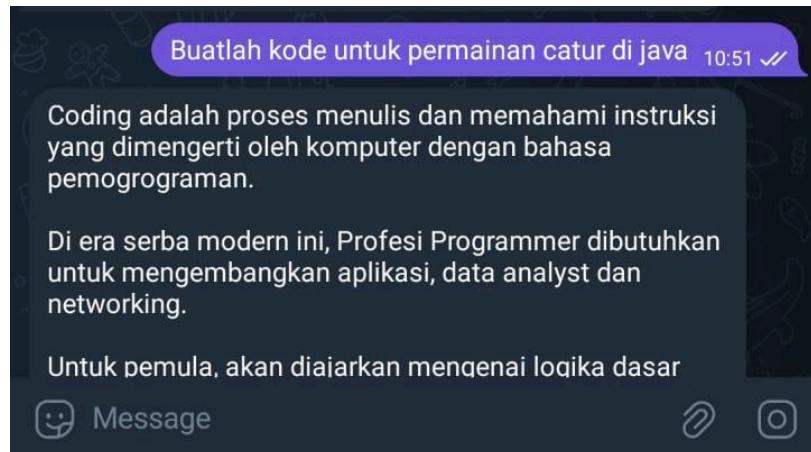
Pada gambar 4.12, Chatbot dapat menerapkan metode Dialogue State Tracking. Chatbot mengindikasikan bahwa pertanyaan user sebelumnya yaitu “biaya brp kak?” masuk kedalam intent “biaya”. Sehingga tercatat bahwa last\_intent pengguna adalah “Biaya”. Sedangkan, ketika pengguna akan memberikan jawaban lanjutan, tanpa mengetik panjang seperti “coding” saja atau “design” saja, Pengguna akan mendapatkan jawaban yang sesuai yaitu Informasi terkait biaya coding.



Gambar 4. 15 Chatbot dapat menjawab pertanyaan diluar konteks kursus

Pada Gambar diatas, Chatbot juga berhasil mendeteksi pertanyaan diluar konteks, seperti “Rekomendasi lipstick bagus dong” adalah pertanyaan yang menjurus pada tema kosmetik, dan Chatbot tidak dapat memahami pertanyaan tersebut. Cara Chatbot dapat mengetahui pertanyaan diluar konteks kursus adalah dengan membuat intent baru yakni Out of Scope yang berisi pertanyaan diluar konteks kursus. Sehingga, Model akan dapat memahami urutan kata yang termasuk pertanyaan out of scope.

Namun, seringkali chatbot masih memiliki sejumlah kekurangan yang perlu diperhatikan. Meskipun teknologi chatbot telah mampu memahami berbagai pertanyaan secara lebih natural, pada kenyataannya chatbot masih sering mengalami keterbatasan dalam menangkap maksud pengguna secara akurat, terutama ketika pertanyaan bersifat ambigu atau sangat bervariasi.



Gambar 4. 16 Chatbot salah menanggapi Pertanyaan

Pada Gambar diatas, Chatbot masih dapat merespon pertanyaan dengan lingkup yang cukup luas, pertanyaan “Buatlah kode untuk permainan catur di Java” tidak ada dalam dataset, namun terprediksi sebagai intent info\_coding dikarenakan adanya kata kunci "kode" dan "Java" yang dianggap relevan dengan topik seputar coding secara umum. Oleh karena itu, untuk menangani pertanyaan-pertanyaan yang lebih detail atau di luar cakupan dataset, chatbot ini tetap memerlukan integrasi rule-based agar dapat memberikan respon yang lebih tepat dan sesuai kebutuhan pengguna.

File bernama fallback.py dibuat untuk mengatur batasan chatbot dan berperan penting untuk menangani pertanyaan atau perintah yang tidak dapat dipahami atau tidak termasuk ke dalam intent yang telah didefinisikan pada dataset.

```
# Fallback keyword detection
FALLBACK_KEYWORDS = ["bingung", "ga ngerti", "tidak tahu", "tidak paham"]
FALLBACK_OUTOFSCOPE = ["buatkan", "contohkan", "contoh", "selesaikan", "rumus", "makan", "tuliskan", "namamu", "tulis", "mengapa", "kenapa"]
# Rule mapping keyword ke intent override
INTENT_KEYWORDS = {
    "pembayaran": ["kredit", "debit"],
    "aplikasi": ["LMS"],
}

def detect_fallback(user_text: str):
    lower_text = user_text.lower()

    # 1. Fallback jika mengandung kata bantu atau membingungkan
```

```

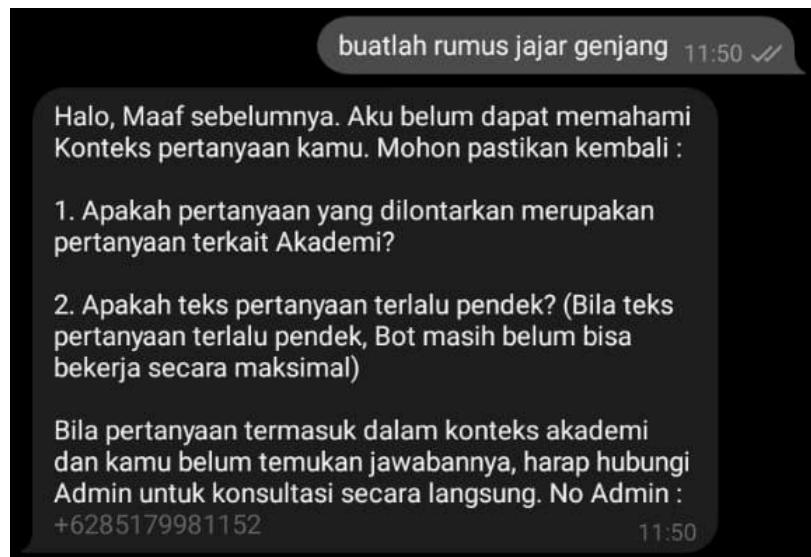
if any(k in lower_text for k in FALBACK_KEYWORDS):
    return "umum"
if any(k in lower_text for k in FALBACK_OUTOFSCOPE):
    return "outofscope"

# 2. Fallback jika input terlalu pendek
if len(lower_text.strip().split()) < 2:
    return "outofscope"

```

Salah satu contoh implementasi fallback yaitu ketika pengguna mengajukan pertanyaan yang mengandung kata "buatkanlah", "buatlah", atau kata serupa yang mengindikasikan permintaan pembuatan sesuatu seperti kode atau gambar. Jika sistem mendeteksi kata-kata tersebut, maka secara otomatis pertanyaan akan diarahkan ke intent out\_of\_scope. Intent ini digunakan sebagai penanda bahwa pertanyaan atau perintah tersebut berada di luar cakupan kemampuan chatbot yang sudah dirancang Chatbot juga mengantisipasi jika terdapat pertanyaan dengan kalimat yang terlalu pendek (kurang dari dua kata), sehingga chatbot akan mendeteksi bahwa pertanyaan tersebut terlalu singkat dan menyatakan bahwa chatbot tidak memahami maksud pertanyaan pengguna secara jelas.

Dengan cara ini, chatbot dapat menghindari memberikan jawaban yang tidak relevan atau bersifat ambigu, serta menjaga keakuratan dan konsistensi respon.



Gambar 4. 17 Chatbot merespon pertanyaan diluar konteks dengan benar.

Pendekatan ini juga membantu meningkatkan pengalaman pengguna, karena chatbot dapat secara eksplisit memberitahukan bahwa pertanyaan tersebut tidak dapat dijawab, hal tersebut lebih baik dibanding chatbot harus dipaksa menjawab dengan respon yang tidak sesuai. Selain itu, penggunaan fallback seperti ini mempermudah pengelolaan logika percakapan dan meminimalkan potensi kesalahan interpretasi.

#### 4.9 Kuisisioner

Sebagai upaya untuk mengevaluasi performa chatbot secara menyeluruh, dilakukan pengujian langsung oleh beberapa responden. Responden diminta untuk mencoba chatbot melalui platform Telegram dengan cara mengajukan pertanyaan sesuai kebutuhan. Setelah berinteraksi dengan chatbot, responden diharapkan mengisi kuesioner yang telah disediakan melalui Google Form. Kuesioner ini dirancang untuk menilai tingkat kepuasan, kemudahan penggunaan, kejelasan jawaban, serta efektivitas chatbot dalam menyampaikan informasi. Total responden yang sudah mencoba chatbot telegram ini sebanyak 39 orang, yang dimulai pada tanggal 25 Juni 2025 hingga 9 Juli 2025.

Tabel 4. 8 Daftar Pertanyaan Kuisisioner

No	Pertanyaan
1	Apakah chatbot memahami konteks pertanyaan Anda dengan baik?
2	Apakah Jawaban dari Chatbot mudah dipahami untuk Anda?
3	Apakah Chatbot dapat memahami pertanyaan berbeda, walau konteks pertanyaan sama?
4	Seberapa akurat Chatbot menjawab pertanyaan Anda?
5	Apakah Anda kesulitan selama menggunakan chatbot?
6	Chatbot memberikan respon dengan cepat setelah saya mengirim pesan.
7	Chatbot dapat memberikan jawaban yang tetap relevan meskipun saya menulis dengan gaya bahasa yang tidak formal.
8	Saat chatbot tidak memahami pertanyaan saya, ia memberikan tanggapan yang sopan dan membantu
9	Chatbot masih kurang memahami konteks pertanyaan bila gaya bahasa Informal
10	Chatbot masih menjawab pertanyaan saya, dengan jawaban yang salah juga. Padahal, saya sudah ketik dengan benar
11	Seberapa puas Anda dengan chatbot ini?
12	Apakah Anda ingin menggunakan chatbot ini lagi?

13	Apakah Anda akan merekomendasikan Chatbot ini kepada Kerabat dan Orang lain?
14	Apakah ada Kritik dan Saran untuk Chatbot ini?
15	Bila anda mengalami kendala pada chatbot, bisa ceritakan disini.

Hasil dari kuesioner akan dijadikan bahan analisis untuk mengidentifikasi kelebihan dan kekurangan chatbot, sekaligus sebagai dasar perbaikan dan pengembangan lebih lanjut agar chatbot dapat berfungsi secara optimal. Jawaban yang disediakan pada Kuisioner adalah menggunakan skala angka 1 sampai 5, dengan keterangan seperti pada tabel dibawah.

Tabel 4. 9 Tabel Keterangan Skala Kuisioner

Skala	Keterangan
1	: Sangat Tidak Setuju / Sangat Tidak Terbantu
2	: Tidak Setuju / Tidak Terbantu
3	: Ragu-ragu / Netral
4	: Setuju / Terbantu
5	: Sangat Setuju / Sangat Terbantu

Tabel 4. 10 Sebaran Jawaban Responden pada Kuisioner Testing Chatbot

No	Pertanyaan	Terbantu	Netral	Tdk Terbantu
1	Apakah chatbot memahami konteks pertanyaan Anda dengan baik?	33	6	0
2	Apakah Jawaban dari Chatbot mudah dipahami untuk Anda?	38	1	0
3	Apakah Chatbot dapat memahami pertanyaan berbeda, walau konteks pertanyaan sama?	30	6	3
4	Seberapa akurat Chatbot menjawab pertanyaan Anda?	31	7	1
5	Apakah Anda kesulitan selama menggunakan chatbot?	34	3	2
6	Chatbot memberikan respon dengan cepat setelah saya mengirim pesan.	38	1	0
7	Chatbot dapat memberikan jawaban yang tetap relevan meskipun saya menulis dengan gaya bahasa yang tidak formal.	34	4	1

8	Saat chatbot tidak memahami pertanyaan saya, ia memberikan tanggapan yang sopan dan membantu	36	3	0
9	Chatbot masih kurang memahami konteks pertanyaan bila gaya bahasa Informal	25	8	6
10	Chatbot masih menjawab pertanyaan saya, dengan jawaban yang salah juga. Padahal, saya sudah ketik dengan benar	27	7	5
11	Seberapa puas Anda dengan chatbot ini?	36	3	0
12	Apakah Anda ingin menggunakan chatbot ini lagi?	33	6	0
13	Apakah Anda akan merekomendasikan Chatbot ini kepada Kerabat dan Orang lain?	33	6	0

Untuk menghitung persentase secara keseluruhan, seluruh jawaban responden dikelompokkan sesuai dengan pertanyaannya terlebih dahulu. Seperti :

- A. Pertanyaan 1 : “Apakah chatbot memahami konteks pertanyaan Anda dengan baik?”
- 1) Total Responden = 39 Orang
  - 2) Responden yang merasa terbantu (Yang memberikan skala 4 & 5) = 33 Orang
  - 3) Responden yang merasa netral (Memberikan skala 3) = 6 Orang
  - 4) Responden yang merasa tidak terbantu (Memberikan skala 1 dan 2) = 0 Orang

Perhitungan persentase secara menyeluruh menggunakan persamaan sebagai berikut :

$$Persentase\ Kategori = \frac{Jumlah\ Responden\ pada\ Kategori}{Total\ Responden} \times 100\%$$

Maka :

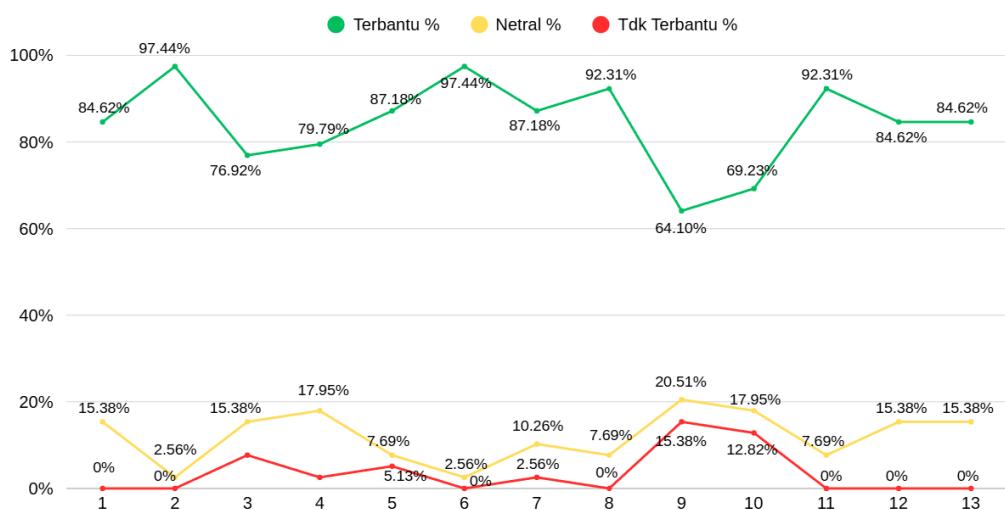
- 1) Terbantu = 33 orang / 39 orang \* 100 = 84,64%
- 2) Netral = 6 orang / 39 orang \* 100 = 15,38%
- 3) Tidak Terbantu = 0

Kesimpulannya adalah pada pertanyaan 1, sebanyak 84,64% merasa bahwa chatbot memahami konteks pertanyaan dengan baik, 15,38% memilih netral, dan tidak ada responden yang merasa bahwa chatbot tidak memahami konteks pertanyaan dengan baik. Secara keseluruhan, berikut tabel yang merepresentasikan jawaban responden pada tiap pertanyaan.

Tabel 4. 11 Tabel Keterangan Persentase Tiap Pertanyaan

No	Pertanyaan	Terbantu	Netral	Tdk Terbantu	Terbantu %	Netral %	Tdk Terbantu %
1	Apakah chatbot memahami konteks pertanyaan Anda dengan baik?	33	6	0	84,62%	15,38 %	0,00%
2	Apakah Jawaban dari Chatbot mudah dipahami untuk Anda?	38	1	0	97,44%	2,56%	0,00%
3	Apakah Chatbot dapat memahami pertanyaan berbeda, walau konteks pertanyaan sama?	30	6	3	76,92%	15,38 %	7,69%
4	Seberapa akurat Chatbot menjawab pertanyaan Anda?	31	7	1	79,49%	17,95 %	2,56%
5	Apakah Anda kesulitan selama menggunakan chatbot?	34	3	2	87,18%	7,69%	5,13%
6	Chatbot memberikan respon dengan cepat setelah saya mengirim pesan.	38	1	0	97,44%	2,56%	0,00%
7	Chatbot dapat memberikan jawaban yang tetap relevan meskipun saya menulis dengan gaya bahasa yang tidak formal.	34	4	1	87,18%	10,26 %	2,56%
8	Saat chatbot tidak memahami pertanyaan saya, ia memberikan tanggapan yang	36	3	0	92,31%	7,69%	0,00%

	sopan dan membantu						
9	Chatbot masih kurang memahami konteks pertanyaan bila gaya bahasa Informal	25	8	6	64,10%	20,51 %	15,38%
10	Chatbot masih menjawab pertanyaan saya, dengan jawaban yang salah juga. Padahal, saya sudah ketik dengan benar	27	7	5	69,23%	17,95 %	12,82%
11	Seberapa puas Anda dengan chatbot ini?	36	3	0	92,31%	7,69%	0,00%
12	Apakah Anda ingin menggunakan chatbot ini lagi?	33	6	0	84,62%	15,38 %	0,00%
13	Apakah Anda akan merekomendasikan Chatbot ini kepada Kerabat dan Orang lain?	33	6	0	84,62%	15,38 %	0,00%
TOTAL	KESELURUHAN	428	61	18	84,42%	12,03 %	3,55%
TOTAL	RATA RATA RESPONDEN	33	5	1			



Gambar 4. 18 Hasil Rangkuman Grafik Persentase Kuisioner

Berdasarkan rekapitulasi dari 13 pertanyaan, diperoleh total sebanyak 428 respon memilih "Terbantu" atau "Setuju", 61 respon memilih "Netral", dan 18 respon memilih "Tidak Terbantu" atau "Tidak Setuju". Hasil ini menunjukkan bahwa rata-rata 84% responden merasa terbantu dengan adanya chatbot, sedangkan 12% responden bersikap netral, dan hanya 3,5% responden merasa tidak terbantu. Untuk rata-rata per-orangan pada responden, yang merasa terbantu dan setuju bahwa Chatbot berjalan dengan baik adalah sebanyak 33 orang, Netral 5 orang dan responden yang merasa sangat tidak setuju adalah 1 orang. Hal ini membuktikan bahwa secara umum, chatbot telah mampu membantu pengguna dengan baik dan efektif, meskipun masih terdapat ruang untuk perbaikan agar dapat memberikan pengalaman yang lebih optimal di masa mendatang.

## **BAB V KESIMPULAN DAN SARAN**

### **5.1 Kesimpulan**

Kesimpulan yang dapat diambil mengenai Implementasi IndoBERT pada Chatbot Akademik Kursus adalah sebagai berikut:

1. Akurasi yang didapatkan berdasarkan data uji dan data latih adalah sebesar 89%. Hasil tersebut membuktikan bahwa IndoBERT sangat baik dalam memahami pertanyaan bahasa Indonesia.
2. Hasil kuisioner untuk Testing Chatbot IndoBERT dan DST kepada publik, diketahui 84,42% merasa terbantu dan setuju bahwa chatbot berjalan baik dan berfungsi sebagaimana mestinya, 12% diantaranya netral, dan 3,55% merasa tidak terbantu dengan chatbot.

### **5.2 Saran**

Berikut adalah beberapa saran yang dapat membantu dalam Implementasi IndoBERT pada Chatbot Akademik Kursus :

1. Mengimplementasikan Chatbot ke WhatsApp agar dapat dijangkau oleh masyarakat secara lebih luas.
2. Menggunakan Bot sebagai sarana atau media pendaftaran secara otomatis.

## DAFTAR PUSTAKA

- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05.
- Alhussayni, K. H., Zamyatin, A., & Alshamery, S. E. (2020). Dialogue state tracking accuracy improvement by distinguishing slot-value pairs and dialogue behaviour. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(2), 1057–1064. <https://doi.org/10.11591/ijeecs.v21.i2.pp1057-1064>
- Cahya Hutama, R., & Titi Komalasari, R. (2021). *STRING (Satuan Tulisan Riset dan Inovasi Teknologi) APLIKASI CHATBOT BERBASIS TEKS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER FAQ GRABADS*.
- Das, R. K., & Pedersen, T. (2024). *Sentiment Analysis in Twitter using BERT*. <http://arxiv.org/abs/2401.07944>
- Diantoni, C., Mufidah, R., Triana Informatika, H., Karawang, S., Ronggo Waluyo, J. H., Timur, T., Karawang, J., & Barat, I. (2024). MEMBANGUN CHATBOT UNTUK INFORMASI MAGANG DAN STUDI INDEPENDEN KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2). <https://kampusmerdeka.kemdikbud.go.id/>
- Dwiyono, A., Fachrurrozi, M., Palembang-Prabumulih, J., Ogan Ilir, K., & Selatan, S. (2024). Analisis Perbandingan Klasifikasi Intent Chatbot Menggunakan Deep Learning BERT, RoBERTa, dan IndoBERT. *Journal of Information System Research*, 6(1), 595–606. <https://doi.org/10.47065/josh.v6i1.6051>
- Erlina, E., Julyanto, J., Rustandi, J., Alexander, A., Francisco, L., Ma'muriyah, N., & Sabariman, S. (2023). Penerapan Artificial Intelligence pada Aplikasi Chatbot sebagai Sistem Pelayanan dan Informasi Online pada Sekolah. *Journal of Information System and Technology*, 4(3), 221–230. <https://doi.org/10.37253/joint.v4i3.6296>
- Indrayani, E. (2024). *Integrasi Sistem Layanan Digital*.
- Inuwa, M. (2025). *Understanding Attention Mechanisms Using Multi-Head Attention*.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. <http://arxiv.org/abs/2011.00677>
- Lin, I.-F., Hasibi, F., & Verberne, S. (2024). *Generate then Refine: Data Augmentation for Zero-shot Intent Detection*. <http://arxiv.org/abs/2410.01953>
- Maleki, F., Muthukrishnan, N., Ovens, K., Reinhold, C., & Forghani, R. (2020). Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment. In *Neuroimaging Clinics of North America* (Vol. 30, Issue 4, pp. 433–445). W.B. Saunders. <https://doi.org/10.1016/j.nic.2020.08.004>
- Maulana, A. E., Rosita, Y. D., & Sari, D. K. (2025). *Pembuatan Chatbot pada Website Persatuan Gizi Kabupaten Karawang menggunakan IndoBERT*.
- Nurizna, F. V. (2021). *PENGGUNAAN BOT TELEGRAM SEBAGAI ANNOUNCEMENT SYSTEM PADA LAYANAN PSIKOTES LEMBAGA KONSELING CAHAYA HATI*.
- Patmawati, S. (2021). *Implementasi algoritman multinomial Naive Bayes dan Chatbot layanan akademik*.
- Sahmoud, T., & Mikki, M. (2022). *Spam Detection Using BERT*.

- Siroj, R. A., Afgani, W., Fatimah, Septaria, D., Zhira, G., & Salsabila. (2024). *METODE PENELITIAN KUANTITATIF PENDEKATAN ILMIAH UNTUK ANALISIS DATA.*
- Sun, Y., Zheng, Y., Hao, C., & Qiu, H. (2022). *NSP-BERT: A Prompt-based Few-Shot Learner Through an Original Pre-training Task--Next Sentence Prediction.* <http://arxiv.org/abs/2109.03564>
- Tarumingkeng, R. C. (2024). *Natural Language Processing (NLP).*
- Venkateswaran, P., Duesterwald, E., & Isahagian, V. (2023). *DiSTRICT: Dialogue State Tracking with Retriever Driven In-Context Tuning.* <http://arxiv.org/abs/2212.02851>
- Wahyudi Harahap, D., & Fitria, L. (2020). APLIKASI CHATBOT BERBASIS WEB MENGGUNAKAN METODE DIALOGFLOW. *Jurnal Informatika Dan Teknologi Komputer*, 01 No. 01.
- Widianto, F., & Nasution, M. A. (2022). PENGARUH PERILAKU KONSUMEN DAN PERUBAHAN PASAR TERHADAP TINGKAT PENJUALAN WULING DI PT ARISTA JAYA LESTARI CABANG SM RAJA MEDAN. In *Journal Economic Management and Business* (Vol. 1, Issue 2).
- Wiharto, M. L. (2024). *SISTEM CHATBOT SEBAGAI LAYANAN TUGAS AKHIR PROGRAM STUDI TEKNIK INFORMATIKA UNISSULA MENGGUNAKAN METODE Bidirectional Encoder Representations from Transformers (BERT).*
- Wijaya, A. (2021). *Implementasi Model Neural Network IndoBERT untuk Klasifikasi Berita Difabel.*
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding.* <http://arxiv.org/abs/2009.05387>
- Xu, L., Peng, N., Zhou, D., Ng, S.-K., & Fu, J. (2024). *Chain of Thought Explanation for Dialogue State Tracking.* <http://arxiv.org/abs/2403.04656>
- Zuniananta, L. E. (2021). *PENGGUNAAN MEDIA SOSIAL SEBAGAI MEDIA KOMUNIKASI INFORMASI DI PERPUSTAKAAN.*

## LAMPIRAN

Lampiran 1 Detail Soal dan Jawaban Responden pada Kuisioner Testing Chatbot

### Form Testing Chatbot Akademik

Halo! Terima kasih sudah bersedia mencoba chatbot akademik ini! t.me/academict\_bot  
Chatbot ini dirancang untuk membantu menjawab berbagai pertanyaan seputar :

- kursus coding
- kursus desain
- kursus hisensei seperti matematika dan bahasa inggris.
- Informasi Kursus (Pendaftaran, Biaya, Metode pembayaran, Pembelajaran, Proses Trial Class, Spesifikasi Laptop, Aplikasi yang digunakan, dsb)

Melalui uji coba ini, peneliti ingin mengetahui apakah chatbot sudah mampu memberikan jawaban yang sesuai, mudah dipahami, dan bisa merespons berbagai gaya pertanyaan dari pengguna.

Silakan gunakan chatbot seperti biasa, lalu isi formulir ini berdasarkan pengalaman kamu. Masukan dari kamu sangat berarti untuk membantu kami menyempurnakan chatbot ini ke depannya. Pengisian formulir ini hanya membutuhkan waktu sekitar 3 sampai 5 menit.

[dhiazrusydanafsyi@gmail.com](mailto:dhiazrusydanafsyi@gmail.com) [Ganti akun](#)



Tidak dibagikan

\* Menunjukkan pertanyaan yang wajib diisi

#### INSTRUKSI

Gunakan Chatbot sebagai Media Layanan Informasi Penyedia Kursus dan **Posisikan dirimu adalah sebagai pendatang baru yang akan mendaftarkan diri atau mendaftarkan kerabat (anak, saudara, teman) ke Kursus ini.** Apa saja yang akan kamu tanyakan?

Contoh : "Ada kursus apa aja di Akademi?" | "Kursus coding ada?" (Boleh gunakan bahasa informal).

Target yang ingin dicapai :

- User mudah mengakses dan menggunakan chatbot
- User dapat menemukan jawaban yang akurat sesuai dengan pertanyaan
- User dapat memahami jawaban dari chatbot.

Kunjungi [t.me/academict\\_bot](https://t.me/academict_bot) di Telegram, selamat mencoba!

Nama (Boleh Samaran) \*

Jawaban Anda

Nomor Telp / Email / Akun Sosial Media \*

ex : twitter (username) | instagram (username) | Telegram (username)

Jawaban Anda

1. Apakah chatbot memahami konteks pertanyaan Anda dengan baik? \*

1    2    3    4    5

Sangat Kurang Memahami

Sangat Memahami Konteks

2. Apakah Jawaban dari Chatbot mudah dipahami untuk Anda? \*

1    2    3    4    5

Sangat Kurang Memahami

Sangat Memahami Jawaban

3. Apakah Chatbot dapat memahami pertanyaan berbeda, walau konteks pertanyaan sama? \*

1    2    3    4    5

Sangat Tidak Memahami

Sangat Bisa Memahami

4. Seberapa akurat Chatbot menjawab pertanyaan Anda? \*

1      2      3      4      5

Sangat Tidak Akurat

Sangat Akurat Sekali

5. Apakah Anda kesulitan selama menggunakan chatbot? \*

1      2      3      4      5

Sangat Sulit Sekali

Sangat Mudah Sekali

6. Chatbot memberikan respon dengan cepat setelah saya mengirim pesan. \*

1      2      3      4      5

Sangat Tidak Setuju

Sangat Setuju

7. Chatbot dapat memberikan jawaban yang tetap relevan meskipun saya menulis \* dengan gaya bahasa yang tidak formal.

1      2      3      4      5

Sangat Tidak Setuju

Sangat Setuju

8. Saat chatbot tidak memahami pertanyaan saya, ia memberikan tanggapan yang sopan dan membantu \*

1      2      3      4      5

Sangat Tidak Setuju

Sangat Setuju

9. Chatbot masih kurang memahami konteks pertanyaan bila gaya bahasa  
Informal

1      2      3      4      5

Sangat Tidak Setuju                               Sangat Setuju

10. Chatbot masih menjawab pertanyaan saya, dengan jawaban yang salah juga. \*  
Padahal, saya sudah ketik dengan benar

1      2      3      4      5

Sangat Tidak Setuju                               Sangat Setuju

1. Seberapa puas Anda dengan chatbot ini? \*

1      2      3      4      5

Sangat Tidak Puas                               Sangat Puas Sekali

2. Apakah Anda ingin menggunakan chatbot ini lagi? \*

1      2      3      4      5

Sangat Tidak Setuju                               Sangat Setuju Sekali

3. Apakah Anda akan merekomendasikan Chatbot ini kepada Kerabat dan Orang lain? \*

1      2      3      4      5

Sangat Tidak Setuju                               Sangat Setuju Sekali

4. Apakah ada Kritik dan Saran untuk Chatbot ini? \*

Jawaban Anda

5. Bila anda mengalami kendala pada Chatbot, bisa ceritakan disini. \*

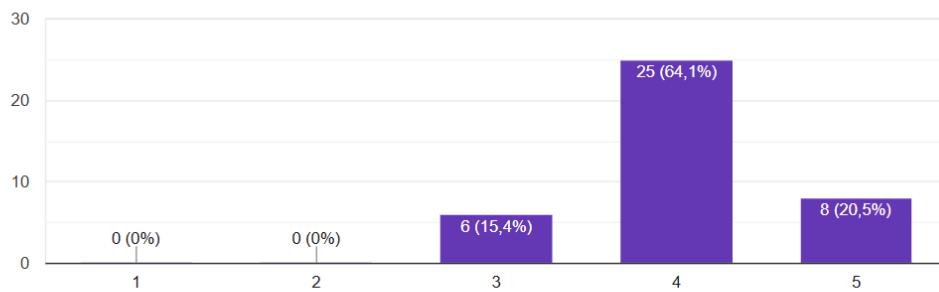
Jawaban Anda

**Form Pertanyaan : "PEMAHAMAN & KETEPATAN JAWABAN"**

1. Apakah chatbot memahami konteks pertanyaan Anda dengan baik?

 Salin diagram

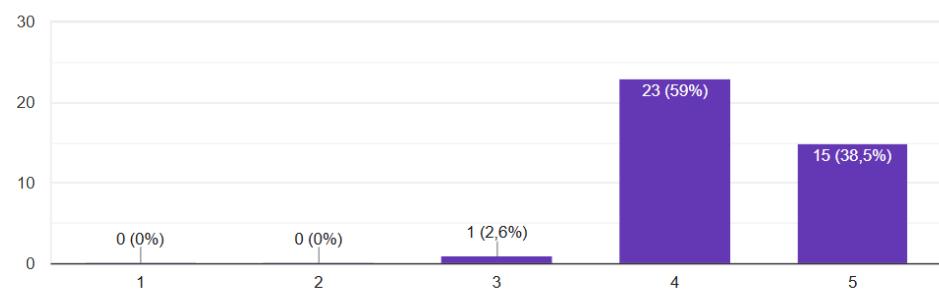
39 jawaban



2. Apakah Jawaban dari Chatbot mudah dipahami untuk Anda?

 Salin diagram

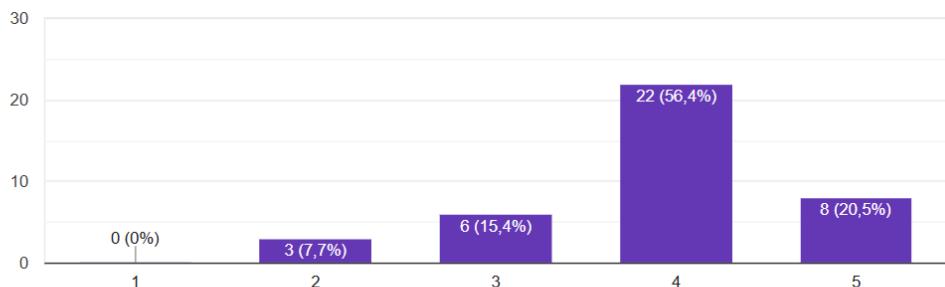
39 jawaban



3. Apakah Chatbot dapat memahami pertanyaan berbeda, walau konteks pertanyaan sama?

 [Salin diagram](#)

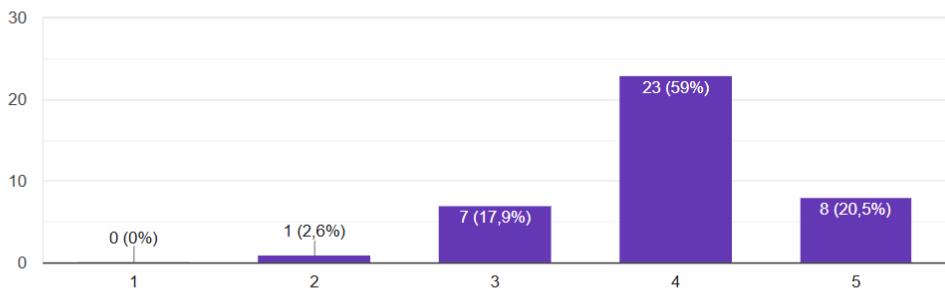
39 jawaban



4. Seberapa akurat Chatbot menjawab pertanyaan Anda?

 [Salin diagram](#)

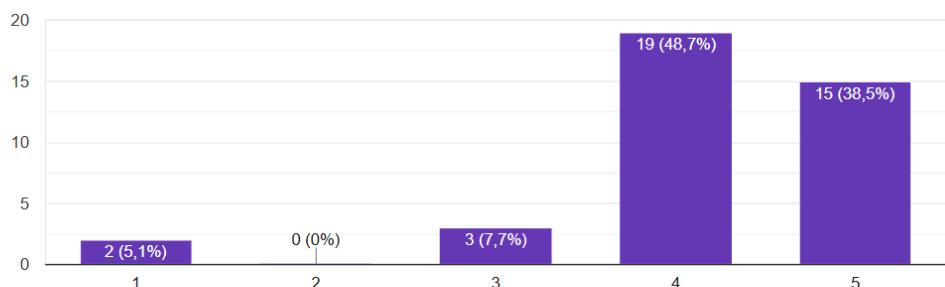
39 jawaban



5. Apakah Anda kesulitan selama menggunakan chatbot?

 [Salin diagram](#)

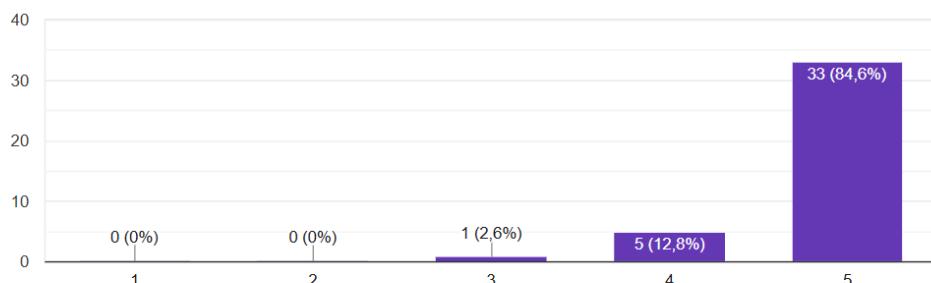
39 jawaban



6. Chatbot memberikan respon dengan cepat setelah saya mengirim pesan.

 Salin diagram

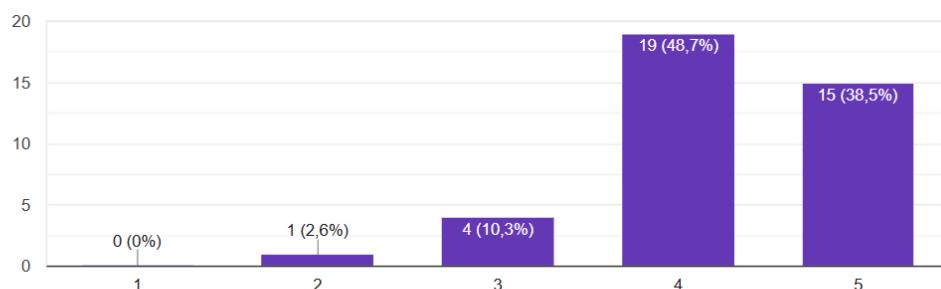
39 jawaban



7. Chatbot dapat memberikan jawaban yang tetap relevan meskipun saya menulis dengan gaya bahasa yang tidak formal.

 Salin diagram

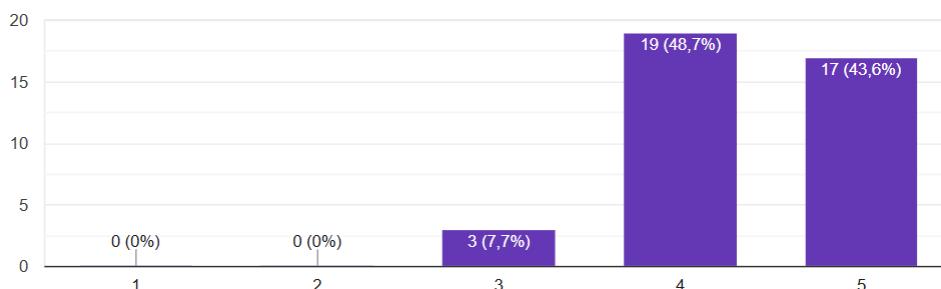
39 jawaban



8. Saat chatbot tidak memahami pertanyaan saya, ia memberikan tanggapan yang sopan dan membantu

 Salin diagram

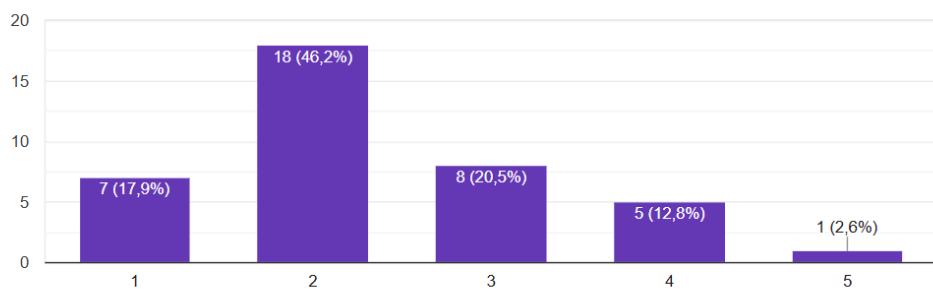
39 jawaban



9. Chatbot masih kurang memahami konteks pertanyaan bila gaya bahasa  
Informal

39 jawaban

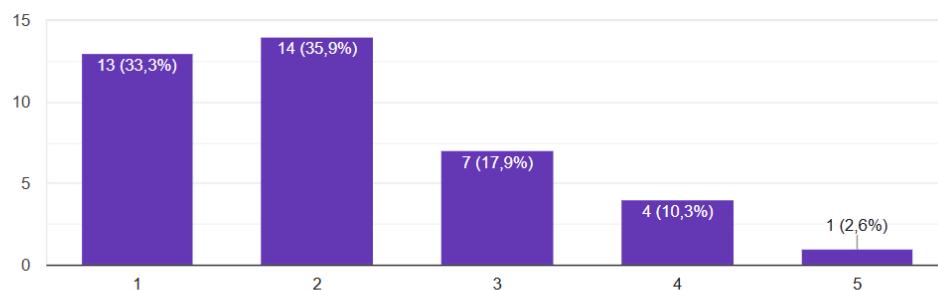
 Salin diagram



10. Chatbot masih menjawab pertanyaan saya, dengan jawaban yang salah juga. Padahal, saya sudah ketik dengan benar

39 jawaban

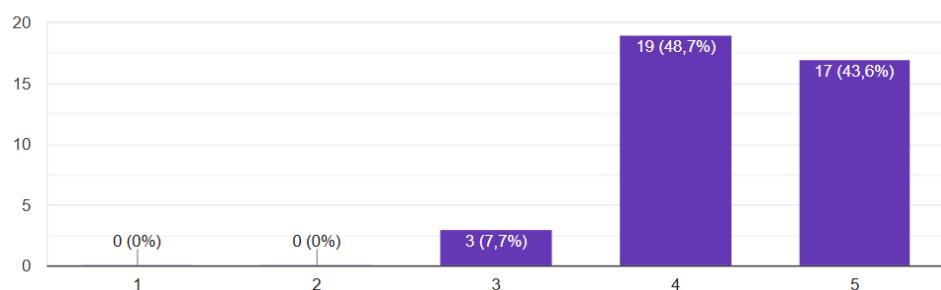
 Salin diagram



1. Seberapa puas Anda dengan chatbot ini?

39 jawaban

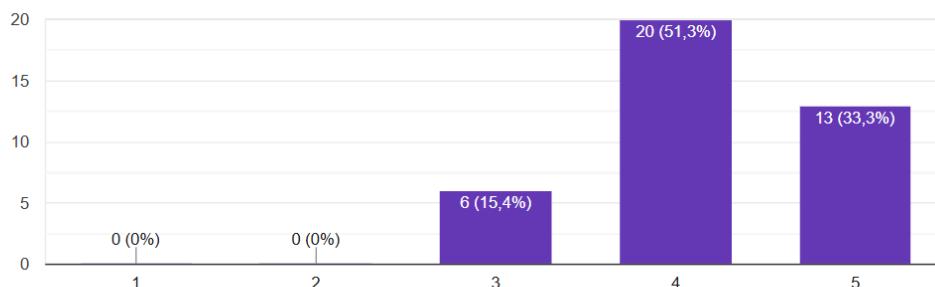
 Salin diagram



2. Apakah Anda ingin menggunakan chatbot ini lagi?

 Salin diagram

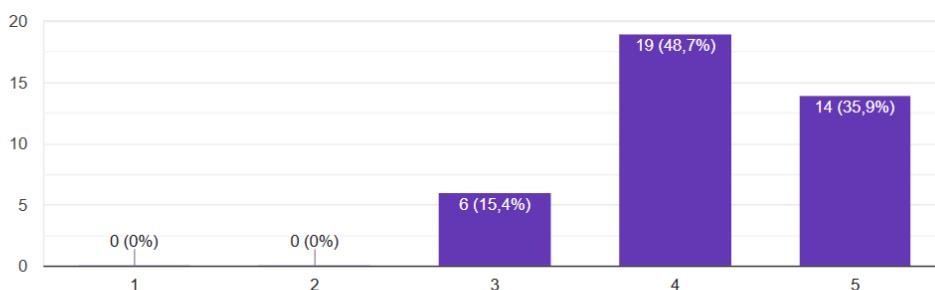
39 jawaban



3. Apakah Anda akan merekomendasikan Chatbot ini kepada Kerabat dan Orang lain?

 Salin diagram

39 jawaban



#### 4. Apakah ada Kritik dan Saran untuk Chatbot ini?

39 jawaban

kalo menurutku bot nya bisa di buat kayak misal kita send "umur 19 bisa lolos ngga?" itu jangan langsung di beri keterangan usianya, bisa di jawab dulu bisa, habis itu jelaskan kenapa bisa lolos dan kasih keterangan usianya deh.

coba pake nlu

ChatBot sudah cukup baik, namun akan lebih baik lagi jika chatbot ini dapat menjawab pertanyaan yang lebih luas

untuk pertanyaan yang out of topic, botnya masih belum paham (tadi minta rekomendasi makan siang tapi dianya bingung), kalau untuk seputar pendaftaran dan pertanyaan terkait kursusnya sudah aman 🙌

mungkin dikasih list pertanyaan/keyword yang bisa digunakan, baik mau di pembukaan bot atau ketika bot memberi tanggapan ketika pertanyaan user tidak dimengerti

Tidak ada sudah bagus 🙌

Makin ditingkatkan lagi untuk ketika menanyakan kenapa dan mengapa soalnya masih belum bisa

#### 5. Bila anda mengalami kendala pada Chatbot, bisa ceritakan disini.

39 jawaban

oke

Saya tidak mengalami kendala pada saat penggunaan ChatBot

tidak ada kendala, chatbotnya mudah dipakai dan mudah dipahami yaa

terkadang bot memberi jawaban yang tidak relevan karna pertanyaan saya menggunakan keyword lain, tapi bukan masalah besar

tidak

Sejauh ini tidak ada, terima kasih 🙌

ndak

pertanyaan di luar topik tidak bisa dipahami oleh bot

Mudah

Lampiran 2 Kartu Asistensi



**KARTU ASISTENSI  
TUGAS AKHIR**

NAMA	:	DHIAZ RUSYDA NAFSYI	SEMESTER	:	8
NIM	:	41521010163	TAHUN AKADEMIK	:	2021
FAKULTAS	:	ILMU KOMPUTER	DOSEN PEMBIMBING	:	WAWAN GUNAWAN
PROGRAM STUDI	:	TEKNIK INFORMATIKA	S.KOM., MT., M.KOM		
JUDUL KP	:	<u>IMPLEMENTASI INDOBERT DAN DIALOGUE STATE TRACKING PADA CHATBOT LAYANAN INFORMASI PENYEDIA KURSUS</u>			

NO.	TANGGAL	KETERANGAN	PARAF	NO.	TANGGAL	KETERANGAN	PARAF
1	17 MAR 2025	konsultasi tema skripsi	WJ	8	17 Juni 2025	Konsultasi terkait bab 3 : Revisi diagram tahapan Penelitian	WJ
2	25 Maret 2025	Konsultasi tema chatbot dengan IndoBERT	WJ	9.	20 Juni 2025	Konsultasi terkait hasil akhir chatbot + debugging	WJ
3	15 April 2025	Konsultasi terkait BAB 1 : Latar belakang	WJ	10	23 juni 2025	Konsultasi terkait Penulisan BAB 4	WJ
4	23 April 2025	Konsultasi terkait BAB 1 : Rumusan masalah,	WJ	11	30 juni 2025	Tanda tangan Surat persetujuan Sidang	WJ
5	2 Mei 2025	Konsultasi terkait Bab 2 : Penelitian terdahulu	WJ				
6	19 Mei 2025	Konsultasi terkait bab 2 : teori utama + Pendukung	WJ				
7	3 Juni 2025	Konsultasi terkait bab 3 : tahapan Penelitian	WJ				

## Lampiran 3 Curiculum Vitae

# CURRICULUM VITAE



**Name** : Dhiaz Rasyda Nafsyi  
**Place and date of birth** : Jakarta, 11 Oktober 2003  
**Sex** : Perempuan  
**Nationality** : Indonesia  
**Marital Status** : Belum Menikah  
**ID Number** : 3671055110030007  
**Pasport Number** : -  
**Email** : dhiazarusyda@gmail.com  
**Religion** : Islam  
**Address** : Taman Buah Puri Permata Blok D3/3 Cipondoh Makmur, Cipondoh, Kota Tangerang, Banten

### EDUCATION

<b>Elementary School</b>	: SDN PORIS GAGA 01
<b>Secondary School</b>	: SMP Muhammadiyah 4 Kota Tangerang
<b>Senior High School</b>	: MAN 2 Kota Tangerang
<b>University</b>	: Universitas Mercu Buana
<b>Degree Awarded</b>	: Sarjana Komputer
<b>Faculty</b>	: Ilmu Komputer
<b>Title of thesis</b>	: Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus
<b>GPA</b>	: 3.81

### SKILLS

**Language** : Score Toefl :

### ORGANIZATION EXPERIENCE

No. of Years	Title	Name of Organization
2024	Sekretaris	Dewan Perwakilan Mahasiswa (DPM) Fakultas Ilmu Komputer
2024	Bendahara	Japan Club Universitas Mercu Buana
2023	Anggota Riset dan Teknologi	HimTi Universitas Mercu Buana

### WORK EXPERIENCE

No. of Years	Position	Employer
2025	Coding Teacher	Timedoors Academy
2024	UI/UX Designer Internship	Writteen

### ACTIVITIES

#### NATIONAL/INTERNATIONAL

No. of Years	Position	Employer

### ACHIEVEMENT

No. of Years	Position	Employer
2022	Juara 3 Lomba Fanart Japanese Culture	Himahi Universitas Budi Luhur

### HOBBIES

**Sports** :  
**Various** : Membuat Ilustrasi 2D

## Lampiran 4 Surat Pernyataan HAKI

### **SURAT PERNYATAAN**

Yang bertanda tangan di bawah ini, pemegang hak cipta:

N a m a : 1. Wawan Gunawan S.Kom., MT., M.Kom  
2. Dhiaz Rusyda Nafsyi

Kewarganegaraan : Indonesia  
Alamat : 1. Alamat Dosen (Tidak perlu di isi oleh Mahasiswa)  
2. Taman Buah Puri Permata Blok D3/3 Cipondoh, Cipondoh  
Makmur, Kota Tangerang, Banten

Dengan ini menyatakan bahwa:

1. Karya Cipta yang saya mohonkan:  
Berupa : Laporan Penelitian  
Berjudul : **Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus**

- Tidak meniru dan tidak sama secara esensial dengan Karya Cipta milik pihak lain atau obyek kekayaan intelektual lainnya sebagaimana dimaksud dalam Pasal 68 ayat (2);
  - Bukan merupakan Ekspresi Budaya Tradisional sebagaimana dimaksud dalam Pasal 38;
  - Bukan merupakan Ciptaan yang tidak diketahui penciptanya sebagaimana dimaksud dalam Pasal 39;
  - Bukan merupakan hasil karya yang tidak dilindungi Hak Cipta sebagaimana dimaksud dalam Pasal 41 dan 42;
  - Bukan merupakan Ciptaan seni lukis yang berupa logo atau tanda pembeda yang digunakan sebagai merek dalam perdagangan barang/jasa atau digunakan sebagai lambang organisasi, badan usaha, atau badan hukum sebagaimana dimaksud dalam Pasal 65 dan;
  - Bukan merupakan Ciptaan yang melanggar norma agama, norma susila, ketertiban umum, pertahanan dan keamanan negara atau melanggar peraturan perundang-undangan sebagaimana dimaksud dalam Pasal 74 ayat (1) huruf d Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta.
2. Sebagai pemohon mempunyai kewajiban untuk menyimpan asli contoh ciptaan yang dimohonkan dan harus memberikan apabila dibutuhkan untuk kepentingan penyelesaian sengketa perdata maupun pidana sesuai dengan ketentuan perundang-undangan.
  3. Karya Cipta yang saya mohonkan pada Angka 1 tersebut di atas tidak pernah dan tidak sedang dalam sengketa pidana dan/atau perdata di Pengadilan.
  4. Dalam hal ketentuan sebagaimana dimaksud dalam Angka 1 dan Angka 3 tersebut di atas saya / kami langgar, maka saya / kami bersedia secara sukarela bahwa:
    - a. permohonan karya cipta yang saya ajukan dianggap ditarik kembali; atau
    - b. Karya Cipta yang telah terdaftar dalam Daftar Umum Ciptaan Direktorat Hak Cipta, Direktorat Jenderal Hak Kekayaan Intelektual, Kementerian Hukum Dan Hak Asasi Manusia R.I dihapuskan sesuai dengan ketentuan perundang-undangan yang berlaku.
    - c. Dalam hal kepemilikan Hak Cipta yang dimohonkan secara elektronik sedang dalam berperkara dan/atau sedang dalam gugatan di Pengadilan maka status kepemilikan surat pencatatan elektronik tersebut ditangguhkan menunggu putusan Pengadilan yang berkekuatan hukum tetap.

Demikian Surat pernyataan ini saya/kami buat dengan sebenarnya dan untuk dipergunakan sebagaimana mestinya.

Jakarta, 11 Juli 2025  
Untuk dan Atas nama  
Universitas Mercu Buana

**SURAT PENGALIHAN HAK CIPTA**

Yang bertanda tangan dibawah ini :

N a m a : 1. Wawan Gunawan S.Kom., MT., M.kom  
2. Dhiaz Rusyda Nafsyi

Alamat : 1. Alamat Dosen (Tidak Perlu Di isi Mahasiswa)  
2. Taman Buah Puri Permata Blok D3/3 Cipondoh, Cipondoh Makmur, Kota  
Tangerang, Banten

Adalah **Pihak I** selaku pencipta, dengan ini menyerahkan karya ciptaan saya kepada :

N a m a : Universitas Mercu Buana  
Alamat : Jl. Meruya Selatan No. 1. Kembangan Jakarta Barat 11650

Adalah **Pihak II** selaku Pemegang Hak Cipta berupa --- **Laporan Penelitian --- JUDUL : "Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus".** untuk didaftarkan di Direktorat Hak Cipta dan Desain Industri, Direktorat Jenderal Hak Kekayaan Intelektual, Kementerian Hukum dan Hak Asasi Manusia RI.

Demikianlah surat pengalihan hak ini kami buat, agar dapat dipergunakan sebagaimana mestinya.

Jakarta, 11 Juli 2025

Pemegang Hak Cipta  
Universitas Mercu Buana  
Rektor

Pencipta

**Prof. Dr. Ir. Andi Adriansyah, M.Eng**

Lampiran 5 Sertifikat BNSP



## Lampiran 6 Surat Ijin Riset Perusahaan

### TIMEDOOR ACADEMY

Blk. A, Rukan Columbus Jl. Green Lake City Boulevard No.5, RT.001/RW.011, Ketapang, Kec. Cipondoh, Kota Tangerang, Banten 15147 | No. Telepon : [0851-7998-1152](tel:0851-7998-1152)

No : -  
Lampiran : -  
Perihal : Surat Izin Riset Perusahaan

Tangerang, 9 Juli 2025

#### SURAT IZIN RISET PERUSAHAAN

Sehubungan dengan izin dari Fakultas Ilmu Komputer Universitas Mercu Buana, Nomor: 999/S-KET/VI/2023, hal: Izin Penyelenggaraan Penelitian tertanggal 03 November 2023, sebagai balasan resmi dari perusahaan dengan ini menyampaikan nama mahasiswa yang akan melakukan penelitian sebagai berikut:

Nama : Dhiaz Rusyda Nafsyi  
NIM : 41521010163  
Jurusan : Teknik Informatika  
Pembimbing : Wawan Gunawan S.Kom, MT., M.Kom.  
Jenjang : S1

Mahasiswa tersebut akan melaksanakan penelitian di Perusahaan ini selama satu semester untuk melengkapi data penulisan skripsi dengan judul: "Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Akademik."

Demikian surat ini dibuat untuk keperluan pengajuan izin penelitian. Semoga kegiatan penelitian ini dapat berjalan dengan lancar dan menghasilkan kontribusi yang positif.

Tangerang, 9 Juli 2025

Branch Manager



Hansen Nurgraha

## Lampiran 7 Hasil Cek Turnitin

### CEK-DHIAZRUSYDA\_-\_Dhiaz\_Rusyda.docx

#### ORIGINALITY REPORT

**1** % SIMILARITY INDEX      0% INTERNET SOURCES      0% PUBLICATIONS      1 % STUDENT PAPERS

#### PRIMARY SOURCES

**1** Submitted to Universidad Politécnica de Madrid  
Student Paper      1 %

---

Exclude quotes	On	Exclude assignment template	On
Exclude bibliography	On	Exclude matches	< 1%

## Lampiran 8 Form Revisi Dosen Pengaji

 UNIVERSITAS <b>MERCU BUANA</b>	<b>FORMULIR REVISI</b> <b>SIDANG TUGAS AKHIR (TA)</b> <b>PROGRAM STUDI TEKNIK INFORMATIKA</b> <b>FAKULTAS ILMU KOMPUTER</b> <b>UNIVERSITAS MERCU BUANA</b>	<b>Q</b>
No.Dokumen	09-2.04.20.001	Distribusi
Tanggal Efektif	29 Mei 2024	ALL

Nama\* : Dhiaz Rusdy Nafsyi  
Nomor Induk Mahasiswa\* : 41521010163  
Judul Tugas Akhir\* : Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot  
Nama Dosen Pengudi 2 : Layanan Informasi Penyedia Kursus  
\*) Wajib diisi mahasiswa : Harwikarya, Dr. M.T.

### Catatan Revisi Lainnya:

Koordinator Sidang : Hadi Santoso, Dr, S.Kom, M.Kom  
Ketua Pengumpul : Afiyati, Dr., S.Si, MT

Tanggal Sidang: 21 July 2025 TA RR3 210725	Tanggal Persetujuan Revisi:
 Tandatangan Anggota Penguji 2 <i>Ditandatangani setelah mahasiswa selesai sidang Tugas Akhir</i>	 Tandatangan Anggota Penguji 2 <i>Ditandatangani setelah mahasiswa telah menyelesaikan revisi Tugas Akhir</i>

## Lampiran 9 Hasil Cek Turnitin



072.423.4.03.01

### **PERNYATAAN SIMILARITY CHECK** */SIMILARITY CHECK STATEMENT*

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh  
*/The undersigned, hereby declare that the scientific work written by*

Nama /Name	:	Dhiaz Rusyda Nafsyi
NIM /Student ID Number	:	41521010163
Program Studi /Study Program	:	Teknik Informatika

Dengan Judul Tugas Akhir

*/The title:*

**“Implementasi IndoBERT dan Dialogue State Tracking pada Chatbot Layanan Informasi Penyedia Kursus”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

*/Similarity checks have been carried out with the Turnitin system on the date:*

Kamis, 24 Juli 2025

dengan nilai persentase sebesar :

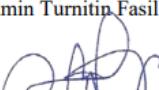
*/with a percentage value of:*

1%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*



Jakarta, 24 Juli 2025  
Admin Turnitin Fasilkom UMB

  
Agung Prawoto  
Agung Prawoto, S.Kom., B.Sc  
NIK : 322970503