

TESIS

**OPTIMASI MODEL *NATURAL LANGUAGE PROCESSING*
UNTUK APLIKASI *QUESTION-ANSWER* KESEHATAN MENTAL:
PERBANDINGAN MODEL INDOBERT DAN M-BERT**



Disusun Oleh:

Fardan Zamakhsyari

22206051005

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA
**PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER
FAKULTAS SAINS DAN TEKNOLOGI
UIN SUNAN KALIJAGA**

YOGYAKARTA

2024

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama : Fardan Zamakhsyari
NIM : 22206051005
Jenjang : Magister
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 25 Juni 2024 Saya
yang menyatakan,



Fardan Zamakhsyari

NIM : 22206051005

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan dibawah ini

Nama : Fardan Zamakhsyari
NIM : 22206051005
Jenjang : Magister
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai dengan ketentuan hukum yang berlaku.

Yogyakarta, 25 Juni 2024

yang menyatakan,



Fardan Zamakhsyari
NIM : 22206051004

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



PENGESAHAN TUGAS AKHIR

Nomor : B-1069/Un.02/DST/PP.00.9/07/2024

Tugas Akhir dengan judul : Optimasi Model Natural Language Processing Untuk Aplikasi Question-Answer Kesehatan Mental Perbandingan Model IndoBERT Dan M-BERT

yang dipersiapkan dan disusun oleh:

Nama : FARDAN ZAMAKHSYARI, S.Kom
Nomor Induk Mahasiswa : 22206051005
Telah diujikan pada : Kamis, 27 Juni 2024
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 668b54ab4f3a



Penguji I

Dr. Ir. Aulia Faqih Rifa'i, M.Kom.
SIGNED

Valid ID: 6686153d4d908



Penguji II

Dr. Ir. Bambang Sugiantoro, M.T., IPU.,
ASEAN Eng.
SIGNED

Valid ID: 6687b6314a7d5



Yogyakarta, 27 Juni 2024
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 668b92de8e5b

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

**OPTIMASI MODEL NATURAL LANGUAGE PROCESSING UNTUK
APLIKASI QUESTION-ANSWER KESEHATAN MENTAL:
PERBANDINGAN MODEL INDOBERT DAN M-BERT**

Yang ditulis oleh:

Nama : Fardan Zamakhsyari
NIM : 22206051005
Jenjang : Magister
Program Studi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Magister Informatika UIN Sunan Kalijaga untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Wassalamu'alaikum wr. wb.

Yogyakarta, 24 Juni 2024
Pembimbing


Dr. Agung Fatwanto, S.Si., M.Kom.
NIP. 19770103 200501 1 003

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

ABSTRAK

Question answering (QA) merupakan salah satu tugas dalam *Natural Language Processing* (NLP), di mana model bahasa BERT telah menunjukkan hasil yang luar biasa di bidang ini. Namun, sejauh ini belum ada penelitian yang membandingkan kinerja model BERT, yaitu M-BERT dan IndoBERT, untuk kasus kesehatan mental dengan dataset bahasa Indonesia. Penelitian ini bertujuan untuk menilai seberapa baik IndoBERT dan M-BERT dalam menyelesaikan tugas QA menggunakan dataset berbahasa Indonesia dengan topik kesehatan mental. Penelitian ini dilakukan dengan membandingkan kinerja model IndoBERT dan M-BERT, baik sebelum maupun sesudah proses *fine-tuning*, untuk tugas QA di domain kesehatan mental. Dataset yang digunakan adalah terjemahan dari dataset Amod/mental_health_counseling_conversations di hugging face. Kinerja model-model ini dievaluasi menggunakan BERTScore. Hasil penelitian ini menunjukkan bahwa kinerja IndoBERT mengungguli M-BERT setelah proses *fine-tuning*, dengan F1-BERTScore 91,8%, recall BERTScore 89,9%, dan precision BERTScore 93,9%, dibandingkan dengan M-BERT yang hanya mendapatkan F1-BERTScore 79,2%, recall BERTScore 73,4%, dan precision BERTScore 86,2%. Temuan ini menyoroti perlunya menggunakan model khusus bahasa, seperti IndoBERT untuk bahasa Indonesia, agar dapat meningkatkan kinerja dan relevansi respons dalam sistem QA. Selain itu, penelitian ini mendemonstrasikan efektivitas metode *fine-tuning* dalam meningkatkan kinerja model di mana dalam hal ini IndoBERT meningkat sebesar 28% sementara MBERT meningkat sekitar 5%. Peningkatan yang lebih tinggi pada IndoBERT menunjukkan bahwa model yang dilatih untuk jenis bahasa tertentu (dalam hal ini bahasa Indonesia) dapat meningkat secara signifikan ketika dioptimalkan untuk tugas NLP dalam bahasa tersebut.

Kata Kunci : *Natural Language Processing*, *Question Answer*, BERT, IndoBERT, MBERT

ABSTRACT

Question answering (QA) is one of the tasks in natural language processing (NLP) where the BERT language model has shown remarkable results in this field. However, so far there has been no research comparing the performance of BERT models, namely M-BERT and IndoBERT, for the case of the mental health domain with the Indonesian dataset. This study aims to assess how well IndoBERT and M-BERT perform in completing QA tasks using Indonesian-language datasets with mental health topics. This research is conducted by comparing the performance of IndoBERT and M-BERT models, both before and after fine-tuning, for QA tasks of the mental health domain. The dataset used is a translation of the Amod/mental_health_counseling_conversations dataset on the hugging face. The performance of these models was evaluated using BERTScore. The results of this study show that the performance of IndoBERT outperforms M-BERT after the fine-tuning process, with F1-BERTScore of 91.8%, recall BERTScore of 89.9%, and precision BERTScore of 93.9%, compared to M-BERT which only gets F1-BERTScore of 79.2%, recall BERTScore of 73.4%, and precision BERTScore of 86.2%. These findings highlight the need to use language-specific models, such as IndoBERT for Indonesians, to improve the performance and relevance of responses in question-answer systems. In addition, this study demonstrates the effectiveness of fine-tuning methods in improving model performance in this case the IndoBERT improved by 28% while MBERT improved by about 5%. The higher improvement of IndoBERT shows that models trained for a specific type of language (in this case the Indonesian language) can improve significantly when optimized for NLP tasks in that specific language.

Keywords : *Natural Language Processing, Question Answer, BERT, IndoBERT, MBERT*

KATA PENGANTAR

Bissmillahirrahmanirrahim,

Assalamualaikum warahmatullahi wabarakatuhu

Alhamdulillahilladzi bini'matihi tatimmush sholihaat, segala puji bagi Allah *Subhanahu wata'ala* yang telah melimpahkan segala rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan tesis dengan judul “Optimasi Model *Natural Language Processing* Untuk Aplikasi *Question-Answer* Kesehatan Mental: Perbandingan Model IndoBERT Dan M-BERT” sebagai salah satu persyaratan mencapai gelar Magister Informatika. Penulis mengucapkan terima kasih kepada semua orang yang berkontribusi pada penyelesaian tesis ini. Ucapan terima kasih tersebut secara khusus disampaikan kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Dr. Bambang Sugiantoro, M.T., selaku Ketua Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Bapak Dr. Agung Fatwanto, S.Si., M.Kom., selaku dosen Pembimbing Tesis yang telah memberikan arahan selama studi sehingga dapat menyelesaikan penyusunan tesis ini.
5. Bapak Ibu dosen dan staf karyawan Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang sudah membagi ilmu yang bermanfaat.
6. Orang tua yang senantiasa memberikan do'a selama studi di

Magister Informatika Universitas Islam Negeri Sunan Kalijaga
Yogyakarta.

7. Teman-teman Magister Informatika angkatan 2022 Universitas
Islam Negeri Sunan Kalijaga.

Akhirnya penulis hanya dapat bersyukur kepada Allah SWT semoga yang telah dilakukan selama ini dapat menjadi amal dan bekal di akhirat nanti. Penulis menyadari masih banyak kesalahan dan kekurangan dalam penulisan tesis ini, maka dari itu saran dan kritik sangat diharapkan demi perbaikan. Semoga tesis ini dapat bermanfaat bagi penulis sendiri khususnya dan bagi pembaca pada umumnya, terimakasih.

Wassalamualaikum warahmatullahi wabarakatuhu

Yogyakarta, 23 Juni 2024



Fardan Zamakhsyari

22206051005

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO



“Solum nati sumus non nobis”
(Kita tidak dilahirkan untuk diri kita sendiri)

- Marco Tulio Cicero-



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

ABSTRAK	I
ABSTRACT	VI
KATA PENGANTAR	VII
MOTTO	IX
DAFTAR ISI	X
DAFTAR TABEL	XII
DAFTAR GAMBAR	XIII
BAB I	1
PENDAHULUAN	1
A. LATAR BELAKANG.....	1
B. RUMUSAN MASALAH	5
C. BATASAN MASALAH	5
D. TUJUAN PENELITIAN	6
E. MANFAAT PENELITIAN.....	6
F. KEASLIAN PENELITIAN.....	7
BAB II	8
TINJAUAN PUSTAKA DAN LANDASAN TEORI	8
A. TINJAUAN PUSTAKA.....	8
B. LANDASAN TEORI	16
1. <i>Machine Learning dan Deep Learning</i>	16
2. <i>Natural Language Processing</i>	19
3. <i>Transformers Model</i>	20
4. <i>Bidirectional Encoder Representations from Transformers</i>	21
a. IndoBERT	24
b. Multilingual BERT	25
5. <i>Question Answering</i>	26
6. <i>BERTScore</i>	27
BAB III	29
METODOLOGI PENELITIAN	29
A. METODOLOGI PENELITIAN.....	29
B. ALAT DAN BAHAN	29
1. <i>Hardware</i>	29

2. <i>Software</i>	30
3. <i>Dataset</i>	30
C. TAHAPAN PENELITIAN	31
1. <i>Pemilihan Model Transformer</i>	32
a. <i>IndoBERT</i>	32
b. <i>Multilingual BERT</i>	33
2. <i>Pre-Processing Data</i>	34
3. <i>Processing Data</i>	34
4. <i>Evaluasi Model</i>	36
BAB IV	40
HASIL DAN PEMBAHASAN	40
A. PENGUMPULAN DATASET	40
1. <i>Dataset</i>	40
2. <i>Translasi Dataset</i>	41
B. PRE-PROCESSING DATA	42
1. <i>Tokenization pertanyaan dan jawaban</i>	43
2. <i>Penambahan Kolom `start_positions` dan `end_positions`</i>	44
C. PROCESSING DATA.....	45
1. <i>Optimasi Model Bahasa</i>	45
2. <i>Fine Tuning</i>	47
D. EVALUASI MODEL	51
E. PEMBAHASAN	56
BAB V	59
PENUTUP	59
DAFTAR PUSTAKA	61

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	12
Tabel 3. 1 Variabel Optimasi Model Bahasa	35
Tabel 4. 1 Hasil Pengujian BERTScore	56



DAFTAR GAMBAR

Gambar 2. 1 Venn diagram konsep machine learning dan deep learning (Christian Janiesch)	17
Gambar 2. 2 Prosedur pre-training dan fine-tuning (Devlin et al., 2019)	22
Gambar 2. 3 Ilustrasi penghitungan BERTScore (T. Zhang et al., 2020)	28
Gambar 4. 1 Hasil Translasi Dataset	42
Gambar 4. 2 BERT Tokenizer untuk IndoBERT dan MBERT	43
Gambar 4. 3 Penambahan <i>start_positions</i> dan <i>end_positions</i> pada dataset	44
Gambar 4. 4 Hyperparameter yang digunakan pada model IndoBERT	45
Gambar 4. 5 Hyperparameter yang digunakan pada model MBERT	47
Gambar 4. 6 Tokenisasi di tahap fine tuning pada model IndoBERT ...	48
Gambar 4. 7 Tokenisasi di tahap fine tuning pada model MBERT	49
Gambar 4. 8 Konversi <i>start_positions</i> dan <i>end_positions</i> menjadi tensor di tahap <i>fine tuning</i> pada model IndoBERT	50
Gambar 4. 9 Konversi <i>start_positions</i> dan <i>end_positions</i> menjadi tensor di tahap <i>fine tuning</i> pada model MBERT	50
Gambar 4. 10 Model training untuk tugas QA di tahap fine tuning pada model IndoBERT	51
Gambar 4. 11 Model training untuk tugas QA di tahap fine tuning pada model MBERT	51
Gambar 4. 12 Load Model untuk evaluasi pada model IndoBERT	52
Gambar 4. 13 Load Model untuk evaluasi pada model MBERT	52
Gambar 4. 14 Mendefinisikan dan memodifikasi fungsi <i>get_answer</i>	52
Gambar 4. 15 Mendefinisikan evaluasi menggunakan BERTScore	54

Gambar 4. 16 Mendefinisikan dan memodifikasi fungsi give_an_answer 54

Gambar 4. 17 Melakukan iterasi menggunakan dataset yang dipilih 55



BAB I

PENDAHULUAN

A. LATAR BELAKANG

Kesehatan mental merupakan komponen mendasar dari definisi kesehatan. Kesehatan mental yang baik memungkinkan orang untuk menyadari potensi mereka, mengatasi tekanan kehidupan yang normal, bekerja secara produktif, dan berkontribusi pada komunitas mereka (Ayuningtyas et al., 2018). Kesehatan mental juga mencakup pada cara mengendalikan diri dari tekanan atau stress, dalam bersosialisasi, dalam penyesuaian diri serta dalam pengambilan keputusan (Yasmin et al., 2022).

Menurut WHO regional Asia Pasifik (WHO SEARO) jumlah kasus gangguan depresi terbanyak di India (56.675.969 kasus atau 4,5% dari jumlah populasi), terendah di Maldives (12.739 kasus atau 3,7% dari populasi). Adapun di Indonesia sebanyak 9.162.886 kasus atau 3,7% dari populasi (Ayuningtyas et al., 2018). Masyarakat Indonesia mengalami masalah psikologis yang dikategorikan sedang, berat, dan sangat berat, masing-masing sebesar 12,4%, 26,3%, dan 16%, selama pandemi COVID-19 di Indonesia (Kelana et al., 2023).

Dalam upaya penanganannya, pemanfaatan teknologi yang digunakan dalam praktik ilmu bimbingan dan konseling bersifat virtual dengan basis internet bisa dilakukan. sehingga proses konseling dapat dilakukan dari jarak jauh (Yulia, 2023). Dengan adanya smarhphone yang fleksibel dan merupakan teknologi yang hampir dimiliki semua orang, aplikasi kesehatan mental dapat menjadi alternatif yang menjanjikan untuk promosi dan pencegahan kesehatan mental di

kalangan masyarakat (Götzl et al., 2022). Sekitar 87% masyarakat Indonesia menggunakan aplikasi untuk telekonsultasi layanan kesehatan mental karena aksesibilitasnya, sementara 63% memilih aplikasi tersebut karena harganya yang terjangkau (Kelana et al., 2023).

Chatbot *artificial intelligence* (AI) adalah sistem komputasi cerdas interaktif, yang menyelesaikan tugas dalam kombinasi dengan manusia atau secara mandiri, menggunakan *big data*, *Natural Language Processing* (NLP), dan algoritma *Machine Learning* (ML) untuk memperluas kemampuan mereka, meningkatkan produktivitas, serta menyediakan percakapan, panduan, dan dukungan (S. Sharma & Chaudhary, 2023).

Aplikasi Chatbot merupakan salah satu bentuk aplikasi *Natural Language Processing* (NLP). NLP sendiri merupakan salah satu bidang ilmu Kecerdasan Buatan yang mempelajari komunikasi antara manusia dan komputer melalui bahasa alami (Mathew et al., 2019). Chatbot dapat memproses input pengguna menggunakan pendekatan NLP, Ketika pengguna mengirim pesan, pemrosesan teks akan dilakukan. Pemrosesan teks dilakukan dengan menggunakan NLP dan kemudian mengasosiasikan input dengan maksud, untuk menghasilkan output yang dapat memproses bahasa manusia (Kumar & Ali, 2020).

Pada penelitian kali ini, akan menelaah mengenai model *question answer* yang berupa *learning-based method* dimana dalam model tersebut mempelajari pola dari data training, yang berisi pertanyaan dengan jawaban, melalui model *deep learning* untuk menghasilkan jawaban yang relevan. Pendekatan ini biasanya didasarkan pada *Seq2Seq learning model* (Kumar, 2015). Dalam kasus kalimat dan percakapan yang panjang (lebih dari 20 kata), semua informasi penting

dari kalimat sumber harus dikompresi menjadi vektor dengan panjang tetap. Model yang disarankan adalah *question answer* dengan model *Bidirectional Encoder Representations from Transformers* (BERT) (Esfandiari et al., 2023).

BERT merupakan penyempurnaan terbaru dari serangkaian *neural models* yang memanfaatkan *pre-training* secara substansial dan telah banyak membantu dalam berbagai NLP tasks, mulai dari klasifikasi teks hingga tugas-tugas seperti *question-answer* dari dataset (Alzubi et al., 2023). BERT menggunakan arsitektur Transformer, yang terdiri dari beberapa lapisan encoder untuk mempelajari representasi kata. Transformer dianggap luar biasa dalam NLP karena memanfaatkan *attention mechanism* untuk memproses data berurutan, seperti kata-kata dalam teks. *Attention mechanism* memungkinkan model untuk secara efektif menangkap hubungan antara kata-kata yang berjauhan, memungkinkan pemahaman yang lebih baik tentang konteks dalam teks (Yunanto et al., 2023).

Kerangka kerja BERT memiliki 2 langkah: *pre-training* dan *fine-tuning*. Selama *pre-training*, model dilatih pada data yang tidak berlabel melalui tugas *pre-training* yang berbeda. Untuk *fine-tuning*, model BERT diinisialisasi terlebih dahulu dengan parameter yang telah dilatih sebelumnya, dan semua parameter tersebut di-*fine-tuning* dengan menggunakan data berlabel dari tugas-tugas sebelumnya. Setiap tugas memiliki model fine-tuned yang terpisah, meskipun diinisialisasi dengan parameter *pre-training* yang sama (Devlin et al., 2019).

Penelitian ini menggunakan dua varian model BERT yang secara khusus dikembangkan untuk *multilingual language* dan bahasa Indonesia, yaitu *indoBERT base uncased* (Rifky/Indobert-QA) dan

BERT *multilingual base* model (cased) (google-bert/bert-base-multilingual-cased). Untuk membandingkan BERT model, peneliti membandingkannya dengan dua model BERT yang sudah ada sebelumnya: Multilingual BERT ("mBERT"), dan BERT monolingual untuk bahasa Indonesia ("IndoBERT"). MBERT dilatih dengan menggabungkan dokumen Wikipedia untuk 104 bahasa, termasuk bahasa Indonesia, dan efektif untuk tugas-tugas multibahasa yang tidak memerlukan banyak penjelasan (Koto, Rahimi, et al., 2020).

IndoBERT merupakan *transformer-based* model dengan BERT (Devlin et al., 2019), model bahasa ini dilatih menggunakan framework Huggingface, mengikuti konfigurasi default dari BERT base (uncased) (Koto et al., 2021). Sedangkan Multilingual-BERT (mBERT) adalah versi dari BERT yang dilatih dalam 104 bahasa, termasuk bahasa Indonesia, dengan dataset Wikipedia terbesar menggunakan masked language modeling (MLM). Dalam menangani data yang tidak seimbang, *oversampling* dilakukan pada bahasa-bahasa kecil dan *undersampling* untuk bahasa-bahasa besar. Hal ini memungkinkan model mBERT untuk digunakan pada dataset bahasa dengan sumber daya rendah seperti bahasa Indonesia (B. Zhang et al., 2020). Kedua model ini sama-sama memiliki konfigurasi dari BERT, 12 hidden layers masing-masing 768, 12 attention head, dan feed-forward hidden layers sebanyak 3.072.

Untuk mengevaluasi dua model yang berbeda, peneliti akan menggunakan dataset *Amod/mental_health_counseling_conversations* yang diambil dari situs huggingface. Dataset ini adalah kumpulan pertanyaan dan jawaban yang bersumber dari dua platform konseling dan terapi online. Pertanyaan-pertanyaan tersebut mencakup berbagai topik kesehatan mental, dan jawabannya diberikan oleh psikolog yang

berkualifikasi. Dataset ini dimaksudkan untuk digunakan untuk fine-tuning language models. Dataset yang diambil masih dalam bahasa Inggris, sehingga peneliti melakukan translasi manual menjadi bahasa Indonesia agar pengujian lebih relevan. Evaluasi dataset akan menggunakan BERT-Score. Metrik evaluasi ini sangat berguna untuk mengevaluasi kinerja algoritma klasifikasi (Alzubi et al., 2023).

Penelitian eksperimental ini bertujuan untuk menunjukkan efisiensi dari penggunaan model M-BERT dan IndoBERT pada tugas *question answer* yang berfokus pada kesehatan mental. Dengan dilakukannya penelitian ini diharapkan dapat mengetahui relevansi model BERT yang sesuai, efisien, dan terukur untuk digunakan dalam *question answer*. Kontribusi lain dari penelitian ini adalah; i) model IndoBERT dan MBERT yang dioptimalkan untuk tugas QA domain kesehatan mental dalam bahasa Indonesia, dan ii) terjemahan dataset Amod/mental_health_counseling_conversations ke dalam bahasa Indonesia.

B. RUMUSAN MASALAH

Penelitian ini berangkat dari permasalahan terkait performa model BERT yang digunakan untuk *task question answer*. Sehingga penelitian ini ditujukan untuk menjawab masalah terkait bagaimana efisiensi performa antara IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model (cased)* (google-bert/bert-base-multilingual-cased) pada *task question answer* dengan dataset terkait *mental health?*.

C. BATASAN MASALAH

Untuk menjaga penelitian yang terarah dan sesuai dengan tujuan

utama, yaitu terkait model BERT yang digunakan untuk *task question answer* dengan menggunakan dataset *mental health*. Sehingga penelitian ini memiliki beberapa batasan masalah, diantaranya;

1. Pemrograman menggunakan bahasa python dengan *tools Google Collab*.
2. Menggunakan IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model (cased)* (google-bert/bert-base-multilingual-cased).
3. Pengujian dilakukan menggunakan metode BERT-Score.

D. TUJUAN PENELITIAN

Sesuai dengan uraian rumusan masalah dan batasan masalah diatas, penelitian ini ditujukan untuk membandingkan performa dari IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model (cased)* (google-bert/bert-base-multilingual-cased) pada *task question answer* dengan dataset terkait *mental health* yang akan diuji dengan menggunakan metode BERT-Score.

E. MANFAAT PENELITIAN

Dari penelitian yang dilakukan diharapkan memberikan akan memberikan manfaat sebagai berikut :

1. Memberikan gambaran terkait model IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model (cased)* (google-bert/bert-base-multilingual-cased) pada *task question answer* dengan dataset terkait *mental health*.
2. Mengetahui perbandingan performa IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model*

- (*cased*) (google-bert/bert-base-multilingual-cased) pada *task question answer* dengan dataset terkait *mental health*.
3. Dapat dijadikan sebagai acuan penggunaan Model BERT untuk tugas *question answer*.
 4. Dapat dikembangkan dan dimanfaatkan dalam bidang konseling kesehatan mental.

F. KEASLIAN PENELITIAN

Dalam peninjauan literatur dan observasi dari beberapa research, Penelitian yang berkaitan dengan perbandingan optimasi model BERT yang ditujukan untuk mengetahui perbandingan performa antara IndoBERT *base uncased* (Rifky/Indobert-QA) dan BERT *multilingual base model (cased)* (google-bert/bert-base-multilingual-cased) pada *task question answer* dengan dataset terkait *mental health*, belum pernah dilakukan.

BAB V

PENUTUP

A. Kesimpulan

Tujuan dari penelitian ini adalah untuk membandingkan performa model BERT dalam tugas *question answer*, khususnya model IndoBERT dan M-BERT, dengan menggunakan dataset bahasa Indonesia dengan topik kesehatan mental. IndoBERT mengungguli M-BERT dalam hal performa, dengan F1-BERTScore sebesar 91,8%, *recall* BERTScore sebesar 89,9%, dan *precision* BERTScore sebesar 93,9%, dibandingkan dengan M-BERT yang mendapatkan F1-BERTScore sebesar 79,2%, *recall* BERTScore sebesar 73,4%, dan *precision* BERTScore sebesar 86,2%. Temuan ini menyoroti perlunya menggunakan model khusus bahasa, seperti IndoBERT untuk bahasa Indonesia, untuk meningkatkan akurasi dan relevansi respons dalam sistem *question answer*. Selain itu, penelitian ini menunjukkan efektivitas metode *fine-tuning* dalam meningkatkan kinerja model. Setelah dilakukan *fine-tuning*, kinerja IndoBERT meningkat sekitar 28%, sementara M-BERT meningkat sekitar 5%. Peningkatan yang kuat dari IndoBERT menunjukkan bahwa model yang dilatih secara eksklusif untuk bahasa Indonesia dapat meningkat secara signifikan ketika dioptimalkan untuk tugas tertentu, seperti menjawab pertanyaan-pertanyaan yang berkaitan dengan kesehatan mental.

B. Saran

Kontribusi lain dari penelitian ini adalah; i) model IndoBERT dan MBERT yang dioptimalkan untuk tugas QA domain kesehatan mental

dalam bahasa Indonesia, dan ii) terjemahan dataset Amod/mental_health_counseling_conversations ke dalam bahasa Indonesia. Sebagai tambahan, penelitian ini tidak hanya menunjukkan keunggulan IndoBERT dalam tugas *question answer* dalam bahasa Indonesia, tetapi juga menekankan pentingnya berinvestasi dalam pengembangan dan optimalisasi model khusus bahasa untuk meningkatkan aksesibilitas dan kualitas layanan digital. Temuan ini membuka jalan bagi penelitian di masa depan untuk mengadaptasi model NLP ke dalam konteks budaya dan domain aplikasi yang berbeda, seperti contoh model NusaBERT atau XLM-R dengan task yang lebih variatif.



DAFTAR PUSTAKA

- Alzubi, J. A., Jain, R., Singh, A., Parwekar, P., & Gupta, M. (2023). COBERT: COVID-19 Question Answering System Using BERT. *Arabian Journal for Science and Engineering*, 48(8), 11003–11013. <https://doi.org/10.1007/s13369-021-05810-5>
- Antoun, W., Baly, F., & Hajji, H. (2020). *AraBERT: Transformer-based Model for Arabic Language Understanding*. <http://arxiv.org/abs/2003.00104>
- Ayuningtyas, D., Misnaniarti, M., & Rayhani, M. (2018). Analisis Situasi Kesehatan Mental Pada Masyarakat Di Indonesia Dan Strategi Penanggulangannya. *Jurnal Ilmu Kesehatan Masyarakat*, 9(1), 1–10. <https://doi.org/10.26553/jikm.2018.9.1.1-10>
- Baharuddin, F., & Naufal, M. F. (2023). Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy. *Journal of Information Systems Engineering and Business Intelligence*, 9(2), 253–263. <https://doi.org/10.20473/jisebi.9.2.253-263>
- Bharadiya, J. P. (2023). A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. *European Journal of Technology*, 7(1), 58–66. www.ajpojournals.org
- Chen, A., Stanovsky, G., Singh, S., & Gardner, M. (2019). Evaluating question answering evaluation. *MRQA@EMNLP 2019 - Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, 119–124. <https://doi.org/10.18653/v1/d19-5817>
- Cresswell, J. w. (2012). *Educational Research Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (4th ed.). Boston: Phoenix Color Corp.
- Dartiko, F., Yusa, M., Erlansari, A., & Basha, S. A. (2024). Comparative Analysis of Transformer-Based Method In A Question Answering System for Campus Orientation Guides. *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, 8(1), 122–139.

<https://doi.org/10.29407/intensif.v8i1.21971>

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1*(Mlm), 4171–4186.
- Esfandiari, N., Kiani, K., & Rastgoo, R. (2023). *A Conditional Generative Chatbot using Transformer Model*.
- Götzl, C., Hiller, S., Rauschenberg, C., Schick, A., Fechtelpeter, J., Fischer Abaigar, U., Koppe, G., Durstewitz, D., Reininghaus, U., & Krumm, S. (2022). Artificial intelligence-informed mobile mental health apps for young people: a mixed-methods approach on users' and stakeholders' perspectives. *Child and Adolescent Psychiatry and Mental Health, 16*(1), 1–19. <https://doi.org/10.1186/s13034-022-00522-6>
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing, 187*, 27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>
- Jazuli, A., Widowati, & Kusumaningrum, R. (2023). Aspect-based sentiment analysis on student reviews using the Indo-Bert base model. *E3S Web of Conferences, 448*, 1–10. <https://doi.org/10.1051/e3sconf/202344802004>
- Kazemi, A., Mozafari, J., & Nematbakhsh, M. A. (2022). PersianQuAD: The Native Question Answering Dataset for the Persian Language. *IEEE Access, 10*, 26045–26057. <https://doi.org/10.1109/ACCESS.2022.3157289>
- Kelana, B., Syaani, R. A., Kristanto, F., & Winata, P. A. (2023). Kepercayaan Generasi Z pada Chatbot Layanan Kesehatan Mental Generation Z ' s Trust in The Chatbot of Mental Health Service. *MAJALAH SAINSTEKES, 10*(2), 136–141.
- Koto, F., Lau, J. H., & Baldwin, T. (2020). *Liputan6: A Large-scale Indonesian Dataset for Text Summarization. 1*. <http://arxiv.org/abs/2011.00679>
- Koto, F., Lau, J. H., & Baldwin, T. (2021). INDOBERTWEET: A Pretrained

- Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 10660–10668. <https://doi.org/10.18653/v1/2021.emnlp-main.833>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 757–770. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Kumar, A. (2015). ISCAIE 2014 - 2014 IEEE Symposium on Computer Applications and Industrial Electronics. *ISCAIE 2014 - 2014 IEEE Symposium on Computer Applications and Industrial Electronics*, 237.
- Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, 470(xxxx), 443–456. <https://doi.org/10.1016/j.neucom.2021.05.103>
- Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019). Chatbot for disease prediction and treatment recommendation using machine learning. *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, Icoei*, 851–856. <https://doi.org/10.1109/ICOEI.2019.8862707>
- Mohiuddin, K., Welke, P., Alam, M. A., Martin, M., Alam, M. M., Lehmann, J., & Vahdati, S. (2023). Retention Is All You Need. *International Conference on Information and Knowledge Management, Proceedings, Nips*, 4752–4758. <https://doi.org/10.1145/3583780.3615497>
- Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2022). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, 216(July), 714–721. <https://doi.org/10.1016/j.procs.2022.12.188>
- Nissa, N. K., & Yulianti, E. (2023). Multi-label text classification of Indonesian

- customer reviews using bidirectional encoder representations from transformers language model. *International Journal of Electrical and Computer Engineering*, 13(5), 5641–5652. <https://doi.org/10.11591/ijece.v13i5.pp5641-5652>
- Puspita, C. E., Pratiwi, O. N., & Sutoyo, E. (2021). Perbandingan Algoritma Klasifikasi Support Vector Machine Dan Naive Bayes Pada Imbalance Data. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 8(1), 11–18. <https://doi.org/10.33330/jurteksi.v8i1.1185>
- Rahali, A., & Akhloufi, M. A. (2023). End-to-End Transformer-Based Models in Textual-Based NLP. *AI (Switzerland)*, 4(1), 54–110. <https://doi.org/10.3390/ai4010004>
- Rakhra, M., Gopinadh, G., Addepalli, N. S., Singh, G., Aliraja, S., Reddy, V. S. G., & Reddy, M. N. (2021). E-Commerce Assistance with a Smart Chatbot using Artificial Intelligence. *Proceedings of 2021 2nd International Conference on Intelligent Engineering and Management, ICIEM 2021*, 144–148. <https://doi.org/10.1109/ICIEM51511.2021.9445316>
- Richardson, B., & Wicaksana, A. (2022). Comparison of Indobert-Lite and Roberta in Text Mining for Indonesian Language Question Answering Application. *International Journal of Innovative Computing, Information and Control*, 18(6), 1719–1734. <https://doi.org/10.24507/ijicic.18.06.1719>
- Rizqullah, M. R., Purwarianti, A., & Aji, A. F. (2023). QASiNa: Religious Domain Question Answering Using Sirah Nabawiyah. 2023 10th International Conference on Advanced Informatics: Concept, Theory and Application, ICAICTA 2023. <https://doi.org/10.1109/ICAICTA59291.2023.10390123>
- Sartika, D. (2020). Implementasi Algoritma K-Nearest Neighbour dalam Menganalisis Sentimen Terhadap Program Merdeka Belajar Kampus Merdeka (MBKM). *Jurnal Buana Informatika*, 69–76.
- Saunders, M., Lewis, P., & Thornhill, A. (2007). Research Methods for Business Students. In *Pearson*.

- https://www.researchgate.net/publication/330760964_Research_Methods_for_Business_Students_Chapter_4_Understanding_research_philosophy_and_approaches_to_theory_development
- Seo, M., Kembhavi, A., Farhadi, A., & Hajishirzi, H. (2017). Bi-Directional Attention Flow for Machine Comprehension. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 1–13.
- Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(07), 3897–3904. <https://ssrn.com/abstract=4458723>
- Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, 2(1), 24–28. <https://doi.org/10.1016/j.gltip.2021.01.004>
- Sharma, S., & Chaudhary, P. (2023). Machine learning and deep learning. *Quantum Computing and Artificial Intelligence: Training Machine and Deep Learning Algorithms on Quantum Computers*, 71–84. <https://doi.org/10.1515/9783110791402-004>
- Tamizharasi, B., Jenila Livingston, L. M., & Rajkumar, S. (2021). Building a medical chatbot using support vector machine learning algorithm. *Journal of Physics: Conference Series*, 1716(1). <https://doi.org/10.1088/1742-6596/1716/1/012059>
- Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet. *International Conference on Interdisciplinary Applications of Artificial Intelligence (ICIDAAI) 2021*. <http://arxiv.org/abs/2102.08036>
- Van Aken, B., Löser, A., Winter, B., & Gers, F. A. (2019). How does BERT answer questions? A layer-wise analysis of transformer representations. *International Conference on Information and Knowledge Management, Proceedings*, 1823–1832. <https://doi.org/10.1145/3357384.3358028>
- von Essen, H., & Hesslow, D. (2020). Building a Swedish Question-Answering

- Model. *Proceedings of the Probability and Meaning Conference (PaM 2020)*, PaM, 117–127. <https://www.aclweb.org/anthology/2020.pam-1.16>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. *EMNLP 2020 - Conference on Empirical Methods in Natural Language Processing, Proceedings of Systems Demonstrations*, 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- Wu, S., & Dredze, M. (2020). Are all languages created equal in multilingual BERT? *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, July, 120–130. <https://doi.org/10.18653/v1/2020.repl4nlp-1.16>
- Xiao, T., & Zhu, J. (2023). *Introduction to Transformers: an NLP Perspective*. <http://arxiv.org/abs/2311.17633>
- Yasmin, Z., Darmalaksana, W., & Ismail, E. (2022). Indeks Kesehatan Mental Menurut Rasulullah Saw: Studi Takhrij dan Syarah Hadis dengan Pendekatan Psikologis. *Gunung Djati Conference Series*, 8(2), 12–24.
- Yulia, H. (2023). Pemanfaatan Artificial Intelligence pada Pelaksanaan Cyber Counseling pada Pengguna Aplikasi Replika. *ULIL ALBAB : Jurnal Ilmiah Multidisiplin*, 2(11), 5321–5328. <https://journal-nusantara.com/index.php/JIM/article/view/2364>
- Yunanto, R., Wibowo, E. P., & Rianto, R. (2023). a Bert Model To Detect Provocative Hoax. *Journal of Engineering Science and Technology*, 18(5), 2281–2297.
- Zhang, B., Dong, N., & Rischmoller, L. (2020). Design thinking in action: A DPR case study to develop a sustainable digital solution for labor resource management. *IGLC 28 - 28th Annual Conference of the International Group for Lean Construction 2020*, 25–36. <https://doi.org/10.24928/2020/0137>
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). Bertscore:

Evaluating Text Generation With Bert. *8th International Conference on Learning Representations, ICLR 2020*, 1–43.

