

**PERBANDINGAN KINERJA 3 MODEL BAHASA INDOBERT  
UNTUK PERINGKASAN TEKS  
PADA DOKUMEN BAHASA INDONESIA**



Disusun Oleh:

**Herwinsyah**

**21206052001**

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

**2023**

## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Herwinsyah  
NIM : 21206052001  
Jenjang : Magister  
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian atau karya tulis saya pribadi. Kecuali, pada bagian – bagian yang dirujuk referensi sumbernya.

Yogyakarta, 6 Maret 2023

Saya yang menyatakan,



**Herwinsyah**

NIM. 21206052001

## PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan dibawah ini:

Nama : Herwinsyah  
NIM : 21206052001  
Jenjang : Magister Program  
Studi : Informatika

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

Yogyakarta, 6 Maret 2023

Saya yang menyatakan,

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



**Herwinsyah**  
NIM. 21206052001



## PENGESAHAN TUGAS AKHIR

Nomor : B-791/Un.02/DST/PP.00.9/03/2023

Tugas Akhir dengan judul : PERBANDINGAN KINERJA 3 MODEL BAHASA INDOBERT UNTUK PERINGKASAN TEKS PADA DOKUMEN BAHASA INDONESIA

yang dipersiapkan dan disusun oleh:

Nama : HERWINSYAH, S.Kom  
Nomor Induk Mahasiswa : 21206052001  
Telah diujikan pada : Kamis, 16 Maret 2023  
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: 6417ec277c525



Penguji I  
Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.  
SIGNED

Valid ID: 64180523db0d9



Penguji II  
Dr. Ir. Bambang Sugiantoro, S.Si., M.T.  
SIGNED

Valid ID: 64155624132fd



Yogyakarta, 16 Maret 2023  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi  
Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 641baae685a84



**KEMENTERIAN AGAMA REPUBLIK INDONESIA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
MAGISTER INFORMATIKA**

Jl. Marsda Adisucipto Yogyakarta. Telp (0274) 515856 Yogyakarta 55281

---

**SURAT PERSETUJUAN TUGAS AKHIR**

Hal : Persetujuan Tugas Akhir

Kepada:  
Yth. Dekan Fakultas Sains dan Teknologi  
Uin Sunan Kalijaga  
Yogyakarta Di Yogyakarta

*Assalamu'alaikum Wr.Wb.*

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul :  
PERBANDINGAN KINERJA 3 MODEL BAHASA INDOBERT UNTUK  
PERINGKASAN TEKS PADA DOKUMEN BAHASA INDONESIA yang di tulis oleh :

Nama : Herwinsyah  
NIM : 21206052001  
Jenjang : Magister  
Program Studi : Informatika

Sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan  
Teknologi UIN Sunan Kalijaga sebagai salah satu syarat untuk memperoleh gelar Magister  
Informatika.

Dengan ini saya mengharap agar tugas tersebut di atas agar dapat segera dimunaqosyahkan.  
Atas perhatiannya saya ucapkan terimakasih.

*Wassalamu'alaikum Wr. Wb.*

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

Yogyakarta, 6 Maret 2023  
Pembimbing,

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

## ABSTRAK

Hasil peringkasan dokumen yang baik yang dilakukan oleh model bahasa adalah hasil peringkasan yang memiliki nilai kemiripan yang besar terhadap hasil peringkasan yang dilakukan manusia, karena dalam peringkasan yang dilakukan manusia ada nilai semantik yang diperhatikan. Biasanya hasil peringkasan sebuah sistem melalui algoritma yang diterapkan, cenderung terpaku pada susunan kata saja, tanpa melihat sisi semantik dari sebuah kalimat. Penelitian ini terfokus kepada pengukuran seberapa tinggi nilai hasil evaluasi dari peringkasan dari 3 model bahasa yaitu IndoBERT.

Pada penelitian ini peneliti meringkas dari sumber yang sama menggunakan 3 model bahasa yang terfokus pada bahasa Indonesia. Hasil peringkasan dari 3 model bahasa tersebut kemudian peneliti bandingkan dengan peringkasan referensi yang merupakan hasil peringkasan manusia dari sumber paragraf tersebut. Tiga *tools* evaluasi digunakan dalam mengukur nilai hasil peringkasan dan hasil yang didapat adalah model bahasa IndoBERT\_3 atau IndoBERT versi Sarah Lintang (UGM Jogja) memiliki hasil peringkasan yang mendekati tingkat kemiripan dengan hasil peringkasan manusia (*gold summarize*).

**Kata Kunci:** Model Bahasa, IndoBERT, ROUGE, BLEU, BERTScore

## **ABSTRACT**

*A good document summary produced by a language model should have a high level of similarity to a summary generated by a human, as the latter considers the semantic value of the text. Typically, a summarization algorithm implemented by a system tends to focus solely on the arrangement of words, without taking into account the semantic aspect of a sentence. The objective of this study is to measure the quality of the summary results obtained from three language models, namely IndoBERT.*

*In this study, the researcher summarized the same source material using three language models that are focused on the Indonesian language. The summary results generated by the three language models were then compared with a reference summary, which is a summary generated by a human from the same paragraph source. Three evaluation tools were used to measure the quality of the summary results. The study found that the IndoBERT\_3 language model, or the IndoBERT version developed by Sarah Lintang (UGM Jogja), produced a summary result that closely approximates the level of similarity to the human-generated summary (gold summary).*

**Keywords:** *Language Model, IndoBERT, ROUGE, BLEU, BERTScore.*

## KATA PENGANTAR

*Bissmillahirrahmanirrahim,*

*Assalamualaikum warahmatullahi wabarakatuhu*

Puji syukur penulis panjatkan kehadirat Allah Subhanahu wata'ala yang telah melimpahkan segala rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan thesis dengan judul “Perbandingan Kinerja 3 Model Bahasa Indobert Untuk Peringkasan Teks Pada Dokumen Bahasa Indonesia” sebagai salah satu persyaratan mencapai derajat Magister Informatika.

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dorongan, semangat, dan ide-ide kreatif sehingga tahap demi tahap penyusunan thesis ini telah selesai. 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 motivasi dan pengarahan selama studi sehingga dapat menyelesaikan penyusunan tesis ini.
5. Bapak dan Ibu dosen Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang sudah membagi ilmu yang sangat bermanfaat.



6. Seluruh Staf Karyawan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta yang telah membantu sehingga penyusunan thesis ini berjalan lancar.
7. Orang tua, Istri dan anak-anakku yang senantiasa memberikan doa selama studi di Magister Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
8. Teman-teman magister informatika angkatan 2021 Universitas Islam Negeri Sunan Kalijaga atas kerjasama, saran, dan bantuannya.
9. Semua pihak yang tidak bisa penulis sebutkan satu persatu atas bantuannya dalam penyelesaian thesis ini.

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

*Wassalamualaikum warahmatullahi wabarakatuhu*

Yogyakarta, 6 Maret 2023



Herwinsyah  
21206052001

## PERSEMBAHAN

Alhamdulillahirobbil alamiin, Puji Syukur atas keridhoan Allah SWT sebagai dzat yang maha kuasa, yang telah memberikan kesehatan, rahmat dan hidayah serta karunia-Nya sehingga salah satu kewajibanku ini dapat diselesaikan. Tak lupa sholawat serta salam kepada junjungan Nabi Besar Muhammad SAW semoga syafa'at diberikan di akhir zaman. Halaman ini saya tujukan terhadap semua pihak yang telah membantu dan Mensupport Penyelesain Tesis ini, sebagai berikut :

- Alm. Bapak dan Ibu yang telah mengisi dunia saya dengan begitu banyak kebahagiaan sehingga seumur hidup tidak cukup untuk menikmati semuanya. Terima kasih atas semua cinta yang telah kalian berikan kepada saya.
- Istriku tercinta Dr. Rubini, S.Pd.I., M.Pd.I atas semua do'a, semangat, dukungan, perhatian, motivasi, serta cinta dan sayang yang tak terhingga selama ini.
- Anak-anaku tercinta Hani Zahrani, S.Pd, Zahi Raihansyah, Dzihni Syahira Ramadhani, Syahizzul Fawwaz, dan Dzakira Raisyah Maulidina atas pengertian serta cinta dan sayang dari kalian yang tak terhingga selama ini.

Sekian halaman persembahan ini saya buat sebagai apresiasi terhadap semua pihak yang telah memberi bantuan secara doa, moral, dan material dalam penunjang tesis saya ini. telah Terima kasih banyak dan mohon maaf apabila saya ada kesalahan dalam penulisan, mohon dimaafkan

## MOTTO

Belum mencapai impian bukan berarti Allah tidak memberi keadilan untuk hamba-Nya, karena

وَعَسَىٰ أَنْ تَكْرَهُوا شَيْئًا وَهُوَ خَيْرٌ لَّكُمْ وَعَسَىٰ أَنْ تُحِبُّوا شَيْئًا وَهُوَ  
شَرٌّ لَّكُمْ وَاللَّهُ يَعْلَمُ وَأَنْتُمْ لَا تَعْلَمُونَ

“Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui.”

QS Al-Baqarah: 216

*Rasulullah bersabda : Barangsiapa menempuh jalan untuk mendapatkan ilmu, Allah akan memudahkan baginya jalan menuju surga. HR. Musilm.*

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR ISI

PERNYATAAN KEASLIAN .....	II
PERNYATAAN BEBAS PLAGIASI .....	III
LEMBAR PENGESAHAN .....	IV
NOTA DINAS PEMBIMBING .....	V
ABSTRAK.....	VI
ABSTRACT .....	VII
KATA PENGANTAR.....	VIII
PERSEMBAHAN .....	X
MOTTO.....	XI
DAFTAR ISI .....	XII
DAFTAR GAMBAR.....	XV
DAFTAR TABEL .....	XVII
BAB I PENDAHULUAN .....	1
A. LATAR BELAKANG .....	1
B. RUMUSAN MASALAH.....	7
C. BATASAN MASALAH.....	7
D. TUJUAN PENELITIAN.....	8
E. MANFAAT PENELITIAN .....	9
F. KEASLIAN PENELITIAN .....	9
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	10
A. TINJAUAN PUSTAKA .....	10
B. LANDASAN TEORI.....	20
1. Python .....	20
2. Google Colaboratory.....	23
3. Library Python .....	25

4.	BERT .....	48
5.	Model Bahasa.....	49
6.	IndoBERT .....	50
7.	Indosum.....	53
8.	Perangkuman / Peringkasan (Summarize).....	55
9.	Dokumen.....	56
10.	Bahasa Indonesia.....	57
11.	ROUGE .....	58
12.	BLEU .....	60
13.	BERTScore.....	62
<b>BAB III METODE PENELITIAN .....</b>		<b>65</b>
A. KAJIAN LITERATUR.....		67
B. DATA PENELITIAN .....		69
C. ALAT PENELITIAN .....		70
1.	Perangkat Keras (Hardware).....	71
2.	Perangkat Lunak (Software).....	71
D. TAHAPAN PENELITIAN .....		72
1.	Cleaning Dataset .....	73
2.	Implementasi Model Bahasa IndoBERT .....	73
3.	Proses Analisa dan Evaluasi .....	74
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>76</b>
A. DATA .....		76
1.	Pengumpulan Dataset.....	76
2.	Pengolahan Dataset.....	78
3.	Implementasi Model Bahasa IndoBERT .....	79
B. UJI HASIL PERINGKASAN.....		81
1.	Uji Evaluasi Menggunakan ROUGE.....	81

2. Uji Evaluasi Menggunakan BLEU .....	83
3. Uji Evaluasi Menggunakan BERTScore .....	85
C. HASIL PENGUJIAN.....	87
1. Hasil Pengujian ROUGE .....	88
2. Hasil Pengujian BLEU.....	90
3. Hasil Pengujian BERTScore.....	92
BAB V PENUTUP .....	94
A. KESIMPULAN .....	94
B. SARAN.....	95
DAFTAR PUSTAKA.....	97
LAMPIRAN .....	104
1. Parsing Data json ke csv .....	104
2. Peringkasan dengan IndoBERT.....	105
3. Evaluasi Tools ROUGE.....	107
4. Evaluasi Tools BLEU .....	108
5. Evaluasi Tools BERTScore .....	109
DAFTAR RIWAYAT HIDUP .....	110

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
 YOGYAKARTA

## DAFTAR GAMBAR

Gambar 2.1 Hierarki tipe standar di Python 3 .....	22
Gambar 2.2 Google Colaboratory .....	24
Gambar 2.3 Library Python, Source : Google.com .....	27
Gambar 2.4 Library NumPy .....	28
Gambar 2.5 NumPy : Accelerated and Distributed Array .....	29
Gambar 2.6 Library Pandas .....	30
Gambar 2.7 Library Tensorflow .....	32
Gambar 2.8 Library SciPy .....	34
Gambar 2.9 Library Keras .....	36
Gambar 2.10 Library Matplotlib .....	37
Gambar 2.11 Library Scikit-learn .....	39
Gambar 2.12 Library PyTorch .....	41
Gambar 2.13 Library Scrapy .....	42
Gambar 2.14 Library Transformer .....	44
Gambar 2.15 Rumus Evaluasi ROUGE .....	59
Gambar 2.16 N-gram .....	61
Gambar 2.17 Rumus Evaluasi BLEU .....	61
Gambar 2.18 BERTScore : Evaluating Text Generation .....	63
Gambar 2.19 BERTScore Formula .....	63
Gambar 3.1 Metode Penelitian .....	67
Gambar 4.1 Dataset IndoSUM .....	77
Gambar 4.2 Convert data json ke csv .....	78
Gambar 4.3 Dataset IndoSUM .....	79
Gambar 4.4 Implementasi 3 Model Bahasa .....	80
Gambar 4.5 Implementasi ROUGE .....	82

Gambar 4.6 Hasil ROUGE IndoBERT_1.....	82
Gambar 4.7 Hasil ROUGE IndoBERT_2.....	83
Gambar 4.8 Hasil ROUGE IndoBERT_3.....	83
Gambar 4.9 Implementasi BLEU .....	84
Gambar 4.10 Hasil Implementasi BLEU.....	85
Gambar 4.11 Implementasi BERTScore .....	86
Gambar 4.12 Hasil BERTScore IndoBERT_1 .....	86
Gambar 4.13 Hasil BERTScore IndoBERT_2.....	87
Gambar 4.14 Hasil BERTScore IndoBERT_3.....	87
Gambar 4.15 Hasil Uji performansi dengan ROUGE .....	88





## DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	14
Tabel 3.1. Alat Penelitian .....	71
Tabel 3.2 Bahan Penelitian.....	72
Tabel 4.1 Dataset IndoSUM.....	77
Tabel 4.2 Analisis Hasil Uji ROUGE.....	89
Tabel 4.3 Hasil Uji BLEU.....	91
Tabel 4.4 Hasil Uji BERTScore .....	93



# **BAB I**

## **PENDAHULUAN**

### **A. LATAR BELAKANG**

Teknologi kecerdasan buatan atau *artificial intelligent* telah mengalami kemajuan dan perkembangan yang sangat pesat dalam beberapa tahun terakhir. Pada awalnya, kecerdasan buatan hanya digunakan sebatas untuk melakukan tugas-tugas yang simple dan sederhana seperti tugas dalam menyortir data, mengurutkan dan tugas-tugas yang sederhana lainnya seperti menjalankan perintah yang telah ditentukan. Namun, seiring dengan kemajuan teknologi pemrosesan data dan pembelajaran mesin, kecerdasan buatan kini mampu melakukan tugas-tugas yang lebih kompleks seperti mengenali wajah, mengidentifikasi suara, membuat kalimat, menjawab pertanyaan dan bahkan melakukan pembelajaran terbuka. Teknologi (*Artificial Intelligence*) AI atau kecerdasan buatan dengan fitur, fungsi, dan tampilan yang baru semakin berdampak pada banyak aspek kehidupan manusia (Luger, George F., 1993).

Dengan menggunakan teknologi pembelajaran mesin, kecerdasan buatan kini mampu memperoleh kemampuan baru dengan memproses dan menganalisis data yang cukup besar (Dwivedi *et al.*, 2021). Hal ini membuat kecerdasan buatan semakin berguna dalam berbagai bidang, seperti perawatan kesehatan, transportasi, dan dalam bidang militer, bahkan dalam beberapa bidang lainnya yang sedang berkembang.

Teknologi kecerdasan buatan telah mengalami perkembangan yang pesat dan luar biasa dalam beberapa tahun terakhir, teknologi tersebut diharapkan akan terus berkembang di masa yang akan datang. Dengan

kecerdasan buatan diharapkan pekerjaan manusia yang lebih kompleks dapat dilakukan oleh teknologi ini sehingga meminimalisir human *error* yang dapat terjadi dan merugikan manusia itu sendiri.

*Artificial Intelligence* (AI) dan *Natural Language Processing* (NLP) merupakan dua bidang yang sangat erat terkait. NLP adalah bidang yang mempelajari bagaimana mesin dapat memahami, mengolah, dan menggunakan bahasa alami manusia seperti bahasa Indonesia, bahasa Inggris atau bahasa lainnya. Sedangkan AI adalah bidang yang mempelajari bagaimana mesin dapat melakukan tugas-tugas yang biasanya dilakukan oleh manusia, seperti pembelajaran, pengambilan keputusan, dan pemecahan masalah. NLP otomatis bertujuan untuk melakukan pekerjaan ini secara efektif dan akurat, seperti yang dilakukan manusia (Chowdhary, 2020).

Dalam proses komunikasi, bahasa alami merupakan cara utama manusia untuk berkomunikasi dan menyampaikan informasi, bahasa alami adalah bahasa sehari-hari yang digunakan dalam komunikasi. Salah satu keberhasilan kecerdasan buatan (AI) adalah mengkolaborasikan antara NLP dan AI, tanpa kemampuan untuk memahami bahasa alami, AI tidak akan dapat mengolah informasi yang disampaikan dalam bentuk teks atau suara.

NLP juga membantu AI untuk menangani masalah yang terkait dengan ambiguitas dan konotasi dalam bahasa alami. NLP dan AI merupakan dua bidang yang erat terkait dan saling melengkapi. Singkatnya NLP membantu AI untuk memahami bahasa alami, sementara AI memanfaatkan kemampuan NLP untuk mengolah dan menggunakan informasi dalam bahasa alami untuk menyelesaikan tugas-tugas yang biasanya dilakukan oleh manusia.

Meskipun kemajuan teknologi AI dalam pemrosesan teks telah

meningkat secara signifikan dalam beberapa tahun terakhir, masih ada beberapa batasan yang harus diperhatikan. Salah satu batasan utama adalah kemampuan AI untuk memahami makna teks secara mendalam (Koleck *et al.*, 2019). AI masih terbatas dalam kemampuan untuk menangkap nuansa, konotasi, dan implikasi dari teks, yang seringkali merupakan bagian penting dari pemahaman manusia terhadap teks. Selain itu, AI juga masih terbatas dalam kemampuan untuk menangani teks yang tidak struktur dan tidak teratur, seperti teks yang terdapat dalam forum online atau teks yang berupa obrolan.

Batasan lain dari AI dalam pemrosesan teks adalah kemampuan untuk memahami bahasa yang tidak standar, seperti bahasa gaul atau bahasa yang digunakan oleh komunitas tertentu. AI juga masih terbatas dalam kemampuan untuk menangani teks yang mengandung ejaan yang salah atau kesalahan tata bahasa, yang sering terjadi dalam teks yang ditulis secara cepat atau dengan kurang teliti. AI juga masih terbatas dalam kemampuan untuk menangani teks yang mengandung metafora atau perbandingan, yang seringkali merupakan bagian penting dari bahasa manusia.

Meskipun demikian, kemajuan teknologi AI dalam pemrosesan teks telah memberikan kemajuan yang signifikan dalam banyak bidang, dan dalam beberapa tahun terakhir, terdapat perkembangan yang signifikan dalam bidang NLP, salah satunya adalah munculnya model BERT (*Bidirectional Encoder Representations from Transformers*). BERT (Koroteev, 2021) merupakan salah satu model terbaru dalam NLP yang mampu menangani masalah pemahaman teks secara lebih mendalam. Model ini menggunakan teknik pemrosesan "*bidirectional*" yang memungkinkan model untuk memahami konteks teks secara lebih baik, dengan mempertimbangkan kata-kata yang ada sebelum dan sesudah kata

yang sedang diproses.

Proses pembelajaran sebuah model bahasa biasanya menggunakan kumpulan dataset yang banyak untuk dapat memahami makna kata. Begitu juga dalam model bahasa BERT dimana dataset di *pre-trained* atau dilakukan proses pembelajaran. Dataset tersebut kemudian akan diubah menjadi sekumpulan token (kata atau frasa yang dianggap penting oleh model) yang akan diolah oleh model. Selanjutnya, model akan mengolah token tersebut dengan menggunakan teknik pemrosesan *bidirectional* sehingga dapat memahami konteks teks secara lebih baik.

Setelah melalui proses pemrosesan, model BERT akan memberikan output berupa sekumpulan vektor (*array numerik*) yang mewakili makna dari setiap token dalam teks. Vektor tersebut kemudian dapat digunakan untuk berbagai keperluan aplikasi NLP, seperti pengklasifikasian teks atau pembuatan jawaban otomatis untuk pertanyaan dan beberapa keperluan lainnya yang berkaitan dengan penggunaan kata pada bahasa alami.

Salah satu model bahasa yang merupakan model bahasa menggunakan model BERT khusus untuk pemrosesan bahasa alami berbahasa Indonesia adalah IndoBERT. Munculnya IndoBERT merupakan hasil pengembangan yang lebih luas dalam pengembangan model bahasa yang memfokuskan pada bahasa Indonesia. Sebelumnya, model bahasa yang tersedia cenderung berfokus pada bahasa Inggris, sehingga model bahasa yang dikembangkan untuk bahasa seperti Indonesia sangat dibutuhkan.

IndoBERT memiliki beberapa kelebihan dibandingkan model bahasa lain. Pertama, IndoBERT telah dikembangkan dengan mempertimbangkan konteks dan kultur bahasa Indonesia, sehingga dapat memahami bahasa Indonesia dengan baik. Kedua, IndoBERT memiliki akurasi yang lebih tinggi dalam memahami bahasa Indonesia, karena model ini dibuat dengan

data-data korpus berbahasa Indonesia.

IndoBERT adalah salah satu model NLP (*Natural Language Processing*) yang dikembangkan untuk memahami bahasa Indonesia secara lebih baik. Model ini merupakan versi Indonesia dari model BERT yang telah terbukti efektif dalam memahami bahasa alami.

Model bahasa berbasis *deep learning* yang telah dilatih pada *unannotated* teks berjumlah besar telah dikembangkan untuk *efisiensi transfer learning* pada pemrosesan bahasa alami. Model berbasis Transformer seperti BERT sangat populer akhir-akhir ini karena performa yang sangat baik, namun hanya sedikit yang berfokus pada bahasa dengan *resource* yang terbatas seperti bahasa Indonesia. Model bahasa IndoBERT merupakan model bahasa yang di *pretrained* dengan berbagai macam kata-kata baik yang formal maupun yang tidak formal, sehingga model ini secara konstan mengungguli *multilingual* BERT model pada aplikasi pemrosesan bahasa alami berbasis bahasa Indonesia seperti sentimen analisis dan peringkasan teks (Sarah Lintang, 2020).

Model bahasa IndoBERT dikembangkan untuk menangani teks dalam bahasa Indonesia dan dapat digunakan untuk berbagai tugas pemrosesan bahasa termasuk peringkasan teks. Dengan menggunakan IndoBERT, kita dapat dengan mudah membuat ringkasan teks yang baik, akurat dan efisien dalam bahasa Indonesia. Selain itu, IndoBERT juga dapat membantu dalam mengidentifikasi tren dan pola dalam data teks, sehingga dapat membantu dalam proses pengambilan keputusan dan merupakan alat yang sangat berguna bagi para peneliti, analis, dan pengembang yang ingin menangani pengolahan teks dalam bahasa Indonesia.

*Summarizing* atau ringkasan teks adalah proses mengurangi teks asli menjadi versi yang lebih ringkas, yang mencakup informasi yang paling

penting dan esensial. *Summarizing* dapat membantu dalam menyimpan waktu dan meningkatkan efisiensi dalam memproses informasi.

IndoBERT merupakan model bahasa alami yang sangat kuat untuk bahasa Indonesia, masih belum banyak aplikasi yang menggunakannya secara langsung. Hal ini disebabkan oleh beberapa faktor seperti sumber daya dan waktu yang dibutuhkan, penyesuaian dengan kebutuhan aplikasi tertentu, model bahasa ini masih tergolong baru dalam penelitian, dan masih banyak peneliti lainnya yang ikut mengembangkan dan menyempurnakan. Namun, dengan adanya peningkatan kesadaran dan pengembangan yang terus dilakukan, diharapkan dalam waktu dekat akan muncul aplikasi yang menggunakan IndoBERT secara luas dan bermanfaat bagi masyarakat.

Ada beberapa alasan mengapa melakukan penelitian tentang IndoBERT sangat penting. Pertama, bahasa Indonesia adalah bahasa yang kompleks dan bervariasi, dengan banyak kosakata dan tata bahasa yang berbeda tergantung pada wilayah dan budaya. Oleh karena itu, model bahasa alami seperti IndoBERT sangat diperlukan untuk memahami dan menghasilkan teks dalam bahasa Indonesia.

Kedua, keberadaan IndoBERT dapat membantu meningkatkan kualitas hasil karya dalam bidang seperti pemrosesan bahasa alami, analisis sentimen, klasifikasi dokumen, dan lain sebagainya. Dalam era digital yang semakin maju, banyak organisasi dan perusahaan yang membutuhkan teknologi bahasa alami yang handal dan efektif untuk memproses data teks dalam bahasa Indonesia.

Ketiga, penelitian tentang IndoBERT dapat membantu dalam pengembangan teknologi bahasa alami yang lebih canggih dan maju di masa depan. Dengan melakukan penelitian tentang model ini, para peneliti dapat menemukan cara baru untuk meningkatkan kinerja dan efisiensi model,

sehingga dapat memberikan manfaat yang lebih besar bagi pengguna di seluruh dunia.

Penelitian tentang IndoBERT sangat penting untuk meningkatkan pemahaman dan penggunaan bahasa Indonesia dalam era digital yang semakin maju. Dengan menggunakan model bahasa alami yang handal seperti IndoBERT, kita dapat memproses data teks dalam bahasa Indonesia dengan lebih mudah dan efektif, kita dapat dengan mudah membuat ringkasan teks yang akurat dan efisien dalam bahasa Indonesia. Model ini juga dapat membantu dalam mengidentifikasi data teks, sehingga dapat membantu dalam proses pengambilan keputusan. Dengan demikian, IndoBERT merupakan alat yang sangat berguna bagi para peneliti, analis, dan pengembang yang ingin menangani teks dalam bahasa Indonesia, sehingga memberikan manfaat yang lebih besar bagi masyarakat Indonesia dan dunia pada umumnya

## **B. RUMUSAN MASALAH**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana perbandingan kinerja (*comparative study*) dari 3 model bahasa IndoBERT terhadap hasil peringkasan teks menggunakan dokumen teks berbahasa Indonesia dengan pendekatan *experimental*.

## **C. BATASAN MASALAH**

Batasan masalah merupakan ruang lingkup dari permasalahan yang akan dibatasi dalam penelitian Perbandingan Kinerja 3 Model Bahasa IndoBERT Untuk Peringkasan Teks Pada Dokumen Bahasa Indonesia. Adapun Batasan masalah yang sudah ditetapkan dalam penelitian ini, antara



lain:

1. Model Bahasa yang digunakan adalah IndoBERT yang merupakan salah satu model NLP (*Natural Language Processing*) versi Indonesia dari model bahasa BERT (*Bidirectional Encoder Representations from Transformers*) yang telah terbukti efektif dalam memahami bahasa alami.
2. Parameter yang diujikan dengan menggunakan hasil peringkasan Indosum, Indosum merupakan hasil proses peringkasan berdasarkan hasil peringkasan manusia (*gold summarize*) yang dinilai *advance* dalam merangkum teks berbahasa Indonesia.
3. Penelitian ini termasuk dalam penelitian kuantitatif dan teknik pengumpulan datanya menggunakan pendekatan *experimental*.
4. Penelitian ini hanya menggunakan bahasa pemrograman python versi 3.8 sampai dengan 3.10 dengan Aplikasi *google colab*
5. Evaluasi pada penelitian ini menggunakan ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*), BERT-Score, dan BLEU (*BiLingual Evaluation Understudy*),

#### **D. TUJUAN PENELITIAN**

Berdasarkan rumusan masalah dan Batasan masalah di atas, maka tujuan dari penelitian ini adalah melakukan studi perbandingan (*comparative study*) Terhadap Kinerja 3 Model Bahasa IndoBERT Untuk Peringkasan Teks Pada Dokumen Bahasa Indonesia dengan metode kuantitatif dan teknik pengumpulan datanya menggunakan pendekatan *experimental*. Proses evaluasi menggunakan tools evaluasi ROUGE, BERT-Score dan BLEU, dengan evaluasi tersebut akan didapatkan hasil

terkait kinerja 3 model bahasa indobert untuk peringkasan teks pada dokumen bahasa Indonesia.

#### **E. MANFAAT PENELITIAN**

Dalam hasil penelitian yang telah dilakukan, diharapkan akan memberikan manfaat sebagai berikut:

1. Memberikan informasi terhadap hasil perbandingan kinerja dari 3 model bahasa IndoBERT terhadap hasil peringkasan teks menggunakan dokumen teks berbahasa Indonesia.
2. Mendapatkan informasi dari ketiga model bahasa, mana model bahasa yang lebih baik untuk digunakan dalam pemrosesan peringkasan teks bahasa Indonesia.

#### **F. KEASLIAN PENELITIAN**

Penelitian yang berkaitan dengan Perbandingan Kinerja 3 Model Bahasa IndoBERT Untuk Peringkasan Teks Pada Dokumen Bahasa Indonesia yang dilakukan dengan menggunakan metode Kuantitatif dan pendekatan *experimental*, sejauh pengamatan yang telah dilakukan dengan mempertimbangkan tinjauan pustaka, penelitian tersebut belum pernah dilakukan pada penelitian sebelumnya.

## **BAB V**

### **PENUTUP**

#### **A. KESIMPULAN**

Hasil Pengujian yang didapatkan dengan melakukan pengujian menggunakan 3 tools yaitu ROUGE, BLEU dan BERTScore didapatkan data perbandingan dari masing-masing model bahasa. Model bahasa IndoBERT\_3 memiliki nilai rata-rata terbesar pada semua pengujian, Pada pengujian dengan tools evaluasi ROUGE dan BERTScore baik *f1-score*, *recall* dan *Precision*. Pada pengujian *tools* evaluasi BLEU IndoBERT\_3 mendapatkan nilai terbesar, baik pada hasil nilai bleu dan juga nilai rata-rata *precisions*.

Pada pengujian ROUGE pada nilai Rouge-1 untuk nilai model bahasa IndoBERT\_3 mendapatkan nilai *f1-score* sebesar 0.552 atau 55.2%, *precisian* sebesar 0.579 atau 57.9%, dan juga *recall* sebesar 0.552 atau 55.2%. Begitu juga dengan pengujian menggunakan BLEU didapatkan hasil terbesar untuk hasil rata-rata nilai adalah model bahasa IndoBERT\_3 dengan nilai bleu nya sebesar 0.4646 atau sebesar 46.46% dan nilai rata-rata *precisions* sebesar 0.4781 atau sebesar 47.81%. Pengujian dengan *tools* BERTScore IndoBERT\_3 juga mandapatkan nilai yang paling besar dengan nilai *f1-score* sebesar 0.8940 atau 89.40%, *precisian* sebesar 0.8976 atau 89.76%, dan juga *recall* sebesar 0.8912 atau 89.12%.

Berdasarkan penelitian ini dapat diambil kesimpulan bahwa model bahasa IndoBERT\_3 atau IndoBERT versi sarah Lintang (UGM Jogja) memiliki hasil peringkasan yang mendekati tingkat kemiripan dengan hasil peringkasan masnusia (*gold summarize*).

Penilaian ROUGE dan BLEU menghasilkan nilai yang lebih kecil

daripada penilaian BERTScore karena keduanya hanya memperhitungkan kesamaan kata atau frasa (*sequential*) yang tepat dalam teks referensi dan teks hasil, sementara itu, BERTScore menggunakan model bahasa alami BERT yang lebih canggih untuk memahami konteks dan makna dari kalimat secara lebih komprehensif (Token). Sehingga, BERTScore dapat menangkap kemiripan dalam kalimat dengan cara yang lebih akurat dan lebih sesuai dengan persepsi manusia tentang kesamaan makna. Walaupun nilai BERTScore lebih tinggi, hal ini tidak berarti bahwa ROUGE dan BLEU tidak berguna, tetapi ketiganya tools evaluasi tersebut dapat memberikan sudut pandang yang berbeda dalam mengevaluasi kesamaan teks.

Penelitian ini memiliki kontribusi penting kelanjutan memperbaiki kualitas ringkasan teks dalam bahasa Indonesia. Hasilnya menunjukkan bahwa ketiga model tersebut dapat menghasilkan ringkasan teks dengan baik. Hasil penelitian ini dapat digunakan sebagai dasar untuk mengembangkan model bahasa IndoBERT yang lebih baik dan lebih akurat dalam menghasilkan ringkasan teks dalam bahasa Indonesia.

## **B. SARAN**

Saran untuk penelitian lebih lanjut pada model bahasa IndoBERT adalah sebagai berikut:

- a. Meningkatkan akurasi model: Meskipun model bahasa IndoBERT telah menghasilkan hasil yang sangat baik, masih ada ruang untuk meningkatkan akurasinya. Penelitian lebih lanjut dapat dilakukan untuk mengevaluasi berbagai teknik *fine-tuning* yang berbeda atau menggunakan dataset yang lebih besar dan lebih spesifik untuk melatih model.
- b. Menerapkan *transfer learning*: Penelitian dapat dilakukan untuk

- menerapkan teknik *transfer learning* pada model bahasa IndoBERT, yaitu memanfaatkan pengetahuan yang telah dipelajari oleh model pada tugas-tugas lain untuk meningkatkan kinerjanya dalam tugas-tugas baru seperti ringkasan teks.
- c. Meningkatkan kecepatan dan efisiensi: Penelitian dapat dilakukan untuk meningkatkan kecepatan dan efisiensi model bahasa IndoBERT. Hal ini dapat dilakukan dengan mengurangi ukuran model atau menggunakan teknik-teknik kompresi model yang efektif.
  - d. Penelitian dapat dilakukan untuk mengembangkan aplikasi yang lebih banyak dan beragam yang memanfaatkan model bahasa IndoBERT, seperti *chatbot*, analisis sentimen, dan penerjemahan otomatis.
  - e. Untuk penelitian selanjutnya bisa membandingkan antara model bahasa IndoBERT dengan hasil peringkasan berbasis BERT lainnya misalnya M-BERT atau model bahasa lainnya yang berbasis transformer, atau bisa juga membandingkan *tool* pengujian lainnya seperti yang belum digunakan lainnya.

## DAFTAR PUSTAKA

- A. Muhandian (2018) *Belajar Pemrograman Python: Pengenalan Dasar Python dan Persiapan Awal*, Online.
- Alagoz, N. K., Score, Bert and Score, Bart (2022) ‘Extractive Summarization of Scientific Paper in Turkish With Deep Learning’, pp. 0–22.
- Alan Tusa Bagus W, D. H. F. (2021) ‘KLASIFIKASI EMOSI PADA TEKS DENGAN MENGGUNAKAN METODE DEEP LEARNING’, *Jurnal Ilmiah Indonesia*, 6(1), pp. 546–553.
- Alfhi Saputra, M. (2021) ‘Peringkasan Teks Otomatis Bahasa Indonesia secara Abstraktif Menggunakan Metode Long Short-Term Memory’, *e-Proceeding of Engineering : Vol.8, No.2 April 2021* |, 8(2), pp. 3474–3488.
- Awalina, A. *et al.* (2022) ‘Perbandingan Pretrained Model Transformer Pada Deteksi Ulasan Palsu Comparison of Pretrained Transformer Models on Spam Review Detection’, 9(3), pp. 597–604. doi: 10.25126/jtiik.202295696.
- Beltagy, I., Lo, K. and Cohan, A. (2019) ‘SCIBERT: A pretrained language model for scientific text’, *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 3615–3620. doi: 10.18653/v1/d19-1371.
- Bisong, E. (2019) ‘NumPy’, in *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Berkeley, CA: Apress, pp. 91–113. doi: 10.1007/978-1-4842-4470-8\_10.
- Cao, Shengjia *et al.* (2021) ‘Research on Python Data Visualization Technology’, *Journal of Physics: Conference Series*, 1757(1), p. 12122. doi: 10.1088/1742-6596/1757/1/012122.
- Chowdhary, K. R. (2020) ‘Natural Language Processing’, in *Fundamentals*

of *Artificial Intelligence*. New Delhi: Springer India, pp. 603–649. doi: 10.1007/978-81-322-3972-7\_19.

Crisis, C. *et al.* (2020) ‘BERTscore’, *Iclr*, 9(2009), pp. 209–224. Available at: <https://arxiv.org/pdf/1904.09675.pdf>.

Devlin, J. *et al.* (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), pp. 4171–4186.

Dhalla, H. K. (2020) ‘A Performance Analysis of Native JSON Parsers in Java, Python, MS.NET Core, JavaScript, and PHP’, in *2020 16th International Conference on Network and Service Management (CNSM)*, pp. 1–5. doi: 10.23919/CNSM50824.2020.9269101.

Dhalla, H. K. (2021) ‘Parsing Performance of Native JSON Libraries in Java, MS.Net Core, and Python: A Comparative Study’, *J. Comput. Sci. Coll.*, 36(3), p. 179.

Dwivedi, Y. K. *et al.* (2021) ‘Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy’, *International Journal of Information Management*, 57, p. 101994. doi: <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>.

Gao, Z. *et al.* (2019) ‘Target-dependent sentiment classification with BERT’, *IEEE Access*, 7, pp. 154290–154299. doi: 10.1109/ACCESS.2019.2946594.

Gholizadeh, S. (2022) ‘Top Popular Python Libraries in Research’, pp. 1–8.

Google (2022) *Google Colab FAQ*. (n.d.). Google Colaboratory, Online. Available at: <https://research.google.com/Colaboratory/intl/id/faq.html>.

Habibi, M., Priadana, A. and Rifqi Ma’arif, M. (2021) ‘Sentiment Analysis and Topic Modeling of Indonesian Public Conversation about COVID-19 Epidemics on Twitter’, *IJID (International Journal on Informatics for Development)*, 10(1), pp. 23–30. doi:

10.14421/ijid.2021.2400.

- Hadiyan, K. P., Arif Bijaksana, M. and Romadhony, A. (2021) 'Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT', *e-Proceeding of Engineering*, Vol.8, No.(2), pp. 3028–3038.
- Hao, J. and Ho, T. K. (2019) 'Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language', *Journal of Educational and Behavioral Statistics*, 44(3), pp. 348–361. doi: 10.3102/1076998619832248.
- Harris, C. R. *et al.* (2020) 'Array programming with NumPy', *Nature*, 585(7825), pp. 357–362. doi: 10.1038/s41586-020-2649-2.
- Imambi, S., Prakash, K. B. and Kanagachidambaresan, G. R. (2021) 'PyTorch BT - Programming with TensorFlow: Solution for Edge Computing Applications', in Prakash, K. B. and Kanagachidambaresan, G. R. (eds). Cham: Springer International Publishing, pp. 87–104. doi: 10.1007/978-3-030-57077-4\_10.
- Ismi, D. P. and Ardianto, F. (2020) 'Peringkasan Ekstraktif Teks Bahasa Indonesia dengan Pendekatan Unsupervised Menggunakan Metode Clustering', *Cybernetics*, 3(02), p. 90. doi: 10.29406/cbn.v3i02.2290.
- 'J. Xu, G. Durrett, Neural extractive text summarization with syntactic compression. arXiv preprint arXiv:1902.00863 (2019)' (no date). Available at: <http://arxiv.org/abs/1902.00863>.
- Jain, R. *et al.* (2022) 'WIDAR - Weighted Input Document Augmented ROUGE BT - Advances in Information Retrieval', in Hagen, M. *et al.* (eds). Cham: Springer International Publishing, pp. 304–321.
- Koleck, T. A. *et al.* (2019) 'Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review', *Journal of the American Medical Informatics Association*, 26(4), pp. 364–379. doi: 10.1093/jamia/ocy173.
- Koroteev, M. V (2021) '{BERT:} {A} Review of Applications in Natural Language Processing and Understanding', *CoRR*, abs/2103.1. Available at: <https://arxiv.org/abs/2103.11943>.



- Koto, F. *et al.* (2020) 'IndoLEM and IndoBERT: {A} Benchmark Dataset and Pre-trained Language Model for Indonesian {NLP}', *CoRR*, abs/2011.0. Available at: <https://arxiv.org/abs/2011.00677>.
- Kurniawan, K. and Louvan, S. (2018) 'Indosum: A New Benchmark Dataset for Indonesian Text Summarization', in *2018 International Conference on Asian Language Processing (IALP)*, pp. 215–220. doi: 10.1109/IALP.2018.8629109.
- Lavie, A., Sagae, K. and Jayaraman, S. (2004) 'The significance of recall in automatic metrics for MT evaluation', in *Machine Translation: From Real Users to Research: 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004, Washington, DC, USA, September 28-October 2, 2004. Proceedings 6*, pp. 134–143.
- Lin, C.-Y. (2004) *ROUGE: A Package for Automatic Evaluation of Summaries*. Available at: <https://aclanthology.org/W04-1013.pdf>.
- Liu, Y. and Lapata, M. (2019) 'Text Summarization with Pretrained Encoders', *CoRR*, abs/1908.0. Available at: <http://arxiv.org/abs/1908.08345>.
- Luger, George F., D. W. A. S. (1993) *Artificial Intelligence Structures And Strategies For Complexmproblem Soving 2nd Edition*. California: The Benjamin/Cumming Publishing Company Inc.
- Luthfi, E. T., Yusoh, Z. I. M. and Aboobaidier, B. M. (2021) 'Enhancing the Takhrij Al-Hadith based on Contextual Similarity using BERT Embeddings', *International Journal of Advanced Computer Science and Applications*, 12(11), pp. 286–293. doi: 10.14569/IJACSA.2021.0121133.
- Mathur, N., Baldwin, T. and Cohn, T. (2020) 'Tangled up in {BLEU:} Reevaluating the Evaluation of Automatic Machine Translation Evaluation Metrics', *CoRR*, abs/2006.0. Available at: <https://arxiv.org/abs/2006.06264>.
- Meng, K. *et al.* (2022) 'Locating and Editing Factual Associations in GPT', (NeurIPS). Available at: <http://arxiv.org/abs/2202.05262>.
- Moolayil, J. (2019) 'An Introduction to Deep Learning and Keras BT - Learn Keras for Deep Neural Networks: A Fast-Track Approach to

- Modern Deep Learning with Python', in Moolayil, J. (ed.). Berkeley, CA: Apress, pp. 1–16. doi: 10.1007/978-1-4842-4240-7\_1.
- Nagpal, A. and Gabrani, G. (2019) 'Python for Data Analytics, Scientific and Technical Applications', in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pp. 140–145. doi: 10.1109/AICAI.2019.8701341.
- Nath, S. *et al.* (2022) 'New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology', *British Journal of Ophthalmology*, 106(7), pp. 889–892. doi: 10.1136/bjophthalmol-2022-321141.
- Nilamsari, N. (2014) 'Memahami Studi Dokumen Dalam Penelitian Kualitatif', *Wacana*, 13(2), pp. 177–181.
- Paszke, A. *et al.* (2019) 'PyTorch: An Imperative Style, High-Performance Deep Learning Library', in Wallach, H. *et al.* (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>.
- Perkovic, L. (2015) *Introduction to Computing Using Python: An Application Development Focus*. USA: John Wiley & Sons, Inc.
- Rahmawati, A., Alamsyah, A. and Romadhony, A. (2022) 'Hoax News Detection Analysis using IndoBERT Deep Learning Methodology', in *2022 10th International Conference on Information and Communication Technology (ICoICT)*, pp. 368–373. doi: 10.1109/ICoICT55009.2022.9914902.
- Reiss, F., Cutler, B. and Eichenberger, Z. (2021) 'Natural Language Processing with Pandas DataFrames', *Proceedings of the 20th Python in Science Conference (Scipy)*, pp. 49–57. doi: 10.25080/majora-1b6fd038-006.
- Richardson, B. and 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), pp. 1719–1734. doi: 10.24507/ijicic.18.06.1719.

- Rossum, G. (1995) *Python Reference Manual*. NLD: CWI (Centre for Mathematics and Computer Science).
- Saadah, S. *et al.* (2022) ‘Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia’, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(4), pp. 648–655. doi: 10.29207/resti.v6i4.4215.
- Sarah Lintang (2020) *IndoBERT*. Universitas Gajah Mada.
- Singh, P. and Manure, A. (2020) ‘Introduction to TensorFlow 2.0 BT - Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python’, in Singh, P. and Manure, A. (eds). Berkeley, CA: Apress, pp. 1–24. doi: 10.1007/978-1-4842-5558-2\_1.
- Smirnov, O. (2021) ‘TensorFlow ManOpt: a library for optimization on Riemannian manifolds’, *CoRR*, abs/2105.1. Available at: <https://arxiv.org/abs/2105.13921>.
- Stančin, I. and Jović, A. (2019) ‘An overview and comparison of free Python libraries for data mining and big data analysis’, in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 977–982. doi: 10.23919/MIPRO.2019.8757088.
- Sugiyono (2012) *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Tran, M.-K. *et al.* (2022) ‘Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery’, *International Journal of Energy Research*, 46(2), pp. 786–794. doi: <https://doi.org/10.1002/er.7202>.
- Vaswani, A. *et al.* (2017) ‘Attention is all you need’, *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), pp. 5999–6009.
- Virtanen, P. *et al.* (2020) ‘SciPy 1.0: fundamental algorithms for scientific computing in Python’, *Nature Methods*, 17(3), pp. 261–272. doi: 10.1038/s41592-019-0686-2.
- Wilie, B. *et al.* (2020) ‘IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding’. Available at:

<http://arxiv.org/abs/2009.05387>.

Wolf, T. *et al.* (2020) ‘Transformers : State-of-the-Art Natural Language Processing’, pp. 38–45.

Yang, F. *et al.* (2021) ‘Transformers-sklearn: a toolkit for medical language understanding with transformer-based models’, *BMC Medical Informatics and Decision Making*, 21(2), p. 90. doi: 10.1186/s12911-021-01459-0.

