

**IDENTIFIKASI POTENSI *LOGICAL FALLACY* PADA KOMENTAR
SOSIAL MEDIA MENGGUNAKAN INDOBERT**

SKRIPSI

FIRDHA ISLY RAMADHANI

181402029



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

IDENTIFIKASI POTENSI *LOGICAL FALLACY* PADA KOMENTAR
SOSIAL MEDIA MENGGUNAKAN INDOBERT

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi

FIRDHA ISLY RAMADHANI

181402029



PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024

PERSETUJUAN

Judul : Identifikasi Potensi *Logical Fallacy* Pada Komentar Sosial Media Menggunakan IndoBERT
Kategori : Skripsi
Nama Mahasiswa : Firdha Isly Ramadhani
Nomor Induk Mahasiswa : 181402029
Program Studi : Sarjana (S-1) Teknologi Informasi
Fakultas : Ilmu Komputer dan Teknologi Informasi
Universitas Sumatera Utara

Medan, 16 Oktober 2024

Komisi Pembimbing:

Pembimbing 2



Dedy Arisandi S.T., M.Kom.

NIP. 197908312009121002

Pembimbing 1



Dr. Baihaqi Siregar S Si., M.T.

NIP. 197901082012121002

Diketahui/disetujui oleh

Program Studi SI Teknologi Informasi

Ketua,



Dedy Arisandi S.T., M.Kom.

NIP. 197908312009121002

PERNYATAAN**IDENTIFIKASI POTENSI LOGICAL FALLACY PADA KOMENTAR
SOSIAL MEDIA MENGGUNAKAN INDOBERT****SKRIPSI**

Dengan ini saya menyatakan bahwa skripsi ini adalah karya yang saya buat sendiri, terkecuali pada beberapa bagian ataupun tulisan yang sumbernya disebutkan pada skripsi ini. ini saya menyatakan bahwa skripsi ini adalah karya yang saya buat sendiri, terkecuali pada beberapa bagian ataupun tulisan yang sumbernya disebutkan pada skripsi ini.

Medan, 16 Oktober 2024



Firdha Isly Ramadhani

UCAPAN TERIMA KASIH

Puji beserta syukur penulis panjatkan kepada Allah SWT atas rahmat dan berkat yang dilimpahkan oleh-Nya sehingga penyusunan skripsi sebagai salah satu syarat dalam menyelesaikan gelar S1 Program Studi Teknologi Informasi Universitas Sumatera Utara dapat penulis selesaikan.

Tanpa adanya bimbingan, bantuan, dorongan, dan doa dari berbagai pihak skripsi ini tidak akan dapat diselesaikan. Maka dari itu, pada kesempatan ini dengan ketulusan juga kerendahan hati penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Kedua orang tua penulis Ayah Iswan dan Ibu Lily Carolina Batubara yang senantiasa melimpahi doa, kasih sayang, juga semangat sehingga skripsi ini dapat diselesaikan oleh penulis. Terima kasih untuk semua yang telah diberikan.
2. Bapak Dr. Baihaqi Siregar S.Si., M.T., selaku Dosen Pembimbing I yang selama ini telah meluangkan waktu dalam membimbing dan memberikan masukan kepada penulis untuk menyelesaikan skripsi ini.
3. Bapak Dedy Arisandi S.T., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan sebagai Dosen Pembimbing II untuk bimbingan dan waktu yang telah diluangkan selama proses penulisan skripsi.
4. Ibu Dr. Erna Budhiarti Nababan M.IT selaku Dosen Pembimbing I dan Ibu Ade Sarah Huzaifah S.Kom., M.Kom selaku Dosen Pembimbing II yang telah memberi masukan guna menyempurnakan skripsi.
5. Bapak Dr. Muryanto Amin, S.Sos., M.Si., selaku Rektor Universitas Sumatera Utara.
6. Ibu Dr. Maya Silvi Lydia B.Sc., M.Sc., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
7. Dosen-Dosen Program Studi Sarjana Teknologi Informasi yang selama ini telah mengajarkan ilmu-ilmu yang berharga selama masa perkuliahan.
8. Pegawai Program Studi Sarjana Teknologi Informasi USU dan Fakultas Ilmu Komputer dan Teknologi Informasi yang selama ini membantu penulis dalam kegiatan akademik maupun administrasi.

9. Kepada penulis sendiri yang tetap berusaha dan tidak menyerah selama proses pengerjaan tugas akhir walaupun banyak rintangan yang dihadapi.
10. Abang-abang tercinta Ai dan Andre yang telah memberikan dukungan kepada penulis selama masa perkuliahan.
11. Kepada teman-teman di kala suka maupun duka selama masa perkuliahan Fuji Alita, Sonia Rai, Willi Nardo, Maisyaroh Ramadhani, dan Aulia Nisa. Terima kasih telah menjadi teman yang selalu membantu juga memotivasi penulis dalam menyelesaikan skripsi ini.
12. Kepada pihak-pihak yang selama ini telah membantu dalam penyelesaian skripsi ini yang tidak dapat disebutkan satu persatu.

Medan, 16 Oktober 2024



Firdha Isly Ramadhani

181402029

ABSTRAK

Penggunaan sosial media dan internet secara umum di Indonesia semakin berkembang, diketahui bahwa sebanyak 60,4% penduduk Indonesia adalah pengguna aktif sosial media aktif. Sosial media memiliki banyak manfaat seperti mencari saran, berbagi informasi, dan bertukar opini dengan sesama pengguna sosial media lainnya. Tetapi, berdasarkan penelitian-penelitian yang sebelumnya telah dilakukan diketahui bahwa sosial media memiliki peran signifikan dalam menyebarkan logical fallacy sehingga misinformasi dapat tersebar dengan lebih mudah, kualitas diskusi semakin buruk, dan dapat mempengaruhi opini publik secara negatif. Ketidakhahaman mengenai *logical fallacy* membuat pengguna sosial media rentan terhadap misinformasi dan manipulasi, hal ini cukup mengkhawatirkan karena komentar-komentar dengan *logical fallacy* yang marak di sosial media dapat menyebabkan miskomunikasi, polarisasi opini, dan memperburuk opini publik. Sehingga, perlu dibuat sebuah pendekatan yang dapat mengidentifikasi potensi *logical fallacy* yang ada pada komentar di sosial media agar dapat meningkatkan kesadaran pengguna sosial media terhadap *logical fallacy* sehingga dapat mencegah misinformasi dan manipulasi. Penelitian yang dilakukan adalah identifikasi menggunakan IndoBERT dengan lebih dari 14000 data yang telah dikumpulkan. Hasil akurasi dari model yang dilatih adalah 78,77%, di mana untuk kelas 'Fallacy' nilai F1-scorenya adalah 0.80 dan untuk kelas 'No fallacy' nilai F1-scorenya adalah 0.78.

Kata kunci: Identifikasi, IndoBERT, Logical Fallacy, Sosial Media.

IDENTIFICATION OF LOGICAL FALLACY POTENTIAL IN SOCIAL MEDIA COMMENT USING INDOBERT

ABSTRACT

Usage of social media and internet in general in Indonesia is growing, it is known that as many as 60,4% of Indonesian population are active sosial media users. Social media has many benefits such as seeking advice, sharing information, and exchanging opinions with other sosial media users. But, based on several studies that previously have been made, it is shown that social media has a notable role in spreading logical fallacy which causes misinformation to spread more easily, worsening quality of online discussions, and can negatively influence public opinion. The lack of understanding about logical fallacy makes social media users vulnerable to misinformation and manipulation, which is quite worrying because comments with logical fallacy that are rampant on social media can lead to miscommunication, polarization of opinions, and worsening public opinion. Thus, it is necessary to create an approach that can identify potential logical fallacy in social media comments in order to prevent misinformation and manipulation. The research conducted is identification using IndoBERT with more than 14000 data that has been collected. The accuracy result of the trained model is 78,77%, with 'Fallacy' class having an F1-score of 0.80 and 'No fallacy' class having an F1-score of 0.78.

Keywords: Identification, IndoBERT, Logical Fallacy, Social Media.

DAFTAR ISI

PERSETUJUAN	i
PERNYATAAN	ii
UCAPAN TERIMA KASIH	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	3
1.6. Sistematika Penulisan	4
BAB II LANDASAN TEORI	5
2.1. <i>Logical Fallacy</i>	5
2.2. Kalimat Singkatan	9
2.3. <i>Bidirectional Encoder Representations from Transformers (BERT)</i>	9
2.4. IndoBERT	17
2.5. Ekstensi Browser	17
2.6. Modelbit	18
2.7. Penelitian Terdahulu	18
BAB III ANALISIS DAN PERANCANGAN	25
3.1. Arsitektur Umum	25
3.2. Pengumpulan Dataset	27
3.2.1. <i>Shaping</i> Dataset	28

3.2.2. Penerjemahan Dataset	29
3.2.3. Pelabelan Dataset	30
3.3. <i>Preprocessing</i> Dataset	32
3.3.1. <i>Punctuation Removal</i>	32
3.3.2. <i>Normalization</i>	34
3.3.3. <i>Tokenization</i>	35
3.4. <i>Finetuning</i> IndoBERT	37
3.5. Evaluasi Model	37
3.6. Testing Model	38
3.7. Pemodelan Sistem	37
3.7.1. Diagram Aktivitas	39
3.7.2. Flowchart Sistem	41
3.8. Perancangan Ekstensi Browser	43
3.8.1. Rancangan Antarmuka <i>Landing Page</i>	43
3.8.2. Rancangan Antarmuka <i>Popup</i> Ekstensi	44
3.8.3. Rancangan Antarmuka Identifikasi <i>Tweet</i> dengan Ekstensi Browser	45
3.8.4. Rancangan Antarmuka Identifikasi Komentar Instagram dengan Ekstensi Browser	45
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	46
4.1. Implementasi Sistem	46
4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak	46
4.2. Implementasi Ekstensi Google Chrome	47
4.2.1. Tampilan <i>Landing Page</i> Ekstensi Browser	47
4.2.2. Tampilan <i>Popup</i> Ekstensi Browser	47
4.2.3. Tampilan Antarmuka Identifikasi <i>Tweet</i> Menggunakan Ekstensi Browser	48
4.2.4. Tampilan Antarmuka Identifikasi Komentar Instagram Menggunakan Ekstensi Browser	49
4.3. Implementasi IndoBERT	50
4.4. Evaluasi Model	52
4.5. Pengujian Sistem	61
BAB V KESIMPULAN DAN SARAN	64
5.1. Kesimpulan	64
5.2. Saran	64
DAFTAR PUSTAKA	65
LAMPIRAN	69

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	20
Tabel 3. 1 List Dataset Berdasarkan Sumber	27
Tabel 3. 2 Dataset Training	31
Tabel 3. 3 <i>Pseudocode Punctuation Removal</i>	33
Tabel 3. 4 Contoh Implementasi <i>Punctuation Removal</i>	34
Tabel 3. 5 <i>Pseudocode Normalization</i>	34
Tabel 3. 6 Contoh Implementasi <i>Normalization</i>	35
Tabel 3. 7 <i>Pseudocode Tokenization</i>	35
Tabel 3. 8 Contoh <i>Tokenization</i>	36
Tabel 3. 9 Contoh Tahap <i>Preprocessing</i>	38
Tabel 3. 10 Contoh Hasil Identifikasi Model	39
Tabel 4. 1 Hasil Evaluasi Model	52
Tabel 4. 2 Hasil Pengujian Ekstensi	61

DAFTAR GAMBAR

Gambar 2. 1 Encoder dan decoder (Vaswani et al., 2017)	10
Gambar 2. 2 Proses <i>self-attention</i> (Alammar, 2018)	12
Gambar 2. 3 Proses <i>encoder</i> (Alammar, 2018)	12
Gambar 2. 4 Perbedaan ukuran BERT (Alammar, 2018)	13
Gambar 2. 5 Arsitektur BERT (Alammar, 2018)	14
Gambar 2. 6 Proses Masked Language Modelling (Fimoza, 2019)	15
Gambar 2. 7 Proses Pre-training pada BERT (Fimoza, 2019)	16
Gambar 2. 8 Representasi Input pada BERT (Devlin, 2019)	16
Gambar 2. 9 Prosedur Pre-training dan Fine-tuning (Devlin, 2019)	16
Gambar 3. 1 Arsitektur Umum	26
Gambar 3. 2 Dataset Sosial Media yang Telah Dikumpulkan	27
Gambar 3. 3 Struktur Asli Dataset	29
Gambar 3. 4 Struktur Asli Setelah Melalui Proses Shaping	29
Gambar 3. 5 Dataset yang Belum Diterjemahkan	29
Gambar 3. 6 Dataset yang Telah Diterjemahkan	30
Gambar 3. 7 Diagram Aktivitas Ekstensi	40
Gambar 3. 8 Flowchart Training	41
Gambar 3. 9 Flowchart Ekstensi Browser	42
Gambar 3. 10 Rancangan Antarmuka Landing Page Ekstensi Browser	44
Gambar 3. 11 Rancangan Antarmuka Popup Ekstensi Browser	44
Gambar 3. 12 Rancangan Antarmuka Identifikasi Potensi Logical Fallacy pada Tweet Menggunakan Ekstensi Browser	45
Gambar 3. 13 Rancangan Antarmuka Identifikasi Potensi Logical Fallacy pada Komentar Instagram dengan Ekstensi Browser	45
Gambar 4. 1 Tampilan Antarmuka <i>Landing Page</i> Ekstensi Browser	47
Gambar 4. 2 Tampilan Antarmuka <i>Popup</i> Ekstensi Browser	48
Gambar 4. 3 Tampilan Antarmuka Identifikasi <i>Tweet</i> Menggunakan Ekstensi Browser	49
Gambar 4. 4 Tampilan Antarmuka Identifikasi Komentar Instagram Menggunakan Ekstensi Browser	49
Gambar 4. 5 Hasil Pelatihan 8 Epoch	50
Gambar 4. 6 Hasil Pelatihan 16 Epoch	51
Gambar 4. 7 Hasil Pelatihan 32 Epoch	51
Gambar 4. 8 Perbandingan Nilai Akurasi Dengan Berbagai <i>Batch Size</i> Dan Epoch	51
Gambar 4. 9 Hasil Performa <i>Hypertuning</i> 32 Epoch <i>Batch Size</i> 16	52
Gambar 4. 10 <i>Confusion Matrix</i>	60

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring dengan perkembangan zaman kemajuan media komunikasi seperti sosial media terus meningkat dengan pesat dan tidak dapat dihentikan. Pengguna internet di Indonesia mencapai 212,9 juta orang dengan 167 juta di antaranya adalah pengguna sosial media aktif, di mana angka tersebut mencakup 60,4% dari populasi Indonesia. Masyarakat Indonesia sendiri juga merupakan pengguna sosial media yang aktif yang pada umumnya menghabiskan rata-rata waktu sebanyak 3 jam 11 menit per harinya (Riyanto, 2024).

Terdapat banyak manfaat yang diberikan oleh sosial media salah satunya dengan memungkinkan penggunanya untuk mengakses dan bertukar informasi dengan sangat mudah dan cepat menjadikan sosial media sebagai tempat yang sesuai untuk bertukar pikiran dengan berdiskusi mengenai berbagai macam hal dengan orang lain. Hal ini juga membantu pengguna sosial media dalam meningkatkan kemampuan berpikir kritis karena memungkinkan penggunanya untuk meningkatkan keterampilan berpikir tingkat tinggi dengan menilai kredibilitas dari sebuah sumber informasi (Hidayat et al. 2021).

Namun, meskipun sosial media menawarkan banyak manfaat dalam meningkatkan kemampuan berpikir kritis penggunanya sosial media tidak lepas dari dampak negatif, tidak sedikit penelitian yang telah dilakukan menyatakan bahwa sosial media memiliki peran yang signifikan dalam menyebarkan logical fallacy atau kesalahan dalam berpikir (Hidayat et al. 2021). Tetapi, banyak pengguna sosial media tidak sadar akan penggunaan logical fallacy di sosial media sebagai taktik untuk memanipulasi, hal ini cukup mengkhawatirkan karena komentar-komentar dengan logical fallacy yang marak di sosial media dapat menyebabkan miskomunikasi, misinformasi, manipulasi, polarisasi opini, memperburuk ruang untuk berdiskusi di internet, dan bahkan berperan dalam penyebaran propaganda (Da San Martino, 2019).

Logical fallacy sendiri adalah kesalahan dalam bernalar, baik disengaja ataupun tidak, yang terjadi ketika suatu argumen tersusun dari premis-premis yang tidak berkaitan dengan kesimpulan yang ada. *Logical fallacy* merupakan taktik yang sering digunakan dalam berargumentasi, ketika digunakan *logical fallacy* tidak selalu terkait dengan isi kalimatnya, tetapi kepada metode yang digunakan ketika berargumentasi.

Penelitian terdahulu yang berkaitan dengan identifikasi *logical fallacy* telah dilakukan sebelumnya, salah satunya ialah penelitian yang dilakukan oleh Ruiz-Dolz et al. (2023). Penelitian dilakukan dengan tujuan untuk mendeteksi *fallacy* pada argumen di luar domain training. Peneliti menggunakan berbagai macam model seperti eSVM, GPT-3.5-TURBO, GPT- 4, dan RoBERTa dengan melakukan tahapan mendeteksi apakah suatu kalimat termasuk *logical fallacy* lalu mengklasifikasikan *fallacy* kalimat tersebut. Hasil penelitian menunjukkan bahwa RoBERTa mendapatkan hasil tertinggi dalam mendeteksi *logical fallacy* dengan akurasi sebesar 79,8% dan klasifikasi *logical fallacy* sebesar 76,2%.

Penelitian terdahulu yang berkaitan dengan identifikasi juga pernah dilakukan oleh Situmorang et al. (2024). Penelitian ini menggunakan IndoBERT untuk mengidentifikasi potensi depresi dari unggahan di sosial media dan mendapatkan hasil F1-score 94,91%. Tetapi penelitian tersebut berfokus pada unggahan yang mengandung potensi depresi, sehingga penelitian ini dapat menjadi kontribusi baru dalam klasifikasi *logical fallacy* berbahasa Indonesia.

Berdasarkan latar belakang dan beberapa penelitian terdahulu yang telah dijelaskan, penulis hendak melakukan penelitian dengan judul “Identifikasi Potensi *Logical Fallacy* Pada Sosial Media Menggunakan IndoBERT”.

1.2. Rumusan Masalah

Sosial media memiliki peran signifikan dalam menyebarkan *logical fallacy*, sebanyak 60% penduduk Indonesia adalah pengguna sosial media aktif yang menghabiskan setidaknya 3 jam perhari menggunakan sosial media, sehingga dampak dari interaksi komentar yang dikonsumsi memiliki pengaruh yang besar dalam mempengaruhi persepsi pengguna sosial media, padahal penyebaran *logical fallacy* menyebabkan pengguna sosial media rentan terhadap misinformasi, manipulasi, dan bahkan propaganda. Oleh karena itu, diperlukan sebuah pendekatan yang membantu pengguna sosial media untuk mengidentifikasi komentar yang berpotensi mengandung *logical fallacy*.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk mengembangkan ekstensi browser yang dapat mengidentifikasi potensi logical fallacy pada komentar di sosial media dengan menggunakan IndoBERT.

1.4. Batasan Masalah

Batasan-batasan masalah yang dicakup pada penelitian ini untuk menghindari perluasan ruang lingkup yang tidak dibutuhkan yaitu sebagai berikut:

1. Dataset yang dipakai berupa teks dalam Bahasa Indonesia.
2. Dataset yang digunakan diperoleh dari penelitian-penelitian sebelumnya dan data yang di-scraping dari sosial media seperti Instagram, Twitter, dan Reddit.
3. Ekstensi browser yang dibuat hanya ekstensi browser Google Chrome.
4. Ekstensi browser hanya bisa dijalankan pada website Instagram dan Twitter.
5. Model hanya akan mengidentifikasi ada tidaknya potensi logical fallacy pada sebuah kalimat.
6. Penelitian ini hanya mengidentifikasi kalimat dalam Bahasa Indonesia.
7. Tipe logical fallacy yang dapat diidentifikasi potensinya oleh model adalah *fallacy ad hominem, appeal to authority, appeal to emotion, appeal to ignorance, appeal to majority, appeal to nature, appeal to tradition, appeal to worse problems, circular reasoning, false causality, false dilemma, hasty generalization, red herring, slippery slope, dan strawman*.

1.5. Manfaat Penelitian

Hasil yang diperoleh dalam penelitian diharapkan untuk mencapai beberapa manfaat sebagai berikut:

1. Membantu pengguna sosial media untuk mengidentifikasi komentar yang berpotensi mengandung *logical fallacy*.
2. Mengetahui tingkat efektifitas IndoBERT dalam mengidentifikasi potensi *logical fallacy* pada komentar ataupun pernyataan dari sosial media.
3. Sebagai referensi penelitian identifikasi potensi *logical fallacy* pada kalimat berbahasa Indonesia.

1.6. Metodologi Penelitian

Metodologi penelitian ini mencakup tahap-tahapan berikut Pada pengerjaan penelitian mengidentifikasi potensi *logical fallacy* di sosial media terdapat beberapa tahapan metode yang digunakan oleh penulis. Tahapan-tahapan yang dilakukan pada penelitian ini meliputi lima bagian sebagai berikut:

Bab I: Pendahuluan

Bab I meliputi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, serta sistematika penulisan.

Bab II: Landasan Teori

Bab II meliputi penjelasan landasan teori mengenai *logical fallacy*, tipe-tipe *logical fallacy*, BERT, IndoBERT, dan lain-lain.

Bab III: Metodologi Penelitian

Bab III meliputi arsitektur umum yang digunakan pada penelitian ini. Pada bab ini akan dijelaskan tahap-tahap perancangan, seperti dataset yang digunakan pada penelitian ini, tahap *preprocessing*, dan juga tahap perancangan antarmuka dari sistem aplikasi yang dijelaskan beserta dengan gambarnya.

Bab IV: Implementasi dan Pengujian Sistem

Bab IV meliputi implementasi serta pengujian sistem dari perancangan pada Bab III yang telah dijelaskan sebelumnya. Pada bab ini dilakukan proses *finetuning* yang dijelaskan dengan grafik untuk mendapatkan model dengan akurasi yang cukup tinggi, hasil dari pengujian juga akan dijelaskan pada bab ini.

Bab V: Kesimpulan dan Saran

Bab V akan meliputi kesimpulan dari seluruh pembahasan bab yang sebelumnya telah dijelaskan beserta dengan saran dari penulis mengenai perbaikan untuk penelitian kedepannya.

BAB II

LANDASAN TEORI

2.1. *Logical Fallacy*

Menurut Copi, *logical fallacy* timbul tidak dikarenakan adanya kesalahan atas ide maupun kenyataan yang disampaikan. Yang membedakan suatu argumen mengandung *logical fallacy* atau tidak ditentukan dari hubungan antara premis dan kesimpulan yang dibuat. Apabila premis dari suatu argumen adalah benar namun kesimpulan yang dihasilkan tidak mendapatkan konklusi yang sesuai maka dapat dikatakan bahwa argumen tersebut tetaplah dianggap sebagai fallacy (Copi, 1990).

Berdasarkan penelitian-penelitian sebelumnya, terdapat beberapa tipe logical fallacy yang sering digunakan dan telah diidentifikasi dalam diskusi pada media daring (Sahai et al., 2021), *logical fallacy* tersebut ialah *ad hominem*, *appeal to authority*, *appeal to emotion*, *appeal to ignorance*, *appeal to majority*, *appeal to nature*, *appeal to tradition*, *appeal to worse problems*, *circular reasoning*, *false causality*, *false dilemma*, *hasty generalization*, *red herring*, *slippery slope*, dan *strawman*. Tipe-tipe *logical fallacy* ini menjadi tipe yang akan diidentifikasikan pada penelitian ini:

a) *Ad Hominem*

Ad hominem merupakan tipe *logical fallacy* yang memiliki ciri khas di mana argumen yang disampaikan bukan menyerang argumen dari lawan bicara tetapi menyerang karakter dari lawan bicara itu sendiri dengan tujuan pertahanan diri ataupun untuk membantah argumen lawan bicara tersebut (Copi, 1990). Contoh: 'Kamu adalah seorang pria. Aku tidak ingin mendengar apa yang kamu katakan. Kamu tidak bisa mengerti apa yang wanita rasakan.' (Mayfield, 2014).

b) *Appeal to Authority*

Appeal to authority merupakan salah satu tipe *logical fallacy* di mana argumen yang disampaikan mengacu kepada sosok atau individu yang dihormati. Tetapi, kesalahannya terjadi ketika seseorang menggunakan pernyataan dari tokoh

yang dihormati tersebut sebagai sebuah kesimpulan tanpa adanya alasan atau dukungan rasional lainnya (Copi, 1990). Contoh: 'Kata dokter kamu harus menggosok gigimu setiap hari dengan Florident.' (Mayfield, 2014).

c) *Appeal to Emotion*

Appeal to emotion adalah sebuah *fallacy* yang terjadi ketika sebuah emosi seperti rasa kasihan, takut, marah, ataupun emosi lainnya digunakan untuk mendukung sebuah argumen tanpa ada alasan logis lain yang mendukung. Contoh: 'Kamu harus menikahiku. Aku tahu kita tidak kompatibel, tapi kamu adalah satu-satunya opsiku.' (Jin et al., 2022).

d) *Appeal to Ignorance*

Merupakan sebuah *fallacy* yang terjadi ketika sebuah hal dianggap sebagai benar hanya karena hal tersebut belum dibuktikan sebagai salah, ataupun hal tersebut dianggap sebagai salah hanya karena belum dibuktikan sebagai benar (Copi, 1990). Contoh: 'Sampai hari ini, sains belum mampu menciptakan kehidupan dari benda mati; oleh karena itu, kehidupan pasti merupakan hasil campur tangan ilahi.' (Logically Fallacious, n.d.).

e) *Appeal to Majority*

Appeal to majority merupakan sebuah *logical fallacy* di mana argumen menyatakan suatu kesimpulan sebagai benar karena sebagian besar atau bahkan kelompok elit berpikir, memercayai, atau merasa bahwa kesimpulan tersebut benar tanpa adanya bukti-bukti yang relevan (Archie, 1997). Contoh: 'Bagaimana mungkin kamu tidak percaya pada kelahiran dari seorang gadis suci? Sekitar dua miliar orang memercayainya, tidakkah kamu pikir kamu harus mempertimbangkan kembali posisi kamu?' (Logically Fallacious, n.d.).

f) *Appeal to Nature*

Appeal to nature adalah sebuah *fallacy* yang digunakan ketika terdapat keyakinan atau saran bahwa sesuatu yang "alami" pasti lebih baik daripada "tidak alami" didasarkan pada sifat kealamiannya. Hal ini terjadi ketika terdapat sebuah keyakinan bahwa apa yang alami pastilah baik dan apa yang tidak alami pastilah. Contoh: 'Saya berbelanja di Natural Happy Sunshine Store (NHSS), yang jauh lebih baik daripada toko bahan makanan kamu karena di NHSS semuanya alami.' (Anonim, 2024).

g) *Appeal to Worse Problems*

Adalah sebuah *fallacy* yang terjadi ketika sebuah argumen berusaha untuk membuat sebuah scenario menjadi lebih buruk atau lebih baik dengan membandingkannya terhadap skenario terburuk. Contoh: ‘Bersyukur saja dengan mobil Chevy Nova 1972 yang kamu kendarai. Banyak orang di negara ini yang bahkan tidak mempunyai mobil.’ (Anonim, 2024).

h) *Appeal to Tradition*

Appeal to tradition adalah sebuah *fallacy* yang terjadi ketika sebuah argumen atau tindakan dianggap sebagai benar hanya karena tindakan ataupun argument tersebut selalu dianggap demikian di masa lalu. Contoh: ‘Menjadi vegan itu tidak masuk akal karena nenek moyang kita sejak dulu adalah pemakan daging.’ (Sahai et al., 2021)

i) *Circular Reasoning*

Circular reasoning terjadi ketika sebuah kesimpulan dari sebuah argumen dinyatakan atau diasumsikan pada salah satu premisnya (Copi, 1990). Contoh: ‘Dia adalah kandidat presiden terbaik karena dia lebih baik dari kandidat lainnya.’ (Jin et al., 2022).

j) *False Causality*

False causality adalah *logical fallacy* yang terjadi ketika sebuah argumen menyatakan bahwa sesuatu yang sebenarnya bukan penyebab dari sebuah hal lain dianggap sebagai penyebabnya (Copi, 1990). *Fallacy* ini memiliki pola tipikal yaitu “ α terjadi dengan $\beta \Rightarrow \alpha$ menyebabkan β .” Contoh: ‘Tiap kali aku mencuci mobilku hujan selalu turun. Kegiatan saya mencuci mobil memiliki pengaruh yang pasti terhadap cuaca.’ (Jin et al., 2022).

k) *False Dilemma*

False dilemma adalah kesalahan logika yang terjadi ketika terdapat dua opsi ekstrem sebagai satu-satunya kemungkinan yang disampaikan, padahal pada kenyataannya lebih banyak pilihan yang ada. Kesalahan ini juga dikenal sebagai *black-or-white fallacy*. *False dilemma* sering digunakan pada argumen untuk memaksakan keputusan di antara dua opsi yang dianggap tidak menarik untuk menyepelekan isu kompleks yang ada. *False dilemma* sering digunakan dalam politik, periklanan, dan debat. Contoh: ‘Kamu harus

mendukung Undang-Undang ini atau dianggap sebagai pengkhianat.' pada contoh ini pembicara mengisyaratkan bahwa hanya ada dua opsi yang layak, padahal pada kenyataannya terdapat banyak pilihan lain seperti abstain dari pemungutan suara, atau menolak RUU tersebut (Logical Fallacies, n.d.).

l) *Hasty Generalization*

Hasty generalization adalah kesalahan dalam penalaran yang menggunakan sampel yang terlalu kecil atau bukti yang tidak representatif untuk membuat penilaian yang tergesa-gesa terhadap suatu kelompok. (Young, 2006). Contoh: 'Tidak ada mekanik mobil yang bisa dipercaya. Mereka hanya ingin menghasilkan uang.' (Mayfield, 2014).

m) *Red Herring*

Red herring terjadi ketika lawan bicara berusaha untuk mengalihkan perhatian dari argumen atau masalah yang sedang didiskusikan (Copi, 1990). Contoh: 'Kita harus memindahkan kantor ke California untuk memperluas potensi pelanggan yang ada. Cuaca di sana juga lebih hangat, yang menjadi alasan kuat untuk pindah ke sana.' (Jin et al., 2022).

n) *Slippery slope*

Slippery slope adalah *logical fallacy* yang terjadi ketika argumen yang menyatakan bahwa suatu peristiwa kecil A memiliki konsekuensi besar yang tidak diinginkan yaitu C. Setidaknya ada satu peristiwa B lagi dalam rantai kausalitas (A akan menyebabkan B, B akan menyebabkan C). Contoh: 'Kalau kamu melanggar diet dan makan satu kue malam ini, besok kamu pasti mau makan 10 kue dan 20 kue pada lusa, dan sebelum kamu sadar, berat badan kamu akan naik kembali setelah turun 15kg.' (Sahai et al., 2021).

o) *Straw Man*

Straw man merupakan sebuah *logical fallacy* yang terjadi ketika suatu argumen dibuat lawan bicara menjadi lebih ekstrim atau bahkan tidak masuk akal daripada argumen yang sebenarnya (Copi, 1990). Contoh: 'Pembicara 1: Saya berpendapat bahwa kita harus memiliki layanan kesehatan universal gratis. Pembicara 2: Negara-negara komunis telah mencobanya. Kita tidak ingin Amerika menjadi negara komunis, jadi kita tidak boleh memiliki layanan kesehatan universal gratis.' (Logically Fallacious, n.d.)

2.2. Kalimat Singkatan

Perkembangan bahasa di sosial media seiring waktu membuat penggunaannya menjadi semakin kasual dan tidak formal dengan adanya penyingkatan terhadap kata formal yang dipakai. Kata yang disingkat tersebut akan tetap memiliki arti yang sama dengan bentuk formalnya, tetapi kata yang disingkat terkadang menjadi suatu faktor dalam terjadinya duplikasi kata yang menyebabkan turunnya akurasi.

Ada tiga metode untuk memendekkan kata menurut Tarigan (1995), yaitu penghilangan fonem, kontraksi, dan akronim. Penghilangan fonem terbagi menjadi 3 yang mencakup Afaresis, yaitu proses pengurangan setidaknya satu fonem pada awal kata, misalnya kata ‘tetapi’ menjadi ‘tapi’, Sinkop yaitu tahap pengurangan satu maupun lebih fonem di tengah kata, contohnya ‘dahulu’ diubah menjadi ‘dulu’, serta Apokop yaitu proses penanggalan fonem pada akhiran kata, contohnya ‘kasih’ menjadi ‘kasi’.

Kontraksi adalah penyingkatan kata atau frasa tanpa memodifikasi artinya. Contoh kontraksi yaitu "tidak ada" menjadi "tiada". Biasanya, dalam bahasa Inggris, kontraksi ditandai dengan penggunaan tanda petik, misalnya "I have" menjadi "I've". Akronim adalah bentuk penyingkatan di mana setiap huruf awal dari sebuah frasa diambil untuk membentuk sebuah kata yang dapat diucapkan atau dibaca. Terkadang, masyarakat lebih mengetahui bentuk akronim dibandingkan dengan bentuk lengkapnya. Contoh dari akronim adalah, mayor jenderal disingkat menjadi mayjen.

2.3. *Bidirectional Encoder Representations from Transformers (BERT)*

Bidirectional Encoder Representation from Transformers atau BERT adalah suatu teknik berbasis neural network yang digunakan untuk melakukan *pre-training* pada *natural language*. BERT dibangun dengan tujuan untuk mempelajari bahasa dengan kalimat yang bersifat ambigu menggunakan teks di sekitarnya untuk memahami konteks dengan lebih jelas (Luo et al., 2019).

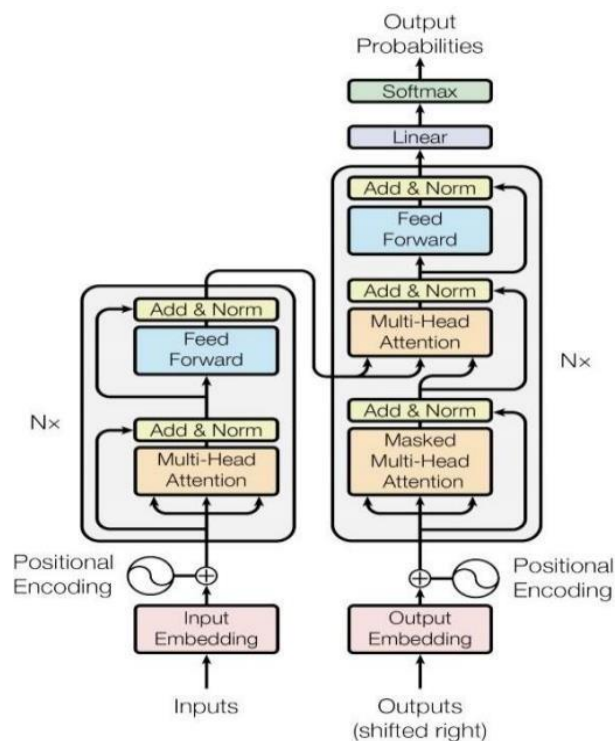
BERT dibangun menggunakan transformer, suatu mekanisme yang dapat mempelajari konteks pada kalimat dengan cara memahami hubungan di antara kata-kata pada teks secara kontekstual. Transformer mempelajari serta memperbaiki pemahaman yang didapatkan melalui mekanisme self-attention. Mekanisme self-attention merupakan cara transformer memodifikasi kata-kata yang berhubungan serta mengubah representasinya berdasarkan pemahaman tersebut. Transformer juga memiliki dua mekanisme yang bernama encoder dan decoder (Ganesh et al., 2021).

a) *Encoder*

Encoder adalah mekanisme yang berfungsi untuk membaca data masukan teks yang memiliki 6 *layer* identik. Semua *layer encoder* terdiri dari *layer self-attention* dan *feed forward neural network*. *Layer self-attention* membantu *encoder* untuk membuat *node* fokus pada kata yang divisualisasikan dan menangkap konteks yang ada pada kata tersebut. Tiap posisi yang ada pada *encoder* juga dapat memproses semua posisi *layer* sebelumnya.

b) *Decoder*

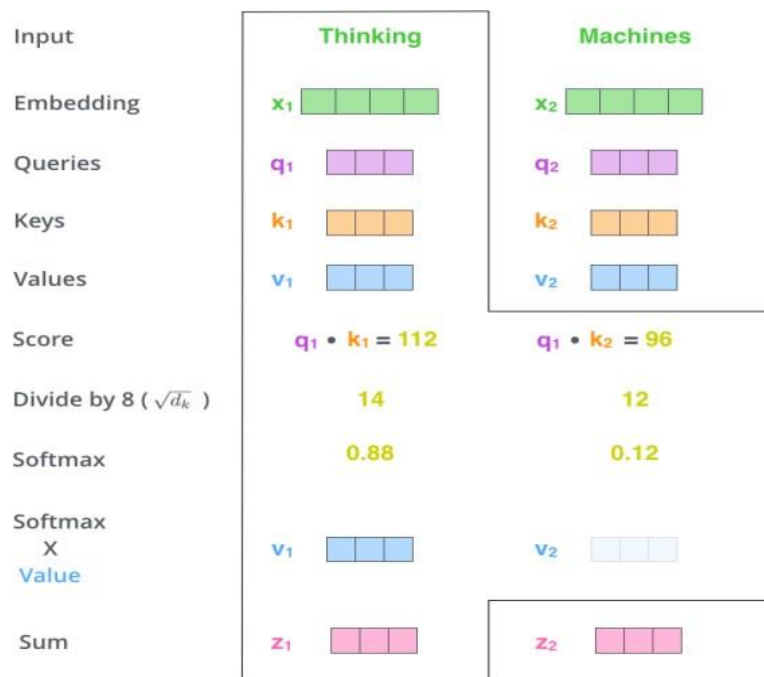
Decoder adalah mekanisme yang digunakan untuk menghasilkan prediksi urutan keluaran. *Decoder* juga terdiri dari 6 *layer* yang bisa diidentifikasi. Mirip dengan *encoder*, tiap *layer* pada *decoder* juga terdiri dari 2 *sublayer*. Adanya tambahan *attention layer* di antara *encoder* dan *decoder* membantu *node* untuk mengakses konten utama yang diinginkan dengan melakukan perhatian *multi-head* pada hasil dari *encoder*. *Layer self-attention* di *decoder* juga memungkinkan tiap posisi untuk memproses posisi saat ini dan semua *layer* sebelumnya di *decoder*.



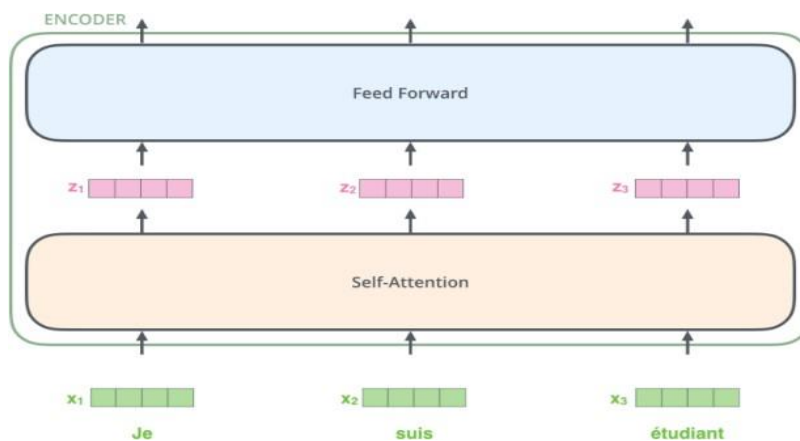
Gambar 2. 1. *Encoder dan decoder* (Khan et al., 2022)

Proses yang terjadi pada encoder dan decoder dapat dilihat seperti pada Gambar 2. 1 (Khan et al., 2022):

1. Tiap input yang melalui proses *encoder* akan menjadi daftar vektor yang menggunakan *embeddings* karena tidak adanya perbedaan urutan dari kata-kata sebuah kalimat pada *layer self-attention*, *positional encoding* digunakan agar posisi pada kata-kata tersebut ditampilkan. Semua vektor berukuran 512. Proses ini terjadi hanya pada *encoder* terbawah, sebab itu *encoder* lainnya dapat mendapatkan hasil *encoder* pertama.
2. Setelah itu, input vektor akan melalui lapisan *self-attention* dan *neural network feed-forward* pada *encoder*. Pada *layer self-attention* akan dibuat tiga vektor berdasarkan input vektor yaitu *query*, *key*, dan *value vector* dengan mengalikan *embedding*. Setiap vektor memiliki 64 dimensi. Kemudian, nilai *self-attention* dihitung untuk setiap istilah dengan mengalikan *query vektor* dan *key vector*, sebagaimana yang diperlihatkan pada Gambar 2.2. Lalu, karena dimensi tiap vektor adalah 64 maka nilai *self-attention* akan dibagi menjadi 8 karena 8 adalah akar kuadrat dari 64. *Softmax* juga digunakan untuk menghitung nilai *self attention*, sehingga tiap *value vector* akan dikalikan sesuai nilai menurut *softmax*. Hasilnya ialah *value vector* tersebut kemudian dijumlahkan untuk menghasilkan output dari *layer self-attention*. Output yang didapatkan dari *layer self-attention* akan masuk ke *layer feed-forward* untuk semua posisi, sebagaimana dalam Gambar 2. 2.



Gambar 2. 2. Proses *self-attention* (Alammar, 2018)

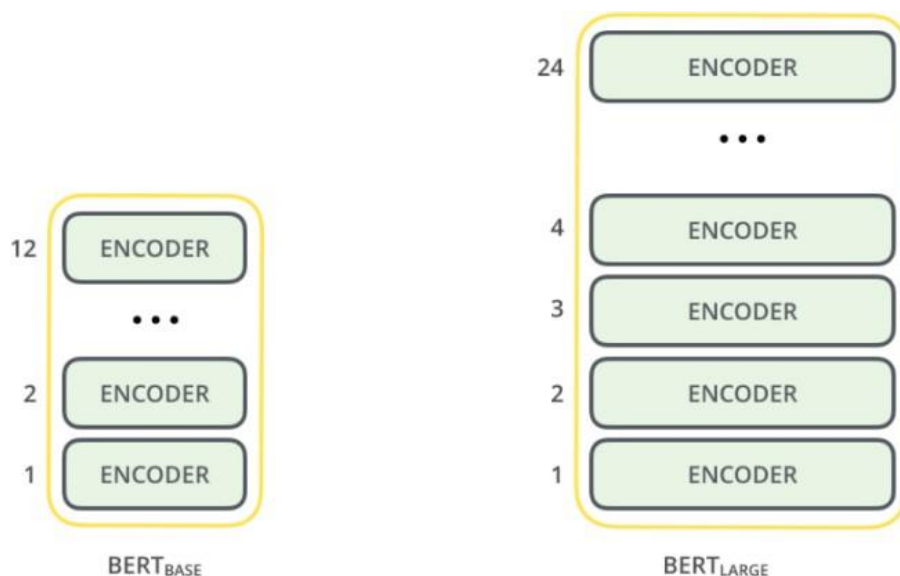


Gambar 2. 3. Proses *encoder* Alammar (2018)

- Setelah seluruh proses di *encoder* selesai, maka hasil yang didapatkan adalah *vector key* dan juga *vector value* yang kemudian akan diteruskan ke decoder. Semua input dan juga output dari encoder dan decoder, termasuk lapisan self-awareness dan neural network feedforward, diproses melalui lapisan add dan norm, yang melibatkan struktur residual dan juga normalisasi. Proses pada decoder mirip dengan proses di encoder, namun dengan penambahan *self*-

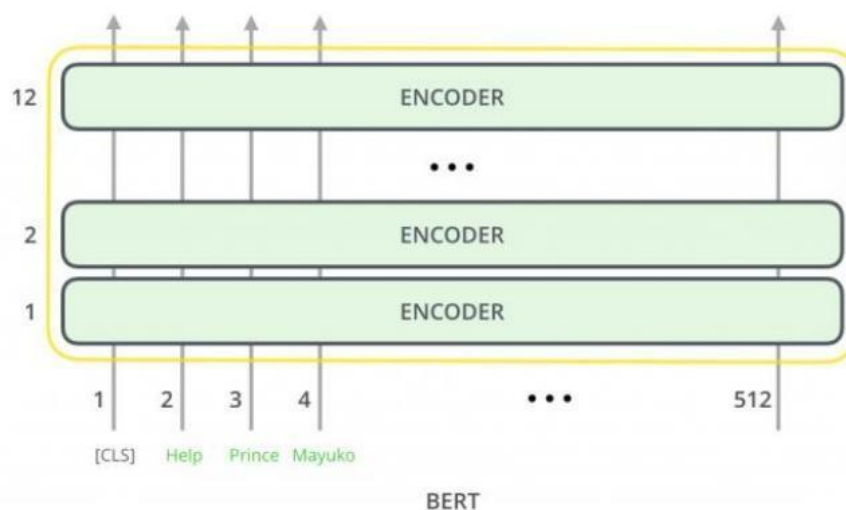
attention di lapisan *self-aware* dan *neural network feedforward* untuk memfokuskan decoder pada kata-kata yang relevan. Lapisan *self-aware* pada decoder hanya memperhatikan posisi output sebelumnya. Setiap hasil dari langkah ini kemudian dikirimkan ke decoder, dan hasil akhir decoder adalah sama dengan hasil dari encoder. Terakhir, didapatkan vektor bernilai *floating point* sebagai *output* dari rangkaian *decoder*. Namun, diperlukan lapisan tambahan dari lapisan terhubung penuh dan *layer softmax* untuk menerapkannya ke dalam kata-kata.

BERT memiliki arsitektur multi-layer biridirectional transformer yang mirip dengan implimentasi asli transformer, tetapi proses yang dilakukan hanya sampai tahap encoder. BERT dibagi menjadi dua model yaitu BERT-large dan BERT-base. BERT-base terdiri dari 12 layer, 12 self-attentions head, hidden size sebanyak 768, juga parameter sebanyak 110M. Untuk BERT-large sendiri terdiri dari 24 layer, 16 self-attention heads, hidden size sebanyak 1024, serta parameter sebanyak 340M.



Gambar 2. 4. Perbedaan ukuran BERT (Alammar, 2018)

Arsitektur BERT terlihat seperti Gambar 2. 5 yang terdiri hanya dari *encoder*, berbeda dengan model terarah atau *bidirectional* yang mementingkan urutan dari teks dari kanan-ke-kiri, kiri-ke-kanan, atau gabungannya.



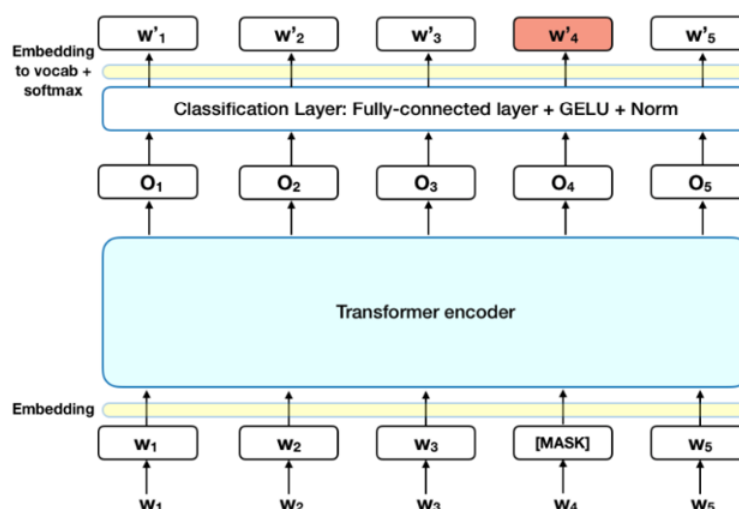
Gambar 2. 5. Arsitektur BERT (Alammar, 2018)

BERT merupakan model yang bisa dilatih agar mampu memahami bahasa dan dapat disempurnakan (*finetuning*) untuk berbagai macam tugas. Proses pelatihan BERT terdiri atas tahap *pre-training* dan tahap *fine-tuning*. Pada tahap *pre-training* BERT dilatih untuk memahami dan mempelajari bahasa beserta dengan konteksnya. Selama pelatihan ini, BERT menggunakan dua tugas *unsupervised* yang secara bersamaan dilakukan, di mana tugas tersebut ialah *Masked Language Model* dan *Next Sentence Prediction*.

1. *Masked Language Modelling (Masked LM)*

Masked Language Modeling bertujuan untuk menutupi beberapa kata dalam sebuah kalimat secara acak dengan probabilitas kecil menggunakan *mask*. Sebelum urutan kata diproses oleh BERT, sekitar 15% kata dalam setiap urutan akan diganti dengan token [MASK]. Model lalu akan berusaha menebak kata asli yang telah diganti dengan [MASK] berdasarkan konteks dari kata-kata lain yang tidak ditutup dengan [MASK] dalam urutan tersebut. Proses prediksi kata-kata ini melibatkan beberapa langkah seperti:

- i. Di atas *output encoder* dibutuhkan lapisan klasifikasi,
- ii. Vektor output dikalikan dengan matriks embedding, setelah itu diubah menjadi *vocabulary dimension*,
- iii. Probabilitas tiap kata pada *vocabulary* perlu dihitung dengan *softmax*.



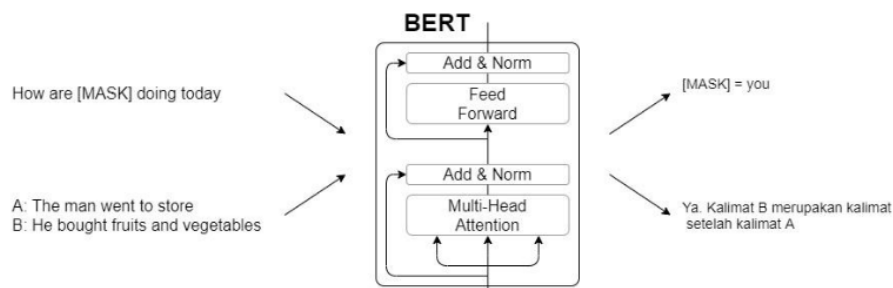
Gambar 2. 6. Proses *Masked Language Modelling* (Fimoza, 2019)

2. *Next Sentence Prediction*

Selama proses pelatihan BERT, model akan diberikan input berupa pasangan kalimat untuk diajarkan memprediksi apakah kalimat kedua dalam pasangan tersebut adalah kelanjutan dari kalimat pertama dalam dokumen asli atau berupa satu kalimat saja. Dalam proses pelatihan, 50% dari input akan berisi kalimat kedua yang merupakan kelanjutan langsung dari kalimat pertama dalam dokumen aslinya, sedangkan 50% sisanya adalah kalimat kedua yang diambil secara acak dari korpus sebagai kalimat kedua.

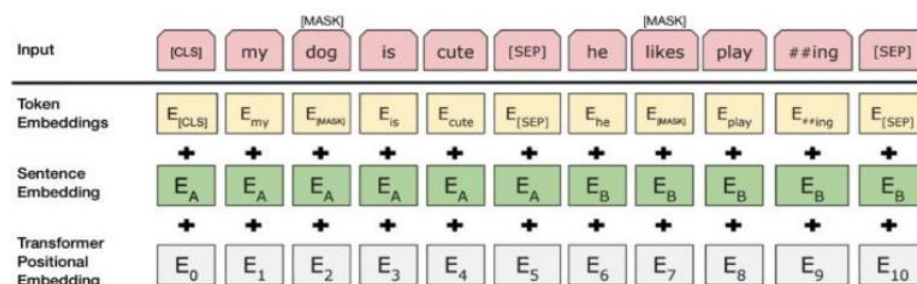
Terdapat tiga *embedding layers* pada BERT sebagai representasi input:

- i. Lapisan pertama yang dilalui oleh token adalah token embeddings, yaitu representasi vektor dari setiap token. Setiap token pada input lalu dihubungkan dengan vektor berdimensi tinggi sesuai dengan token tersebut. Kemudian, setiap token akan digantikan dengan sebuah id yang ditentukan berdasarkan kosakata.
- ii. *Sentence embeddings*, yang menandakan apakah token tersebut berasal dari kalimat pertama atau kedua, akan disertakan pada setiap token untuk membedakan antara dua kalimat jika terdapat lebih dari satu kalimat. Lapisan ini terdiri dari dua representasi: yaitu A sebagai token dalam kalimat pertama, dan B untuk token pada kalimat kedua.



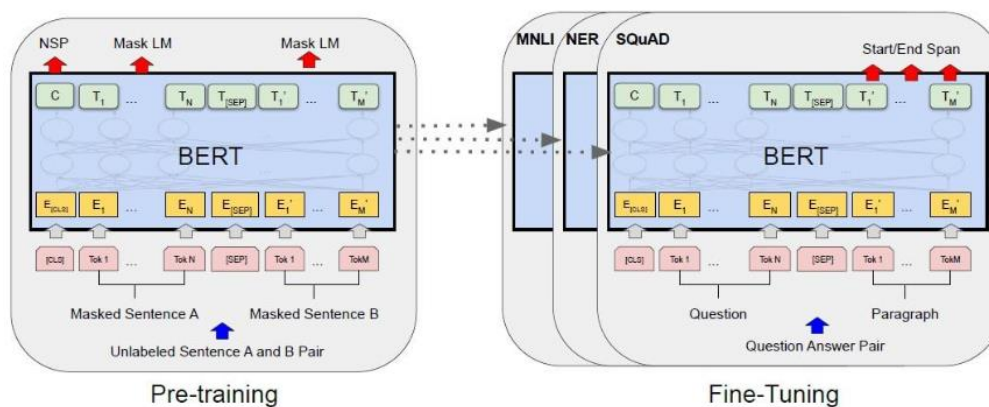
Gambar 2. 7. Proses Pre-training pada BERT (Fimoza, 2019)

- iii. *Positional Embedding* disertakan pada setiap token untuk menyimpan informasi tentang letak kata dalam urutan.



Gambar 2. 8. Representasi Input pada BERT (Devlin, 2019)

Dalam proses *training* model bahasa, classifier dilatih dengan adanya sedikit penyesuaian pada model BERT saat dilakukan *finetuning*. Pada penelitiannya, Devlin mengusulkan *hyperparameters* yang dapat dimodifikasi agar mendapatkan hasil yang optimal. Mekanisme *self-attention* di Transformer sangat memudahkan *finetuning*, sehingga membuat BERT optimal dalam membuat model untuk berbagai macam tugas.



Gambar 2. 9. Prosedur Pre-training dan Fine-tuning (Devlin, 2019)

2.4. IndoBERT

IndoBERT merupakan sebuah model natural language processing khusus Bahasa Indonesia yang dikembangkan berdasarkan arsitektur BERT (*Bidirectional Encoder Representations from Transformers*). IndoBERT dikembangkan atas kerjasama antara Institut Teknologi Bandung (ITB), Universitas Kristen Petra (UK Petra), dan Institut Teknologi Sepuluh Nopember (ITS).

IndoBERT terdiri dari 12 layers encoder dan lebih dari 100 juta parameter. IndoBERT dilatih dengan menggunakan korpus Bahasa Indonesia yang luas, termasuk data dari Twitter, Wikipedia, serta sumber-sumber online lainnya. Pada proses pelatihannya model IndoBERT menggunakan teknik transfer learning, dimana model BERT yang sudah dilatih dengan korpus Bahasa Inggris (BERT-base) digunakan sebagai fondasi dasar untuk melatih IndoBERT dengan korpus Bahasa Indonesia (Koto et al., 2020).

Dalam training menggunakan transfer learning, model IndoBERT melalui dua tahapan. Tahapan awal melibatkan model BERT-base Bahasa Inggris, di mana parameter model dimodifikasi agar mampu memproses dan memahami teks dalam Bahasa Indonesia. Tahapan berikutnya terdiri dari fine-tuning pada korpus Bahasa Indonesia, dimana model disesuaikan untuk tugas spesifik seperti analisis sentimen atau identifikasi teks.

IndoBERT adalah model yang optimal dalam melakukan berbagai macam tugas natural language processing seperti analisis sentimen, identifikasi topik, peringkasan teks, dan lain-lain. Kinerja IndoBERT ini dapat ditingkatkan dengan melakukan fine-tuning menggunakan data khusus untuk tugas spesifik, sehingga hasilnya menjadi lebih akurat.

2.5. Ekstensi Browser

Ekstensi browser ataupun disebut juga dengan add-on dan plug-in ialah modul perangkat lunak pada browser yang memiliki tujuan untuk memperluas fungsionalitas yang pada sebuah browser. Banyak browser yang telah mendukung pengembangan ekstensi pada aplikasinya seperti Mozilla Firefox, Google Chrome, dan Microsoft Edge. Pada proses pengembangan ekstensi diperlukan beberapa tools seperti HTML, CSS, dan juga JavaScript (Wikipedia, 2024).

2.6. Modelbit

Modelbit adalah platform yang berfungsi untuk melakukan deployment terhadap model machine learning melalui Git dan notebook python menjadi REST API yang dapat digunakan (Modelbit). Adapun fungsi-fungsi pada Modelbit sebagai berikut:

- Memudahkan *deployment* dengan otomasi
- Membuat *deployment* lebih aman
- Mengintegrasikan CI/CD untuk pipeline *machine learning*
- Pengawasan API yang dapat diandalkan

2.7. Penelitian Terdahulu

Penelitian terdahulu terkait dengan identifikasi *logical fallacy* di Indonesia belum pernah dilakukan, tetapi penelitian serupa telah dilakukan dalam bahasa Inggris. Salah satu penelitian terkait pernah dilakukan oleh Vorakitphan et al. (2022) berjudul *Protect: A Pipeline for Propaganda Detection and Classification* yang mengusulkan sistem untuk mengidentifikasi dan mengklasifikasikan propaganda secara otomatis dengan memanfaatkan fitur semantik dan argumentatif. Penelitian ini mendapatkan skor F1 sebesar 72% menggunakan BERT dalam tahap *binary classification*, dan skor mikro F1 sebesar 64% dalam tugas pengklasifikasian *fallacy* menggunakan arsitektur berbasis RoBERTa.

Sebelumnya penelitian mengenai klasifikasi pernah dilakukan oleh Dharmawan et al. (2023) dengan judul ‘Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (IndoBERT)’ yang menerapkan IndoBERT untuk klasifikasi menggunakan dataset dari sosial media Twitter dan Youtube. Pada penelitian ini terdapat kelas “Normal” dan kelas “Ujaran Kebencian”. Hasil dari penelitian menunjukkan bahwa finetuning dengan learning rate $3e-5$ dan batch sebanyak 32 mendapatkan tingkat akurasi sebesar 89,52%.

Penelitian lainnya dilakukan oleh Jin et al. (2022) dengan judul ‘*Logical Fallacy Detection*’ mengusulkan arsitektur berbasis klasifier sederhana dengan menggabungkan informasi struktural dari *fallacy* untuk mengidentifikasi *logical fallacy*. Penelitian ini mencapai skor F1 sebesar 58,77% dalam mengklasifikasikan 13

tipe *logical fallacy*. Dataset yang digunakan dalam penelitian ini terdiri dari bahan pengajaran online yang dirancang untuk menguji pemahaman siswa tentang *logical fallacy*.

Penelitian yang dilakukan oleh Sahai et al. (2021) melakukan identifikasi *logical fallacy* pada tipe *logical fallacy* seperti *Appeal to authority*, *Appeal to majority*, *Appeal to nature*, *Appeal to tradition*, *Appeal to worse problems*, *Black or white*, *Hasty generalization*, dan *Slippery Slope* terhadap dataset yang dikumpulkan dari Reddit. Penelitian ini menggunakan MGN (*Multi-Granularity Network*) dan BERT dan mendapatkan skor F1 sebesar 67,41%.

Penelitian selanjutnya dilakukan oleh Ruiz-Dolz et al. (2023) dengan judul *Detecting Argumentative Fallacies in the Wild: Problems and Limitations of Large Language Models*. Penelitian ini mendeteksi dan mengklasifikasikan *logical fallacy* ke dalam 5 kelas yaitu *Appeal to Authority*, *Appeal to Majority*, *Ad Hominem*, *Slippery Slope*, dan *Neither*. Penelitian ini menggunakan beberapa macam algoritma dan model seperti *eSVM*, RoBERTa, GPT-3.5 TURBO, dan GPT-4. Dari evaluasi yang dilakukan hasil yang didapatkan ialah dari semua algoritma dan model lainnya RoBERTa mendapatkan skor Macro-F1 tertinggi sebesar 79.6%, skor *precision* sebesar 79.8%, dan skor *Recall* sebesar 79.6% dalam mendeteksi *logical fallacy*, sementara dalam mengklasifikasikan tipe *logical fallacy* RoBERTa mendapatkan skor Macro-F1 sebesar 76.2%, Recall 78%, dan *precision* 75.4%

Adapun penelitian terkait lainnya dilakukan oleh Goffredo et al. (2023) dengan judul *Argument- based Detection and Classification of Fallacies in Political Debates*. Penelitian ini menggunakan dataset yang terdiri dari debat politik pada kampanye presiden Amerika Serikat dari tahun 1960 sampai 2016 dan mengklasifikasikan dataset tersebut menjadi 6 tipe *logical fallacy* seperti *Ad Hominem*, *Appeal to Authority*, *Appeal to Emotion*, *False Cause*, *Slippery Slope*, dan *Slogans*. Penelitian ini menggunakan MultiFusion BERT, sebuah arsitektur berbasis transformer yang mengkombinasikan teks debat, fitur argumentative, serta mengidentifikasi dan mengklasifikasikan *logical fallacy*. Arsitektur yang digunakan dalam penelitian ini mendapatkan rata rata skor F1 sebesar 74%.

Penelitian selanjutnya mengenai identifikasi *logical fallacy* dilakukan secara Anonim (2024) dengan judul *CoCoLoFa: News Comment Sections with Common Logical Fallacies*. Penelitian ini menggunakan dataset yang dikumpulkan dari kolom

komentar pada situs berita, total dataset yang digunakan sebanyak 5772 komentar yang didapatkan dari 647 artikel berita. Terdapat beberapa tipe *logical fallacy* yang diidentifikasi pada penelitian ini adalah *Appeal to Authority*, *Appeal to Majority*, *Appeal to Nature*, *Appeal to Tradition*, *Appeal to Worse Problems*, *False Dilemma*, *Hasty Generalization*, dan *Slippery Slope*. Penelitian ini menggunakan BERT, NLI, dan juga RoBERTa. Hasil dari penelitian ini mendapatkan skor F1 sebesar 0.65% dalam mendeteksi *logical fallacy* dan persetujuan κ Cohen di antara para ahli sebesar 0.6-0.8 untuk semua kelas dengan rata-rata 0.65.

Tabel 2. 1. Penelitian Terdahulu

No	Peneliti/ Tahun	Judul	Metode	Keterangan
1	Situmorang et al., (2024)	Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan Teknik NLP dan Model IndoBERT	NLP, IndoBERT	Penelitian ini melakukan identifikasi menggunakan IndoBERT terhadap tweet yang mengandung potensi depresi. Dataset pada penelitian ini terbagi menjadi "Normal" dan "Berpotensi Depresi". Hasil dari penelitian menunjukkan <i>finetuning</i> dengan <i>learning rate</i> 2e-5, <i>batch</i> sebesar 128, dan epoch sebanyak 5 mendapatkan tingkat akurasi sebesar 94,91%.

Tabel 2. 1. Penelitian Terdahulu (Lanjutan)

No	Peneliti/ Tahun	Judul	Metode	Keterangan
2	Vorakitphan et al., (2022)	<i>Protect: A Pipeline for Propaganda Detection and Classification</i>	BERT dan RoBERTa	<p>Penelitian ini mengusulkan sistem pendeteksi untuk mengidentifikasi dan mengklasifikasi propaganda. Tahap <i>binary classification</i> untuk mendeteksi propaganda mendapatkan skor F1 72% menggunakan BERT dan skor mikro F1 sebesar 64% dalam mengklasifikasikan jenis propaganda menggunakan arsitektur berbasis RoBERTa.</p> <p>Menggunakan arsitektur berbasis klasifier yang menggabungkan informasi struktural dari sebuah <i>fallacy</i>. Ditambahkan model bahasa standar yang telah dilatih untuk mengidentifikasi dengan template yang mencakup asumsi dan premis yang dimodifikasi berdasarkan skema masking yang didefinisikan. Mengidentifikasi 13 tipe <i>logical fallacy</i> dan mendapatkan skor F1 sebesar 58,77%.</p>
3	Jin et al., (2023)	<i>Logical Fallacy Detection</i>	<i>Structure- aware classifier</i>	

Tabel 2. 1. Penelitian Terdahulu (Lanjutan)

No	Peneliti/ Tahun	Judul	Metode	Keterangan
4	Sahai et al., (2022)	<i>Breaking Down the Invisible Wall of Informal Fallacies in Online Discussions</i>	BERT	Mengidentifikasi dan mengklasifikasikan 5 tipe <i>logical fallacy</i> terhadap dataset yang dikumpulkan dari Reddit. Penelitian ini menggunakan MGN (<i>Multi-Granularity Network</i>) dan BERT dan mendapatkan skor F1 sebesar 67,41%.
5	Ruiz-Dolz et al., (2023)	<i>Detecting Argumentative Fallacies in the Wild: Problems and Limitations of Large Language Models</i>	RoBERTa, eSVM, GPT-3.5-TURBO, GPT-4	Mengidentifikasi dan mengklasifikasikan <i>logical fallacy</i> ke dalam 5 kelas yaitu <i>Appeal to Authority</i> , <i>Appeal to Majority</i> , <i>Slippery Slope</i> , <i>Ad Hominem</i> , dan <i>None</i> . Penelitian ini menggunakan algoritma dan model RoBERTa, eSVM, GPT-3.5-TURBO, GPT-4. Hasil evaluasi RoBERTa mendapatkan Macro-F1 tertinggi 79.6%, skor precision sebesar 79.8%, dan skor Recall sebesar 79.6% dalam mendeteksi <i>logical fallacy</i> , dalam mengklasifikasikan tipe <i>logical fallacy</i> RoBERTa mendapatkan skor Macro-F1 sebesar 76.2%, Recall 78%, dan precision 75.4%.

Tabel 2. 1. Penelitian Terdahulu (Lanjutan)

No	Peneliti/ Tahun	Judul	Metode	Keterangan
6	Gofredo et al., (2023)	<i>Argument-based Detection and Classification of Fallacies in Political Debates</i>	MultiFusion BERT	Penelitian ini menggunakan dataset debat politik pada kampanye presiden Amerika Serikat dan mengidentifikasi 6 tipe <i>logical fallacy</i> yang ada pada dataset. Penelitian ini menggunakan pendekatan MultiFusion BERT dan mendapatkan skor F1 sebesar 74%.
7.	Anonim (2024)	<i>CoCoLoFa: News Comment Sections with Common Logical Fallacies</i>	BERT, NLI, dan RoBERTa	Menggunakan 5772 dataset yang dikumpulkan dari kolom komentar situs berita. Mengidentifikasikan 9 tipe <i>logical fallacy</i> . Penelitian menggunakan BERT, NLI, dan RoBERTa. Penelitian ini mendapatkan skor F1 sebesar 65% dalam mengidentifikasi <i>fallacy</i> .

Adapun beberapa perbedaan penelitian yang akan dilakukan oleh penulis dengan penelitian-penelitian sebelumnya yaitu penulis akan menggunakan dataset Bahasa Indonesia dan pendekatan yang digunakan oleh penulis menggunakan machine learning dengan IndoBERT, dataset yang akan digunakan sebanyak lebih dari 14000, dengan ± 2000 data di antaranya adalah dataset yang dikumpulkan dan secara manual dilabel, selain itu model yang dilatih akan diimplementasikan dalam bentuk ekstensi Google Chrome yang dapat digunakan ketika pengguna menggunakan sosial media Instagram ataupun Twitter.

BAB III

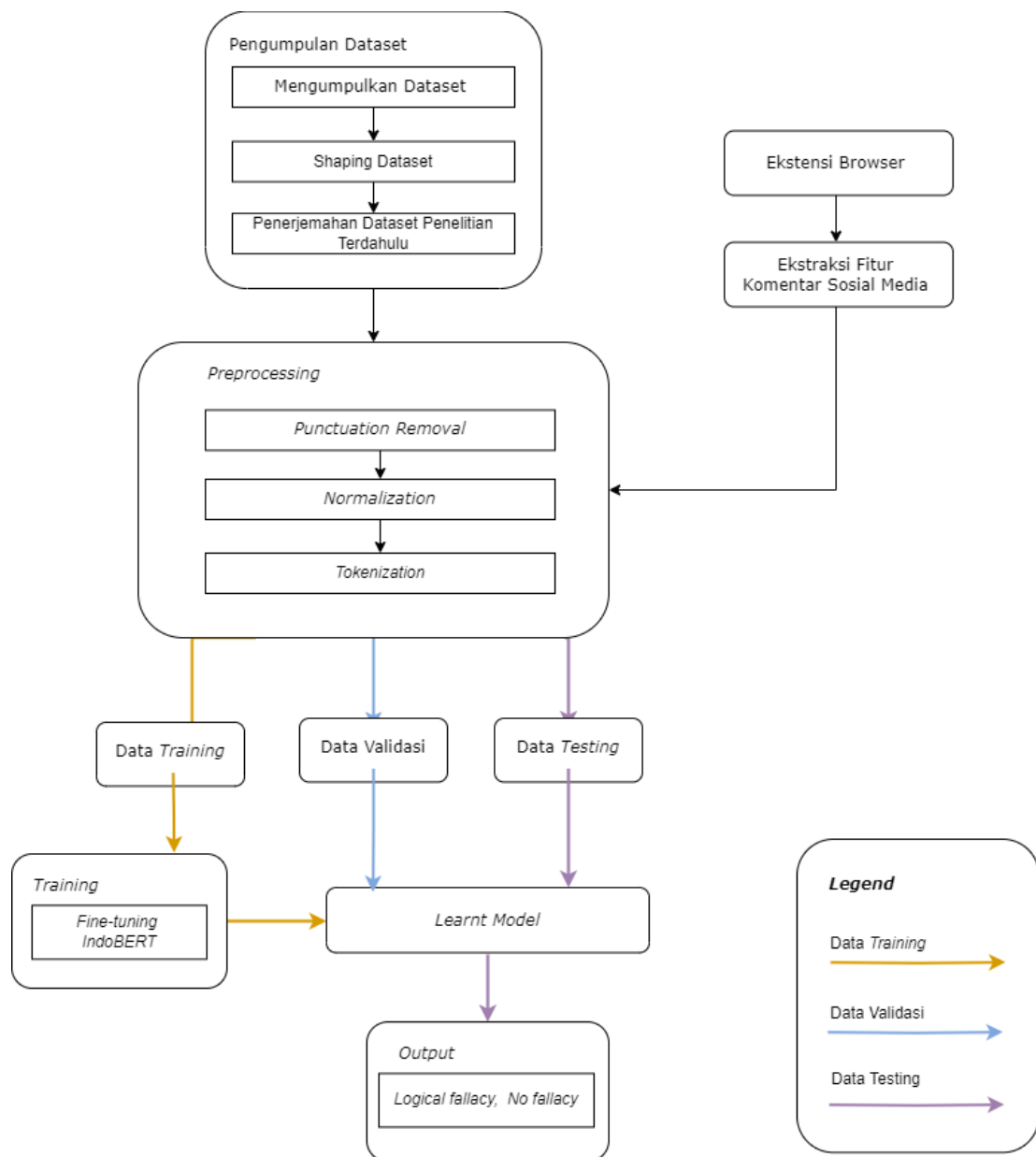
ANALISIS DAN PERANCANGAN

3.1. Arsitektur Umum

Penelitian ini terdiri dari beberapa tahap, tahap awal akan dimulai dengan mengumpulkan dataset yang dijadikan sumber terbuka oleh penelitian-penelitian terkait *logical fallacy* yang telah dilakukan sebelumnya, selain itu akan dikumpulkan juga dataset yang terdiri dari komentar sosial media yang diambil dari berbagai macam platform oleh penulis. Dataset yang telah dikumpulkan kemudian akan melalui proses *shaping dataset* untuk merapikan struktur dataset agar lebih konsisten dan seragam, setelah itu dataset yang didapatkan dari penelitian terdahulu akan diterjemahkan terlebih dahulu menggunakan Google Translate API. Lalu, dataset akan melalui tahap *preprocessing* yang terdiri dari berbagai tahap seperti tahap *punctuation removal* untuk membersihkan dataset dari *noise* yang ada, tahap *normalization* untuk mengubah kata-kata pada dataset yang tidak terstruktur ataupun disingkat menjadi bentuk aslinya, dan juga tahap *tokenization* di mana kalimat akan dipecah menjadi token, token-token ini lalu akan dipetakan berdasarkan index pada kosakata *tokenizer*. Setelah melewati tahap *preprocessing* dataset dibagi menjadi data *training* dan data validasi dengan pembagian 80% untuk *data training* dan 20% untuk data validasi. Kemudian, data *training* akan melakukan tahap pelatihan dengan melakukan *fine-tuning* dari *pre-trained* model IndoBERT, sementara itu data validasi akan memvalidasi kinerja model selama proses pelatihan.

Sedangkan untuk data testing dilakukan dengan menggunakan ekstensi browser yang akan dikembangkan. Ekstensi browser yang diaktifkan ketika membuka situs sosial media akan melakukan ekstraksi fitur untuk mendapatkan komentar yang ada pada suatu post di sosial media ketika cursor diarahkan pada komentar. Komentar yang didapatkan kemudian akan dilanjutkan ke tahap *preprocessing* untuk membersihkan *noise* yang ada. Setelah itu, komentar yang telah di-preprocess akan diproses oleh *learnt* model untuk diidentifikasi potensi *logical fallacy*-nya. Ketika

proses identifikasi berhasil dilakukan, hasil yang ditampilkan yaitu Logical fallacy apabila terdapat potensi logical fallacy pada komentar dan No fallacy apabila potensi logical fallacy tidak diidentifikasi oleh model. Arsitektur umum dari proses yang telah dijabarkan dapat dilihat pada Gambar 3. 1.



Gambar 3. 1. Arsitektur Umum

3.2. Pengumpulan Dataset

Dataset yang akan digunakan dalam penelitian ini berupa file excel dengan total lebih dari 14000 data berbahasa Indonesia yang didapatkan dari berbagai macam sumber. Selain itu, penulis juga menambahkan dataset pribadi dalam bentuk komentar beserta dengan labelnya yang telah dikumpulkan dari sosial media seperti Instagram, Reddit, dan Twitter. Penulis melakukan *scraping* untuk mendapatkan dataset dari Instagram dari berbagai macam akun. Selain itu, penulis juga mengambil dataset Reddit dalam bentuk NDJson dari file *torrent* yang telah diekstrak dengan post, komentar, dan informasi lainnya dari berbagai macam subreddit yang dikumpulkan oleh pengguna bernama Watchful dan Raiderbdev (2024), penulis hanya mengambil data dari subreddit r/Indonesia karena lebih mencerminkan perilaku warga Indonesia dalam bersosial media. Dataset ini lalu difilter terlebih dahulu untuk mencari data komentar yang mengandung kata kunci seperti ‘ad hominem’, ‘generalisasi’, ‘hasty generalization’, ‘ad populum’, ‘false dilemma’, ‘black or white fallacy’, dan ‘appeal to authority’, kemudian data komentar yang mengandung kata kunci tersebut beserta dengan *parent* komentarnya akan disimpan ke bentuk excel untuk dilabel secara manual oleh penulis dan menghasilkan sebanyak ± 2000 data

1	Upvote	Date	Author	Comment	Parent Comment ID	Parent Comment	Link
2	3	2017-10-1	u/km0bandokf84p	Mungkin gw mikirnya dokdu1s		yep kita memang gitu haha	
3	-1	2017-10-2	u/kellhusi dooxq8	Oi, argumen mas, jang doox7xi		Tuntutan mahasiswa tolol m	https://www.reddit.com/r/indonesia/comments/777tgw/can_you_be_racist_if_youre_a
4	0	2017-10-2	u/namaku dopabs5	Halah. Ini internet, bu dooxq8		Oi, argumen mas, jangan ad	https://www.reddit.com/r/indonesia/comments/77sury/demo_jokowijk_hingga_tengah
5	1	2017-10-2	u/kekekm doznvpb	Generalisasinya janga doycpwd		terlihat stupid kampusnya. S	https://www.reddit.com/r/indonesia/comments/79185d/kuliah_nggak_lulus_young_lex
6	3	2017-10-2	u/honeyb dp09nfu	Emang bener ad some dp032mu		emang bener sih	https://www.reddit.com/r/indonesia/comments/7939qp/aksara_jawa/dp09nfu/
7	4	2017-10-2	u/how11 dp0e9ao	What he said may see dp09nfu		Emang bener ad some truth.	https://www.reddit.com/r/indonesia/comments/7939qp/aksara_jawa/dp0e9ao/
8	-5	2017-10-3	u/Emman dp4xhy5	nggak pernah nyetir d dp4xfz3		[removed]	https://www.reddit.com/r/indonesia/comments/79tze4/nggak_terima_disalip_berantem
9	1	2017-10-3	u/somebcdp5a9mh	Berujung ga mutu kari dp57y4d		Aneh sekali. Bagaimana cara	https://www.reddit.com/r/indonesia/comments/79m8zx/pemprov_dki_tidak_akan_perp

Gambar 3. 2. Dataset Sosial Media yang Telah Dikumpulkan

Hasil dataset yang diambil dari Reddit dataset dapat dilihat pada Gambar 3.2 dan pada Tabel 3. 1 dapat dilihat list dataset yang dikumpulkan dari berbagai sumber.

Tabel 3. 1. List Dataset Berdasarkan Sumber

Sumber Dataset	Jumlah Dataset Yang Diambil
Paper Breaking Down the Invisible Wall of Informal Fallacies in Online Discussions (Sahai et al., 2021)	6756
Paper Logical Fallacy Detection (Jin et al., 2022)	2164

Tabel 3. 1. List Dataset Berdasarkan Sumber (Lanjutan)

Sumber Dataset	Jumlah Dataset Yang Diambil
Paper Detecting Argumentative Fallacies in the Wild: Problems and Limitations of Large Language Models (Ruiz-Dolz et al., 2023)	1223
Paper Argotario: Computational Argumentation Meets Serious Games (Habernal et al., 2017)	1032
Paper Beyond The Imitation Game: Quantifying And Extrapolating The Capabilities Of Language Models (Srivastava et al., 2022)	1012
Scraping Dataset Reddit	1210
Scraping Dataset Instagram	794
Scraping Dataset Twitter	507
Total:	14701

Sebelum data diproses lebih lanjut, terdapat beberapa tahapan yang akan dilakukan untuk mempersiapkan dataset terlebih dahulu. Adapun proses dalam tahap ini yaitu: shaping dataset, cleaning dataset, penerjemahan dataset, dan labeling dataset.

3.2.1. *Shaping* Dataset

Tahap shaping dataset dilakukan untuk memastikan dataset memiliki struktur yang rapi, seragam, dan memuat data yang relevan. Karena dataset yang digunakan diambil dari banyak penelitian yang berbeda maka struktur dataset dari tiap penelitian juga berbeda, karena itu perlu dilakukan shaping dataset agar semua dataset memiliki struktur yang sama.

Shaping dataset akan dilakukan menggunakan `shape_dataframe` yang disediakan oleh *library* pandas untuk membuat dan menyimpan DataFrame yang baru dari data yang dimasukkan ke dalam format excel. Kolom yang disimpan hanya kolom yang relevan seperti kolom `text`, `is_fallacy`, `fallacy_type`, dan `dataset_source`. *Shaping* dataset ini dilakukan berulang pada setiap dataset penelitian terdahulu yang dikumpulkan, lalu disimpan dan disatukan ke dalam satu file yang sama. Contoh salah satu struktur asli dataset dapat dilihat pada Gambar 3. 3 dan hasil dari *shaping* dataset pada Gambar 3. 4.

Unnamed: 0	comment_id	title	comment	parent	grandparent	child	fallacy_exists	fallacy_highlighted	fallacy_highlighted_indices	claim	
473	473	ecpfjja	Having to reinterpret your religion to keep up...	An interesting version of an argument from aut...	If your understanding of religion is as sophom...	NaN	NaN	0	NaN	Religious words are set in stone.	
474	474	egwneen	CMV: There are no "human races" anymore and we...	I call your argument a fallacy (appeal to auth...	My biology textbook (tenth grade) says that ra...	There is a biological difference, and it's way...	NaN	1	My biology textbook (tenth grade) says that ra...	[(0, 2), (3, 10), (11, 19), (20, 21), (21, 26)]	There are no separate races.
475	475	ewh2tkp	What screams "I am very smart"?	That in fact, is an appeal to authority logica...	Using your IQ as an argument to validate your ...	NaN	NaN	1	I have an IQ of 130. That's in the top 2% of h...	[(127, 128), (129, 133), (134, 136), (137, 139)]	Smart people don't have to mention that they a...
476	476	f0mjrtb	U.S. official says drone and missile attack on...	Implication being soon to be "former"? Article...	That's "acting" U.S. official.	"U.S. official" has lost all credibility to me...	NaN	0	NaN	NaN	That they are acting U.S. official
477	477	f58vum2	Being A Victim Of Something Does Not Make You ...	Its a logical fallacy called an argument from ...	Conversely, being an expert on something doesn...	NaN	NaN	0	NaN	NaN	That being an expert on something doesn't mean

Gambar 3. 3. Struktur Asli Dataset

	text	is_fallacy	fallacy_type	dataset_source
0	The green party in Germany has the opinion, th...	Fallacy	Appeal to authority	Argotario Dataset
1	Yes, whoever drinks and drives has a bad reac...	No fallacy	No fallacy	Argotario Dataset
2	the natural habitat of the guerrilla is in jeo...	No fallacy	No fallacy	Argotario Dataset
3	Computers are bad for the development of child...	No fallacy	No fallacy	Argotario Dataset
4	The global taskforce for human rights is doing...	Fallacy	Hasty generalization	Argotario Dataset

Gambar 3. 4. Struktur Asli Setelah Melalui Proses *Shaping*

3.2.2. Penerjemahan Dataset

Beberapa dataset Reddit dan dataset penelitian-penelitian terdahulu yang dikumpulkan menggunakan Bahasa Inggris, dataset ini kemudian akan diterjemahkan ke dalam Bahasa Indonesia terlebih dahulu menggunakan API dari Google bernama Googletrans agar dataset dapat digunakan dalam proses *training*. Dataset yang telah diterjemahkan akan disimpan ke kolom 'text' menggantikan kolom 'text' yang lama. Gambar 3. 5 menampilkan dataset yang belum diterjemahkan dan Gambar 3. 6 menampilkan dataset yang telah diterjemahkan.

	text	is_fallacy	fallacy_type	dataset_source
0	The green party in Germany has the opinion, th...	Fallacy	Appeal to authority	Argotario Dataset
1	Yes, whoever drinks and drives has a bad reac...	No fallacy	No fallacy	Argotario Dataset
2	the natural habitat of the guerrilla is in jeo...	No fallacy	No fallacy	Argotario Dataset
3	Computers are bad for the development of child...	No fallacy	No fallacy	Argotario Dataset
4	The global taskforce for human rights is doing...	Fallacy	Hasty generalization	Argotario Dataset

Gambar 3. 5. Dataset yang Belum Diterjemahkan

	text	is_fallacy	fallacy_type	dataset_source
0	Partai hijau di Jerman berpendapat, reaktor nu...	Fallacy	Appeal to authority	Argotario Dataset
1	Ya, siapa pun yang minum dan mengemudi memilik...	No fallacy	No fallacy	Argotario Dataset
2	habitat alami para gorila terancam punah. Kebu...	No fallacy	No fallacy	Argotario Dataset
3	Komputer berdampak buruk bagi perkembangan ota...	No fallacy	No fallacy	Argotario Dataset
4	Gugus tugas global untuk hak asasi manusia tel...	Fallacy	Hasty generalization	Argotario Dataset

Gambar 3. 6. Dataset yang Telah Diterjemahkan

3.2.3. Pelabelan Dataset

Pada dataset yang didapatkan dari *scraping* komentar sosial media perlu dilakukan pelabelan terlebih dahulu sebelum digabungkan dengan dataset penelitian terdahulu. Dataset yang dikumpulkan melalui proses *scraping* ini terdiri data komentar yang termasuk dalam 5 tipe *logical fallacy* (*ad hominem*, *ad populum*, *appeal to authority*, *false dilemma*, *hasty generalization*) beserta juga dengan komentar yang tidak termasuk *logical fallacy*.

Sama seperti dengan dataset pada penelitian terdahulu, dataset yang akan dilabel akan terdiri dari beberapa kolom yaitu kolom ‘text’ yang berisi komentar. Kolom ‘is_fallacy’ di mana pada kolom ini pelabelan dilakukan, data komentar yang mengandung potensi *logical fallacy* akan dilabel dengan Fallacy dan data komentar yang tidak memiliki potensi *logical fallacy* akan dilabel dengan No fallacy. Pada kolom ‘fallacy_type’ apabila tidak ada potensi *logical fallacy* pada data komentar maka akan dilabel sebagai No fallacy dan apabila data komentar mengandung potensi *logical fallacy* maka akan dilabel sesuai dengan tipe *logical fallacy*-nya.

Proses pelabelan menjadi tahap yang cukup penting untuk mendapatkan hasil yang lebih akurat dengan menyiapkan dataset yang telah dilabel dengan teliti. Dengan adanya ketentuan label yang jelas pada proses pelabelan, diharapkan identifikasi potensi *logical fallacy* pada korpus berbahasa Indonesia bisa dilakukan dengan lebih optimal dan mendapatkan hasil yang lebih akurat. Adapun ketentuan pelabelan sebagai berikut:

1. Data yang tidak mengandung *logical fallacy* akan dilabel dengan *No fallacy* pada kolom ‘is_fallacy’ dan ‘fallacy_type’
2. Data yang mengandung potensi *logical fallacy Ad Hominem* akan diberi ‘Ad hominem’ pada kolom ‘fallacy_type’ dan label ‘Fallacy’ pada kolom ‘is_fallacy’.
3. Data yang mengandung potensi *logical fallacy Ad Populum* akan diberi label ‘Ad populum’ pada kolom ‘fallacy_type’ dan label ‘Fallacy’ pada kolom ‘is_fallacy’.

4. Data yang mengandung potensi *logical fallacy Hasty Generalization* akan diberi label 'Hasty generalization' pada kolom '*fallacy_type*' dan label 'Fallacy' pada kolom '*is_fallacy*'.
5. Data yang mengandung potensi *logical fallacy Appeal to Authority* akan diberi label 'Appeal to authority' pada kolom '*fallacy_type*' dan label 'Fallacy' pada kolom '*is_fallacy*'.
6. Data yang mengandung potensi *logical fallacy False Dilemma* akan diberi label 'False dilemma' pada kolom '*fallacy_type*' dan label 'Fallacy' pada kolom '*is_fallacy*'.

Dataset sosial media yang telah dilabel akan digabungkan dengan dataset dari penelitian-penelitian sebelumnya, dataset tersebut akan dibagi 80% menjadi dataset training dan 20% menjadi dataset validasi. Dataset *training* akan terdiri dari beberapa kolom seperti text sebagai kolom berisi teks komentar, kolom *is_fallacy* sebagai keterangan apakah komentar mengandung potensi *logical fallacy* atau tidak, *fallacy_type* sebagai keterangan dari jenis *logical fallacy* yang ada, dan *dataset_source* untuk memperjelas sumber dataset yang digunakan. Adapun contoh data yang digunakan pada dataset training yang dapat dilihat pada Tabel 3. 2.

Tabel 3. 2. Dataset *Training*

<i>text</i>	<i>is_fallacy</i>	<i>fallacy_type</i>	<i>dataset_source</i>
Entah semua orang bisa, atau tidak ada yang bisa.	Fallacy	False dilemma	Breaking Down the Invisible Wall of Informal Fallacies Paper
Kakak ipar gue yang berprofesi sebagai guru mengatakan bahwa sekolah ini bukanlah tempat yang gue inginkan untuk menyekolahkan anak gue.	Fallacy	Appeal to authority	Logical Fallacy Detection Paper

Tabel 3. 2. Dataset *Training* (Lanjutan)

<i>text</i>	<i>is_fallacy</i>	<i>fallacy_type</i>	<i>dataset_source</i>
hah.. argumen bocah. Gak usah ngepost kalau mental bocah. siapa juga yg larang ² situ?? yang mulau juga siapa. Bocah amat lu anjing.	Fallacy	Ad hominem	Scraped Social Media Dataset
Saya mengambil kesimpulan karena kamu sebagai umat beragama tidak dapat memberikan bukti atas klaim mereka.	Fallacy	Hasty generalization	Scraped Social Media Dataset
Ya, siapa pun yang minum dan mengemudi memiliki waktu reaksi yang buruk. Orang bisa mati.	No fallacy	No fallacy	Argotario Dataset

3.3. Preprocessing Dataset

Preprocessing merupakan proses yang dilakukan untuk membersihkan dataset dari *noise* yang tidak diperlukan pada tahap *training* sehingga membuat dataset menjadi lebih bersih dan mengurangi beban komputasi. Tetapi, berdasarkan penelitian yang dilakukan oleh Alzahrani et al. (2021), BERT memberikan hasil akurasi tertinggi ketika *preprocessing* tidak dilakukan dan memberikan akurasi 1% lebih rendah ketika *stopword* pada dataset dibuang, sehingga proses *preprocessing* penelitian ini hanya akan meliputi *punctuation removal*, *normalization* untuk menormalisasi kata singkatan yang sering kali muncul di sosial media, dan *tokenization* agar mendapatkan akurasi setinggi mungkin.

3.3.1. Punctuation Removal

Punctuation Removal adalah tahap yang dilakukan untuk menghapus tanda baca dari teks yang tidak diperlukan pada proses *training* untuk mengurangi *noise*, adapun tanda baca yang dihapus pada penelitian ini adalah tanda *hyperlink*, *quote*, *emote*, dan *tag* yang mungkin ada pada dataset yang dikumpulkan dari sosial media. Tabel 3. 3 menampilkan

pseudocode punctuation removal beserta contoh dari *punctuation removal* dapat dilihat pada Tabel 3. 4

Tabel 3. 3. Pseudocode Punctuation Removal

<i>Punctuation Removal</i>
<pre> IMPORT emoji IMPORT regex DEFINE FUNCTION clean_text(text: String, newline=Boolean, quote=Boolean, bullet_point= Boolean, link=Boolean, spoiler=Boolean, emote=Boolean, tag=Boolean): IF newline: text = regex(r'\n+', ' ', text) text = text.strip() text = regex(r'\s\s+', ' ', text) IF quote: text = regex (r'"\?\\&?gt;?', " ", text) IF bullet_point: text = regex (r'*', " ", text) text = regex('&#x200B;', " ", text) IF link: text = regex(r'http\S+', " ", text) text = regex(r'www\S+', " ", text) IF spoiler: text = regex('&lt;', " ", text) text = regex(r'!(.*?)!', r'\1', text) IF emote: text = emoji.replace_emoji(text, replace="") IF tag: text = regex(r"@S+", "", text) RETURN text.lower() </pre>

Tabel 3. 4. Contoh Implementasi *Punctuation Removal*

Sebelum <i>punctuation removal</i>	Setelah <i>punctuation removal</i>
Waktu dulu di kelas periklanan dosen menyarankan web ini (https://yourlogicalfallacyis.com/), sangat ringkas dan jelas.	Waktu dulu di kelas periklanan dosen menyarankan web ini sangat ringkas dan jelas.

3.3.2. *Normalization*

Normalization adalah tahap yang dilakukan untuk menormalisasi kata-kata pada dataset yang tidak terstruktur ataupun disingkat sehingga dapat meningkatkan kualitas dari dataset yang digunakan dan memperkecil kemungkinan adanya duplikasi data. Penulis menggunakan korpus bahasa gaul (Salsabila et al., 2018) untuk menggantikan kata-kata *slang* yang disingkat atau tidak standar seperti “bgt”, “gpp”, “pgn”, dan lain-lain menjadi bentuk standar dari kata-kata tersebut. Apabila *normalization* tidak dilakukan maka sistem akan menganggap kata-kata bgt”, “gpp”, “pgn” memiliki arti yang berbeda dengan “banget”, “tidak apa-apa”, dan “pengen”, padahal kata-kata tersebut memiliki arti yang sama. Sehingga, *normalization* tetap dilakukan untuk memastikan bahwa dataset yang dilatih menggunakan bentuk standar Bahasa Indonesia sesuai dengan aturan ejaan yang diberlakukan. *Pseudocode normalization* dapat dilihat pada Tabel 3. 5 dan implementasi *normalization* seperti pada Tabel 3. 6.

Tabel 3. 5. *Pseudocode Normalization*

<i>Normalization</i>
<pre> DEFINE FUNCTION normalization(input_text: String, kamus_gaul: FilePath) kamus_bahasa_gaul = OPENFILE(kamus_gaul) slang_dict = {} FOR EACH row IN kamus_bahasa_gaul ADD TO slang_dict: KEY: row['slang'] VALUE: row['formal'] words = SPLIT(input_text, ' ') normalized_words = [] </pre>

```

FOR EACH word IN words
  IF word EXISTS IN slang_dict
    normalized_words.append(slang_dict[word] )
  ELSE
    normalized_words.append(word)

normalized_text = JOIN(normalized_words, '')
RETURN normalized_text

```

Tabel 3. 6. Contoh Implementasi *Normalization*

Sebelum <i>Normalization</i>	Setelah <i>Normalization</i>
kata orang2 brand itu bagus makanya gue beli	kata orang-orang brand itu bagus makanya gue beli

3.3.3. *Tokenization*

Tahap *tokenization* adalah proses untuk mengubah teks mentah dalam bentuk kata menjadi representasi numerik yang bisa diproses oleh model. Tahap ini menggunakan metode *tokenization* khusus yang disebut *WordPiece tokenization* di mana kata akan dipecah menjadi sub-kata, serta adanya penambahan token khusus, *padding*, dan *truncation*, dan juga pembuatan attention mask. *Pseudocode tokenization* dapat dilihat pada Tabel 3. 7 dan contoh implementasi *tokenization* yang telah diubah *id*-nya ke token seperti pada Tabel 3. 8.

Tabel 3. 7. *Pseudocode Tokenization*

<i>Tokenization</i>
<pre> IMPORT Tokenizer DEFINE FUNCTION tokenize_text(text: String, max_len: Integer): encoded_kalimat = Tokenizer.encode(text, tokens: Boolean, max_length = Integer, pad_max_length = Boolean, truncation = Boolean, return_attention_mask = Boolean) </pre>

3.4. Finetuning IndoBERT

Pada penelitian ini, penulis memutuskan untuk menggunakan model IndoBERT sebagai kerangka kerja analisis. Keputusan untuk menggunakan model ini dikarenakan IndoBERT telah dilatih menggunakan lebih dari 220 juta kata Bahasa Indonesia sehingga akurasi model IndoBERT pada penggunaan dataset berbahasa Indonesia lebih tinggi dibandingkan jika menggunakan model BERT lainnya seperti bert-base-uncased. Setelah tahap *preprocessing* dilewati, dataset akan dikonversi ke format vektor representasi kata menggunakan *tokenizer* yang telah disediakan oleh IndoBERT agar dataset dapat diterima dan diolah menggunakan IndoBERT. Tahap selanjutnya yaitu tahap *finetuning* terhadap *hyperparameter*, di mana model IndoBERT akan disesuaikan untuk tugas identifikasi potensi *logical fallacy*, hasil dari *finetuning* ini adalah learnt model yang dapat mengidentifikasi potensi *logical fallacy* pada komentar.

3.5. Evaluasi Model

Tahap ini dilakukan menggunakan data evaluasi agar performa dari model yang telah dilatih dapat diketahui. Pada penelitian ini akan digunakan accuracy metrics sebagai *performance metrics*. Nilai yang dikeluarkan oleh *accuracy metrics* adalah nilai akurasi subset, yang akan mengembalikan nilai 1 apabila data yang diprediksi bernilai sama dengan nilai aktual, dan nilai 0 apabila sebaliknya. Adapun rumus dari menghitung *accuracy* seperti pada persamaan 3. 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \quad (3.1)$$

Di mana:

- TP (*True Positive*) = Jumlah data dengan nilai positif dari kalimat berlabel positif
- TN (*True Negative*) = Jumlah data dengan nilai negative dari kalimat berlabel negatif
- FP (*False Positive*) = Jumlah data dengan nilai negative dari kalimat berlabel positif
- FN (*False Negative*) = Jumlah data dengan nilai positif dari kalimat berlabel negatif


```
'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, device='cuda:0')
```

Setelah kalimat selesai di-*preprocess* kalimat kemudian diuji dengan model yang sebelumnya telah dilatih. Contoh hasil dari *testing* komentar ditulis pada Tabel 3.10.

Tabel 3. 10. Contoh Hasil Testing Model

No	Komentar	Prediction	Actual	Match
1	CEO itu pernah dihukum krn mengemudi dlm keadaan mabuk saat kuliah. Kita hrs meragukan laporan CEO	Fallacy	Fallacy	True

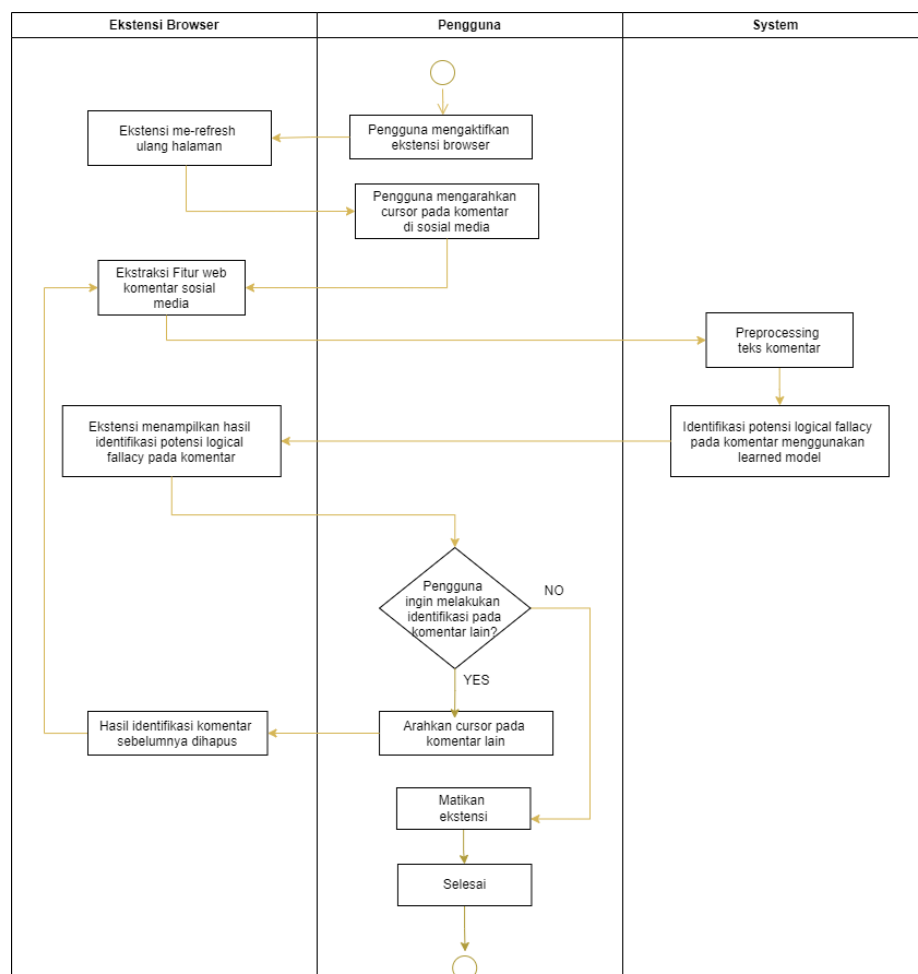
3.7. Pemodelan Sistem

Pada tahap pengujian *testing* data *testing* yang digunakan akan diambil ketika pengguna menggunakan ekstensi pada situs sosial media. Ketika pengguna mengarahkan *cursor* pada komentar di sosial media ekstensi akan melakukan ekstraksi fitur untuk mendapatkan komentar yang dipilih. Komentar ini kemudian akan melalui tahap *preprocessing* terlebih dahulu untuk mengubah kalimat dari kemungkinan adanya kata-kata yang tidak ditulis dalam bentuk standar ke bentuk standar dan mengubah bentuk kalimat ke dalam format yang dapat diterima IndoBERT. Tabel 3.9 menampilkan tahap *preprocessing*.

3.7.1. Diagram Aktivitas

Diagram aktivitas ekstensi menggambarkan aktivitas yang terjadi ketika pengguna berinteraksi dengan ekstensi. Aktivitas dimulai ketika pengguna mengaktifkan ekstensi

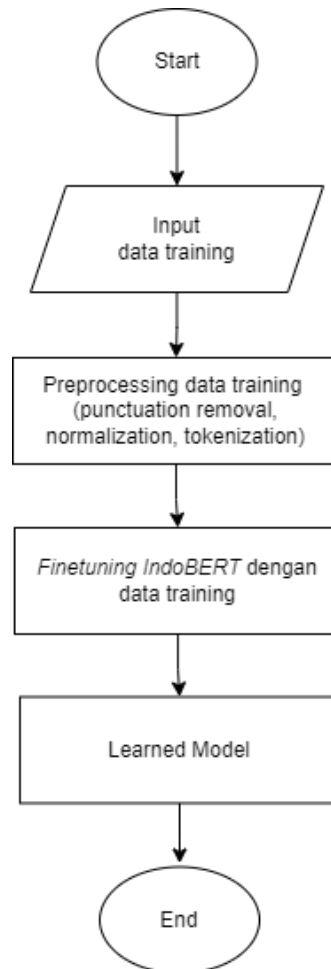
pada situs sosial media Instagram atau Twitter. Setelah diaktifkan maka ekstensi akan *me-refresh* ulang halaman, selanjutnya pengguna dapat mengarahkan cursor pada komentar yang ada di sosial media Instagram atau Twitter. Kemudian, ekstensi browser akan melakukan ekstraksi fitur pada web untuk mendapatkan komentar yang diinginkan pengguna. Setelah komentar didapatkan maka system akan melakukan *preprocessing* untuk membersihkan komentar agar dapat dengan mudah diproses oleh sistem. Lalu, dilakukan proses identifikasi potensi *logical fallacy* pada komentar yang telah di-preprocess dan kemudian akan diproses oleh learned model yang telah dilatih menggunakan IndoBERT. Setelah identifikasi potensi *logical fallacy* berhasil dilakukan maka hasil identifikasi dengan probabilitas tertinggi akan ditampilkan dalam bentuk hover effect oleh ekstensi browser. Lalu, jika pengguna ingin melakukan identifikasi pada komentar lain di sosial media maka identifikasi sebelumnya akan dihapus dan sistem akan melakukan ulang tahapan *preprocessing* dan mendapatkan hasil identifikasi potensi *logical fallacy* pada komentar baru yang dipilih. Jika pengguna tidak ingin melakukan identifikasi pada komentar lagi maka pengguna dapat memilih untuk mematikan ekstensi browser. Gambar 3.7 menjelaskan Diagram Aktivitas ekstensi.



Gambar 3. 7. Diagram Aktivitas Ekstensi

3.7.2. Flowchart Sistem

Flowchart Sistem merupakan alur diagram yang menggambarkan urutan proses yang terjadi pada sistem. Pada bagian ini akan ditampilkan visualisasi flowchart *training* dan flowchart ekstensi browser yang dapat dilihat pada Gambar 3.8 dan Gambar 3.9.

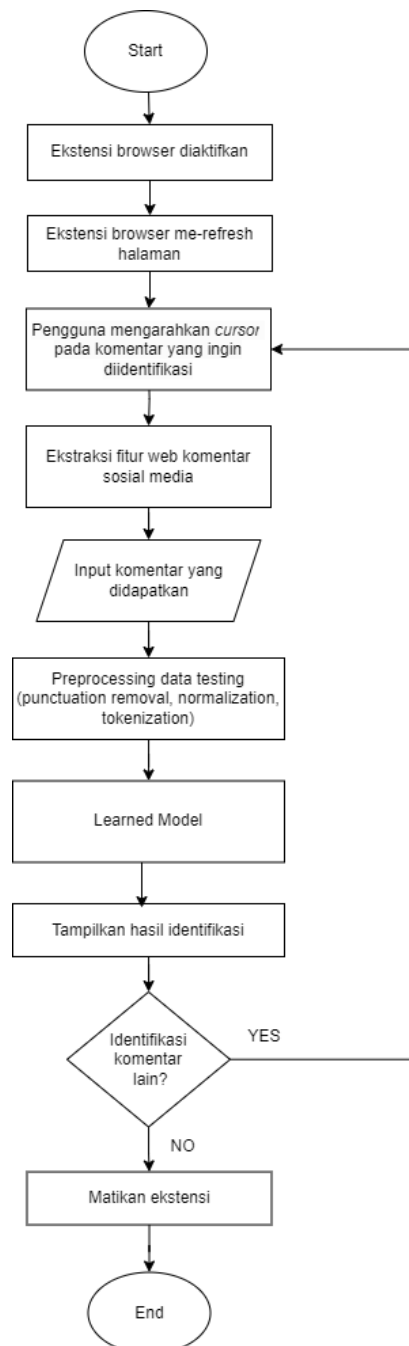


Gambar 3. 8. Flowchart Training

Berikut adalah urutan proses pada sistem yang akan dilakukan pada tahap training:

1. Hal pertama yang dilakukan adalah memasukkan dataset yang akan digunakan dalam proses *training*. Dataset ini terdiri dalam bentuk excel yang terdiri dari kolom 'text', 'is_fallacy', 'fallacy_type', dan juga 'dataset_source'.
2. Dataset yang dimasukkan kemudian akan melalui tahap *preprocessing* seperti *punctuation removal* untuk menghilangkan tanda-tanda baca yang tidak diperlukan, *normalization* untuk mengubah bentuk kata pada dataset yang tidak formal menjadi formal, dan juga *tokenization* untuk mengubah dataset ke dalam bentuk yang dapat diolah oleh IndoBERT.

3. Setelah preprocessing dilakukan, tahap selanjutnya adalah dilakukan training terhadap dataset dengan melakukan finetuning model IndoBERT untuk dipelajari sistem.
4. Hasil model yang telah dilatih kemudian akan disimpan agar dapat digunakan untuk mengidentifikasi potensi logical fallacy komentar di sosial media.



Gambar 3. 9. Flowchart Ekstensi Browser

Berikut adalah alur proses pada flowchart ekstensi browser:

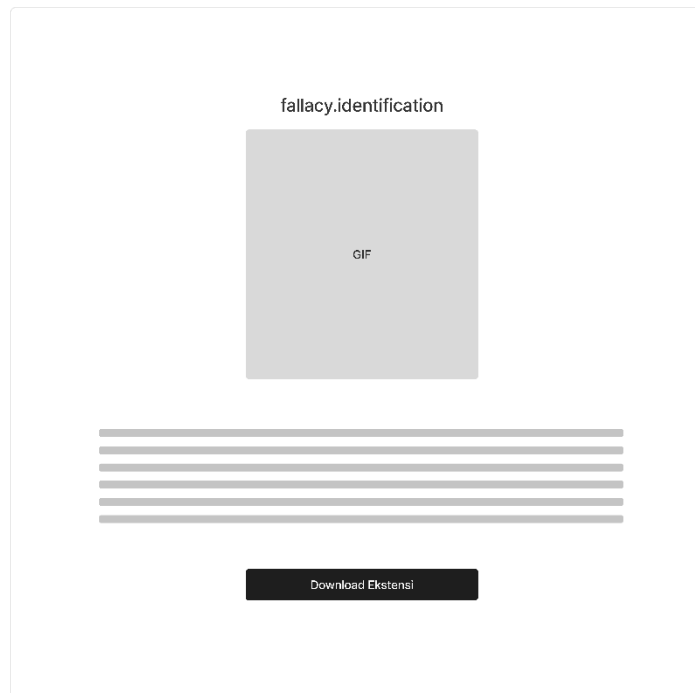
1. Pengguna menjalankan ekstensi dengan mengaktifkan ekstensi browser.
2. Ekstensi browser yang diaktifkan kemudian akan me-*refresh* halaman situs sosial media.
3. Setelah itu pengguna dapat menggunakan ekstensi dengan mengarahkan *cursor* pada komentar yang diinginkan.
4. Selanjutnya, ekstensi akan melakukan ekstraksi fitur untuk mendapatkan komentar yang telah dipilih.
5. Komentar yang telah berhasil diambil selanjutnya akan dimasukkan dan melalui tahap *preprocessing* untuk dibersihkan dari *noise-noise* yang ada.
6. Kemudian, identifikasi potensi *logical fallacy* pada komentar akan diproses oleh learned model yang telah dilatih.
7. Hasil dari identifikasi potensi *logical fallacy* oleh model ditampilkan ekstensi browser dalam bentuk *hover effect*.
8. Pengguna dapat memilih untuk melanjutkan identifikasi atau menghentikan ekstensi dengan mematikannya.

3.8. Pemodelan Sistem

Pada tahap ini akan dijelaskan rancangan antarmuka ekstensi Google Chrome untuk mengidentifikasi potensi *logical fallacy* pada komentar sosial media yang akan dikembangkan. Tujuan dari dibangunnya ekstensi ini adalah untuk memudahkan pengguna dalam menggunakan model.

3.8.1. Rancangan Antarmuka *Landing Page*

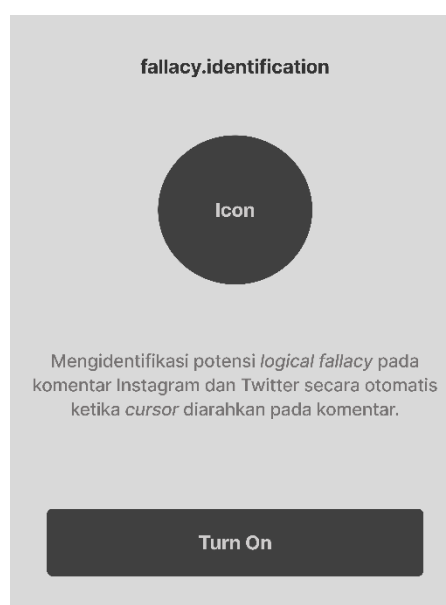
Landing page akan dijadikan sebagai halaman yang akan menampilkan beberapa informasi seperti GIF yang menampilkan cara ekstensi bekerja, tombol *url* untuk mengunduh ekstensi, dan juga penjelasan fungsi dari ekstensi. Gambar 3.10 menampilkan rancangan antarmuka *landing page* ekstensi.



Gambar 3. 10. Rancangan Antarmuka *Landing Page* Ekstensi Browser

3.8.2. Rancangan Antarmuka *Popup* Ekstensi

Tampilan *popup* yang dirancang akan menampilkan beberapa informasi seperti nama ekstensi, ikon ekstensi, penjelasan ekstensi, dan juga tombol *turn on* untuk mengaktifkan ekstensi. Gambar 3.11 menampilkan rancangan antarmuka *popup extension*.



Gambar 3. 11. Rancangan Antarmuka *Popup* Ekstensi Browser

3.8.3. Rancangan Antarmuka Identifikasi *Tweet* dengan Ekstensi

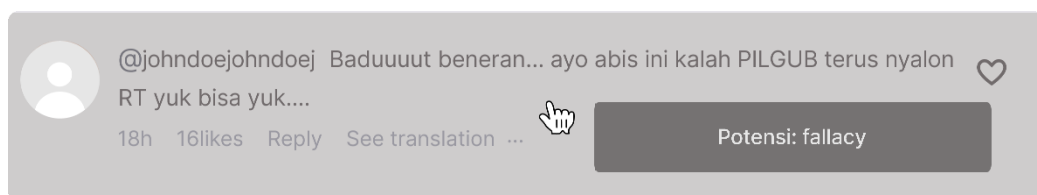
Rancangan ini akan menampilkan hasil identifikasi potensi *logical fallacy* dalam bentuk *hover effect* pada suatu *tweet* ketika pengguna telah mengaktifkan ekstensi dan mengarahkan *cursor* pada sebuah *tweet*, ketika *cursor* sudah tidak diarahkan pada *tweet* maka *hover effect* tersebut akan hilang. Gambar 3.12 menampilkan rancangan antarmuka identifikasi potensi *logical fallacy* pada *tweet*.



Gambar 3. 12. Rancangan Antarmuka Identifikasi Potensi *Logical Fallacy*
pada *Tweet* Menggunakan Ekstensi Browser

3.8.4. Rancangan Antarmuka Identifikasi Komentar Instagram dengan Ekstensi

Identifikasi potensi *logical fallacy* dilakukan dalam bentuk *hover effect* pada sosial media Instagram ketika ekstensi telah diaktifkan dan *cursor* diarahkan pada komentar, hasil identifikasi potensi *logical fallacy* akan hilang ketika *cursor* tidak mengambang di atas komentar. Gambar 3.13 menampilkan rancangan antarmuka identifikasi pada komentar Instagram.



Gambar 3. 13. Rancangan Antarmuka Identifikasi Potensi *Logical Fallacy*
pada Komentar Instagram dengan Ekstensi Browser

BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Implementasi Sistem

Tahap ini menjelaskan perancangan sistem untuk identifikasi potensi logical fallacy pada sosial media menggunakan IndoBERT yang diimplementasikan dengan perangkat keras dan perangkat lunak sebagai berikut:

4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Perangkat keras dan perangkat lunak yang digunakan dalam implementasi penelitian ini telah dipilih untuk memastikan proses pelatihan dan analisis data dilakukan secara optimal. Adapun spesifikasi perangkat keras yang digunakan sebagai berikut:

1. *Processor* : 11th Gen Intel® Core™ i7-1165G7 @ 2.80GHz (8CPUs), ~2.8GHz
2. *RAM* : 16384MB RAM
3. *SSD* : 1024GB

Spesifikasi perangkat lunak yang digunakan dalam implementasi penelitian sebagai berikut:

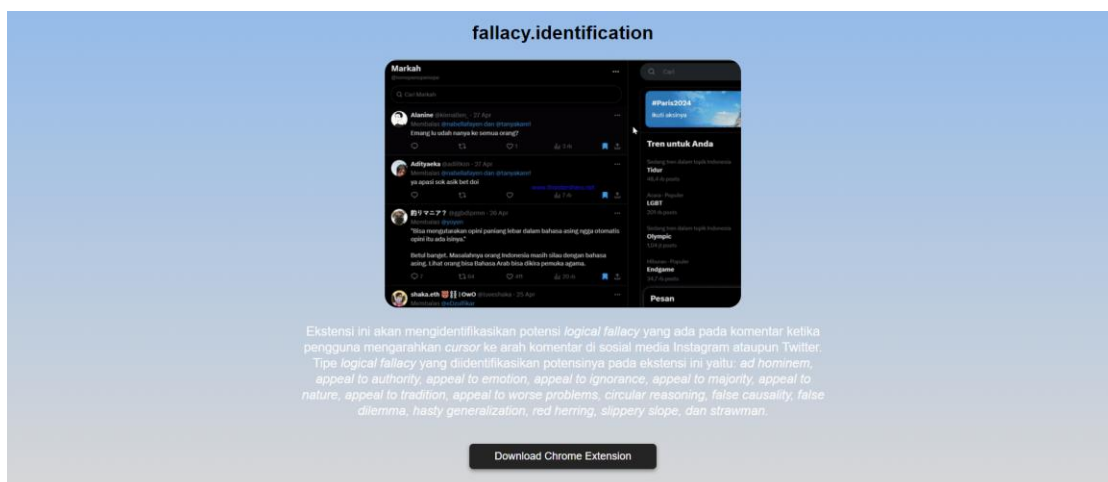
1. Windows 11 Pro 64-bit
2. Google Colab
3. Python 3.10
4. JavaScript
5. CSS & HTML
6. Modelbit
7. Vercel

4.2. Implementasi Ekstensi Google Chrome

Implementasi perancangan ekstensi Google Chrome pada penelitian ini menggunakan HTML, CSS, dan Vanilla JavaScript untuk membangun tampilan antarmuka ekstensi, Modelbit digunakan untuk deployment model, sementara untuk *deployment landing page* digunakan Vercel. Pada bagian implementasi perancangan antarmuka ekstensi ini akan dijelaskan per bagian-bagian yang dilakukan.

4.2.1. Tampilan *Landing Page* Ekstensi Browser

Tampilan landing page ekstensi yang dikembangkan menampilkan informasi berupa nama ekstensi, penjelasan mengenai ekstensi, dan juga tombol Download Chrome Extension. Tombol download ini akan mengarahkan pengguna ke link tempat mendownload ekstensi. Tampilan dari landing page ekstensi bisa dilihat pada Gambar 4.1.



Gambar 4. 1 Tampilan Antarmuka *Landing Page* Ekstensi Browser

4.2.2. Tampilan *Popup* Ekstensi Browser

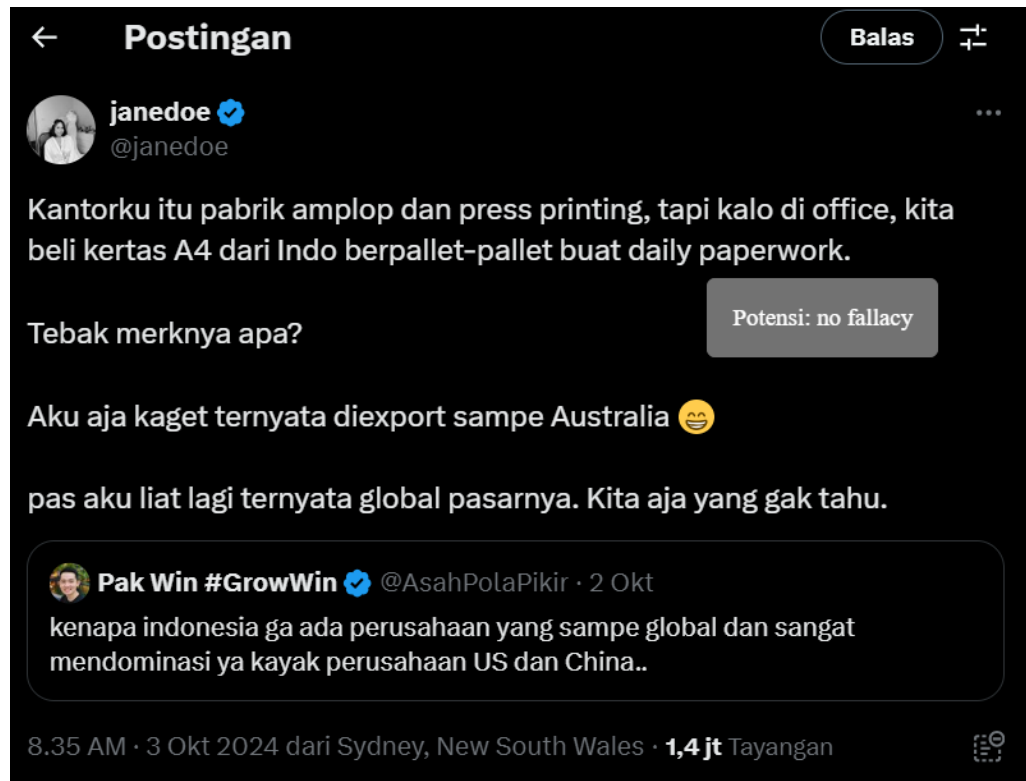
Tampilan *popup* ekstensi browser yang dikembangkan berbentuk seperti *popup* yang menampilkan informasi berupa nama ekstensi, penjelasan mengenai ekstensi, dan juga tombol *Turn On*. Tombol *Turn On* ini memiliki fungsi untuk menjalankan identifikasi potensi *logical fallacy* pada komentar sosial media Instagram ataupun Twitter ketika diaktifkan. Tampilan dari *popup* ekstensi bisa dilihat pada Gambar 4. 2.



Gambar 4. 2 Tampilan Antarmuka *Popup* Ekstensi Browser

4.2.3. Tampilan Antarmuka Identifikasi *Tweet* Menggunakan Ekstensi

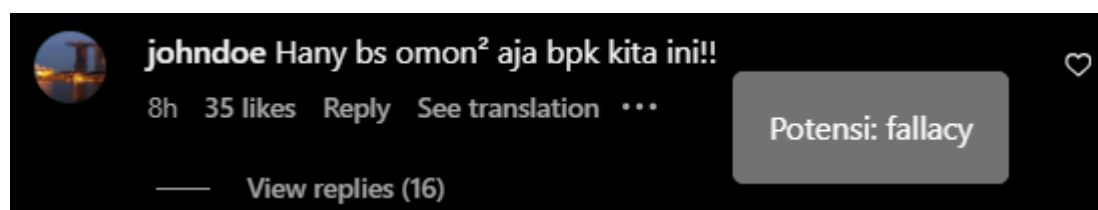
Ketika ekstensi telah diaktifkan dan pengguna membuka situs Twitter, ekstensi dapat melakukan identifikasi ketika pengguna mengarahkan cursor pada sebuah *tweet* ataupun komentar balasan suatu *tweet*, hasil dari identifikasi potensi *logical fallacy* akan ditampilkan dalam bentuk *hover effect* yang akan menghilang ketika *cursor* dipindahkan dari *tweet* atau komentar tersebut. Tampilan dari antarmuka identifikasi pada *tweet* menggunakan ekstensi dapat dilihat pada Gambar 4. 3.



Gambar 4. 3 Tampilan Antarmuka Identifikasi Potensi *Logical Fallacy* pada *Tweet* Menggunakan Ekstensi

4.2.4. Tampilan Antarmuka Identifikasi Komentar Instagram Menggunakan Ekstensi

Setelah ekstensi diaktifkan, pengguna sosial media Instagram dapat melakukan identifikasi potensi *logical fallacy* terhadap sebuah komentar ketika *cursor* diarahkan di atas komentar. Seperti pada penggunaan pada *tweet*, hasil identifikasi hanya akan ditampilkan sepanjang *cursor* tetap berada di atas komentar tersebut dan akan menghilang ketika dipindahkan dari komentar. Tampilan antarmuka identifikasi komentar Instagram menggunakan ekstensi dapat dilihat pada Gambar 4. 4.



Gambar 4. 4 Tampilan Antarmuka Identifikasi Potensi *Logical Fallacy* pada Komentar Instagram Menggunakan Ekstensi

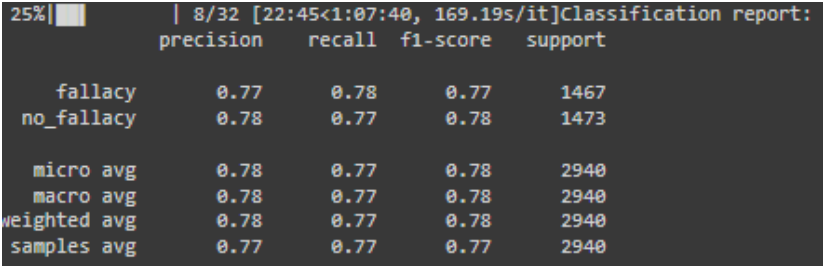
4.3. Implementasi IndoBERT

Pada implementasi pelatihan menggunakan model IndoBERT *resource* yang dibutuhkan cukup tinggi sehingga perlu dilakukan optimalisasi. Optimalisasi yang dilakukan adalah dengan menggunakan *data loader* sehingga seluruh dataset tidak dimuat sekaligus dalam memori, dataset kemudian dibagi menjadi beberapa bagian dan diproses dalam batch yang lebih kecil. Panjang maksimal untuk dataset yang digunakan adalah 210 kata. Optimizer yang digunakan adalah AdamW sesuai dengan *default* model BERT, digunakan juga *dropout layer* sebesar 0.1. Sementara untuk tahap *finetuning* beberapa *hyperparameter* yang diatur adalah sebagai berikut:

1. *Epoch* : 32
2. *Learning rate* : $2e-5$
3. *Batch size* : 16

Epoch sebanyak 32 kali memiliki arti bahwa seluruh dataset *training* yang melewati model akan memperbarui bobot sebanyak 32 kali iterasi. Dengan *learning rate* sebesar $2e-5$, menandakan besarnya perubahan bobot yang dilakukan model ketika melakukan pembaruan berdasarkan gradient fungsi *loss*. Apabila nilai *learning rate* cukup kecil maka nilai dari fungsi error akan turun setelah iterasi, tetapi apabila nilai *learning rate* yang diatur terlalu kecil fungsi error tetap akan turun namun proses pelatihan akan berjalan dengan sangat lambat, sehingga digunakan nilai *learning rate* $2e-5$ karena cukup ideal dalam proses pelatihan model. *Batch size* adalah jumlah sampel data yang digunakan pada setiap iterasi dalam pelatihan model. Semakin besar *batch size* yang digunakan maka akan semakin cepat proses pelatihan dilakukan karena jumlah sampel yang dievaluasi lebih banyak, tetapi semakin besar *batch size* semakin banyak juga jumlah memori yang digunakan dalam proses pelatihan.

Percobaan pertama pada epoch 8 kali dengan *batch size* 16 mendapatkan hasil *F1-score macro average* sebesar 0.78, *micro average* 0.78, *weighted average* sebesar 0.78, dan *samples avg F1-score* sebesar 0.77.



	precision	recall	f1-score	support
fallacy	0.77	0.78	0.77	1467
no_fallacy	0.78	0.77	0.78	1473
micro avg	0.78	0.77	0.78	2940
macro avg	0.78	0.77	0.78	2940
weighted avg	0.78	0.77	0.78	2940
samples avg	0.77	0.77	0.77	2940

Gambar 4. 5. Hasil Pelatihan 8 Epoch

Gambar 4. 6 menampilkan hasil dari pelatihan 16 epoch dengan *batch size* 16 mendapatkan *F1-score macro average* sebesar 0.78, *micro average* sebesar 0.78, *weighted avg* sebesar 0.78, dan *samples avg* sebesar 0.78.

```

50%|███████| 16/32 [42:23<47:37, 168.07s/it]Classification report:
precision    recall  f1-score   support

  fallacy    0.77    0.79    0.78     1467
 no_fallacy  0.79    0.76    0.77     1473

   micro avg    0.78    0.78    0.78     2940
   macro avg    0.78    0.78    0.78     2940
weighted avg    0.78    0.78    0.78     2940
 samples avg    0.78    0.78    0.78     2940

```

Gambar 4. 6. Hasil Pelatihan 16 Epoch

Lalu pada Gambar 4.7 menunjukkan pelatihan dengan 32 epoch dengan *batch size* 16 mendapatkan *F1-score macro average* sebesar 0.79, *micro average* sebesar 0.79, *weighted avg* sebesar 0.79, dan *samples avg* sebesar 0.79.

```

100%|██████████| 32/32 [1:30:25<00:00, 169.54s/it]Classification report:
precision    recall  f1-score   support

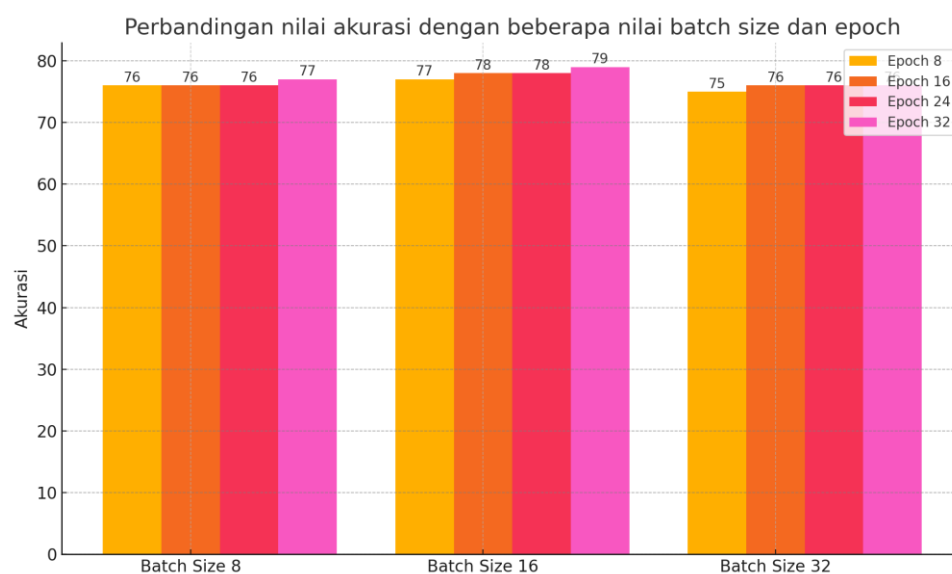
  fallacy    0.77    0.83    0.80     1467
 no_fallacy  0.81    0.75    0.78     1473

   micro avg    0.79    0.79    0.79     2940
   macro avg    0.79    0.79    0.79     2940
weighted avg    0.79    0.79    0.79     2940
 samples avg    0.79    0.79    0.79     2940

```

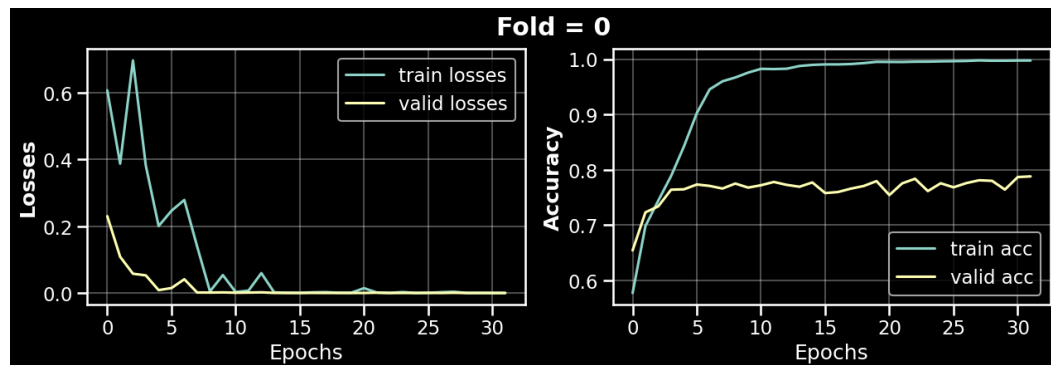
Gambar 4. 7. Hasil Pelatihan 32 Epoch

Pada implementasi ini juga dilakukan beberapa percobaan dengan parameter yang berbeda-beda yaitu digunakan beberapa *batch size* 8, 16, dan 32 dengan jumlah epoch mulai dari 8, 16, 24, dan 32. Hasil dari beberapa percobaan dapat dilihat pada Gambar 4. 8.



Gambar 4. 8. Perbandingan Nilai Akurasi Dengan Berbagai *Batch Size* Dan Epoch

Hasil akurasi yang didapatkan dari beberapa percobaan dengan jumlah epoch dan *batch size* yang berbeda-beda mendapatkan hasil bahwa *batch size* sebesar 16 dan epoch 32 memberikan nilai akurasi tertinggi yaitu mencapai angka 79% pada F1-score dibandingkan dengan parameter-parameter lainnya. Pada Gambar 4.9 dapat dilihat grafik performa pelatihan epoch 32 dengan *batch size* 16.



Gambar 4. 9. Hasil Performa *Hypertuning* 32 Epoch *Batch Size* 16

4.4. Evaluasi Model

Model yang telah dilatih akan dievaluasi untuk mengetahui performa dari model IndoBERT dalam mengidentifikasi potensi *logical fallacy* pada kalimat. Evaluasi model akan menggunakan confusion matrix dan divisualisasikan menggunakan heatmap. Dataset yang digunakan pada tahap pengujian model adalah dataset validasi yang akan memprediksi dataset setelah melalui tahap *preprocessing*, setelah itu hasil prediksi akan dibandingkan dengan label dataset sebenarnya. Tabel 4. 1 menampilkan hasil evaluasi model sementara Gambar 4. 10 menunjukkan *confusion matrix* dari performa model.

Tabel 4. 1. Hasil Evaluasi Model

No	Komentar	Prediction	Actual	Match
0	partai hijau di jerman berpendapat, reaktor nuklir berdampak buruk bagi kita.	Fallacy	Fallacy	True
1	kita tidak boleh membiarkan lakilaki menindas perempuan.	Fallacy	No fallacy	False
2	hanya argumen uji	No fallacy	No fallacy	True
3	dia memiliki pengaruh yang besar	Fallacy	No fallacy	False
4	ya, mereka memang membuat sekolah lebih efektif.	No fallacy	No fallacy	True

5	ya, dan lu adalah pria yang menyukai perang, itu saja. lu suka kalau orang mati.	Fallacy	Fallacy	True
6	anak yang cacat akan menghambat pendidikan anak-anak normal	Fallacy	Fallacy	True
7	semua pengungsi adalah penjahat dan lu tahu itu	Fallacy	Fallacy	True
8	tidak. microsoft jauh lebih baik daripada apple	No fallacy	No fallacy	True
9	ya, karena dia memblok tembakan lebih banyak dibandingkan kiper lainnya.	Fallacy	No fallacy	False
10	gue tidak suka heavy metal secara umum	No fallacy	No fallacy	True
11	gue jelas menentang legalisasi.	Fallacy	No fallacy	False
12	menurutku tidak demikian karena selalu ada bahaya jika anak normal menindas anak cacat karena kelemahan mentalnya.	No fallacy	No fallacy	True
13	ya, orang-orang ini semuanya sakit.	Fallacy	Fallacy	True
14	uli hoene masuk penjara. oleh karena itu selebriti diperlakukan sama seperti nonselebriti.	Fallacy	Fallacy	True
15	oh sekarang, gue tidak akan berdebat dengan lu... apakah lu sudah membaca instruksinya	Fallacy	Fallacy	True
16	informasi di jejaring sosial seringkali tidak dapat dilulka.	No fallacy	No fallacy	True
17	ya, karena lu akan menjadi diktator yang menyedihkan.	Fallacy	Fallacy	True
18	ya. filipi 47 mengatakan bahwa allah akan mengaruniai kita damai sejahtera yang melampaui segala akal.	No fallacy	No fallacy	True
19	akan melanggar hak asasi manusia jika tidak mengizinkan mereka berada di negara yang kata sandinya mengizinkannya.	No fallacy	No fallacy	True
20	semua pendukung teori konspirasi sebaiknya membaca pendulum foucault	Fallacy	No fallacy	False

	terlebih dahulu			
21	menurut gue homeschooling adalah hal yang hebat, baca saja beberapa forum diskusi dan lu akan mengetahuinya sendiri...	Fallacy	Fallacy	True
22	kamu sangat lembek dan lemah. gue kira lu tidak pernah berolahraga, bukan	Fallacy	Fallacy	True
23	gorila seharusnya hidup bebas di alam.	Fallacy	No fallacy	False
24	gue benarbenar merasa jika gue tetap menjadi lajang, gue tidak akan siap menghadapi kehidupan setelah sekolah menengah, secara sosial.	No fallacy	No fallacy	True
25	menurut gue, anakanak sekolah tidak boleh menghabiskan waktu sebanyak biasanya hanya karena karena usia, apalagi karena jenis kelamin.	Fallacy	No fallacy	False
26	seragam terbukti berguna dan diterima dengan baik di sekolah swasta st. blasien. oleh karena itu semua sekolah harus memperkenalkannya.	Fallacy	Fallacy	True
27	pekerjaan tidak boleh disubkontrakkan ke negaranegara berkembang.	No fallacy	No fallacy	True
28	teroris islam merupakan ancaman serius bagi setiap bentuk masyarakat progresif.	Fallacy	No fallacy	False
29	lu adalah satusatunya yang berpikir untuk melecehkan orang-orang ini.	Fallacy	Fallacy	True
30	dengan tujuan memperbaiki situasi setiap orang, pertumbuhan adalah hal yang baik bagi suatu negara.	Fallacy	No fallacy	False
31	manusia sudah berada di bulan.	Fallacy	No fallacy	False
32	gue tidak makan daging satu kali pun selama dua minggu dan gue merasa lemas sepanjang waktu. ini buruk bagi semua orang.	Fallacy	Fallacy	True
33	pola makan vegan umumnya dianggap	No fallacy	No fallacy	True

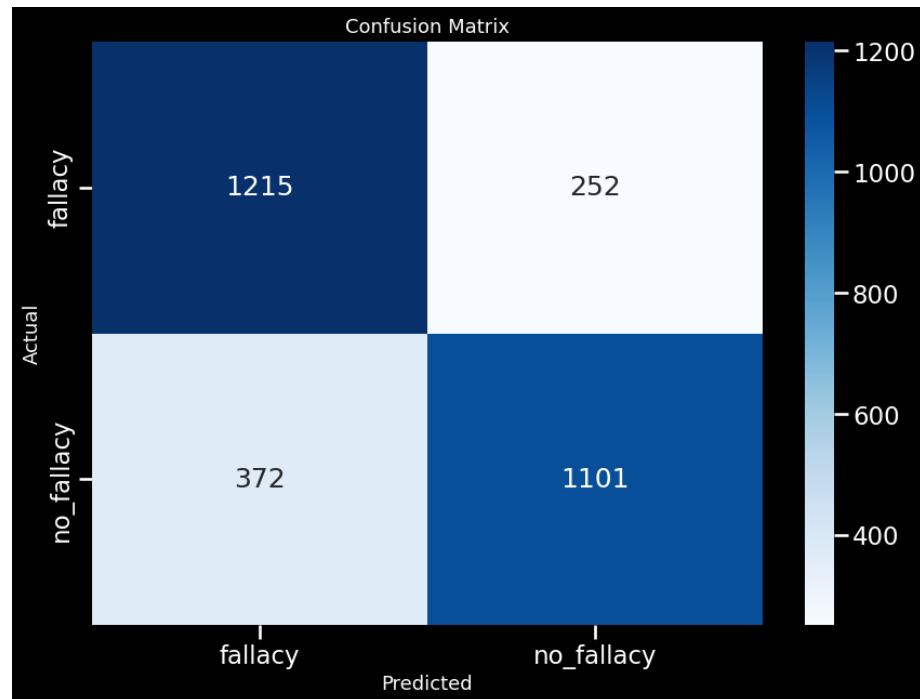
	mengandung lebih banyak sereal, buahbuahan, sayuran, kacang-kacangan dan bijibijian dibandingkan pola makan nonvegetarian.			
34	dan lu sama bodohnya dengan mereka... logika lu salah dan lu memerlukan bimbingan perdebatan.	Fallacy	Fallacy	True
35	yah, kamu jelek.	Fallacy	Fallacy	True
36	tentu saja tidak, ahli macam apa yang lu sarankan untuk melakukan hal ini	Fallacy	Fallacy	True
37	guru gue selalu mengambil ponsel kami ketika kami menggunakannya di kelas. gue yakin dia punya alasan.	Fallacy	Fallacy	True
38	tentu saja mereka bisa lebih kejam lagi. masalahnya dengan quentin tarantino adalah dia memahami pemisahan antara kehidupan nyata dan film. dia memahami bahwa film seharusnya bersifat liar dan keterlaluhan dan itulah yang membuat film-filmnya terkenal, terutama yang berkaitan dengan kekerasan. adegan dari film seperti django unchained mempunyai bagian tubuh yang meledak setelah ditembak dengan pistol sederhana. itu seharusnya untuk hiburan, dan tidak lebih.	No fallacy	No fallacy	True
39	mereka membawa beberapa kilo batuan bulan kembali ke bumi, dan lu tidak bisa begitu saja memalsukan material batuan berumur empat miliar tahun.	Fallacy	No fallacy	False
40	microsoft bukanlah apel baru	No fallacy	No fallacy	True
41	tidak dapat disangkal bahwa emisi co2 dan suhu rata-rata atmosfer sama-sama meningkat, banyak penelitian mengenai hal tersebut.	Fallacy	No fallacy	False

42	imigran noneropa menghancurkan budaya kami.	Fallacy	Fallacy	True
43	taliban, iis dll harus dimusnahkan sampai ke akarakarnya.	Fallacy	No fallacy	False
44	tidak masalah hewan tidak bisa berpikir sendiri.	Fallacy	No fallacy	False
45	lu hanya idiot kompetitif.	Fallacy	Fallacy	True
46	hitler menyensor media. kim yong il menyensor media. hanya orang jahat yang menyensor media.	Fallacy	Fallacy	True
47	orang-orang membuang sampah ke luar jendela. setiap orang yang membuang sampah ke luar jendela akan menumpuk di jalanan.	Fallacy	Fallacy	True
48	mereka memang memberikan dampak positif pada anak gue, dia lebih bersosialisasi sekarang. oleh karena itu, mereka juga baik untuk masyarakat.	Fallacy	Fallacy	True
49	lu jelas tidak tahu tentang film. lebih baik beli tv yang lebih besar, mungkin lu bisa melihat lebih baik apa yang terjadi di layar.	Fallacy	Fallacy	True
50	tidak. harga susu sudah terlalu tinggi	Fallacy	Fallacy	True
...
2889	biar adil, putusan maka dan putusan baleg di pakai semua.	No fallacy	No fallacy	True
2890	menteri ini makan uang pensiun keringat uang rakyat	Fallacy	Fallacy	True
2891	ketika nafsu kekuasaan diatur sedemikian rupa	No fallacy	Fallacy	False
2892	orang-orang islam itu perasaannya sangat sensitif tapi selalu menghina penganut agama lain, terutama kristen.	Fallacy	Fallacy	True
2893	tapi tidak seheboh dan sedramatis ini sampai buat thread sendiri kan padahal lu	Fallacy	Fallacy	True

	juga yang perdebatan aneh2. liat saja lu kalo di thread debat serius juntrungannya tidak jelas puol.			
2894	wkwkwkwk kocak entar tiba berubah lagi dia	Fallacy	Fallacy	True
2895	culas dan tidak tau malu	Fallacy	Fallacy	True
2896	sok santun itu orang heh	Fallacy	Fallacy	True
2897	gue seratus persen sudah kagak percaya sama penguasa.	Fallacy	No fallacy	False
2898	omongane koyo belut, licin	Fallacy	Fallacy	True
2899	batupahit semakin pahit	No fallacy	No fallacy	True
2900	sekedar menghormati tanpa menjalankan apa yang jadi putusan mk..	No fallacy	No fallacy	True
2901	maka cukup dihormatin, yang dijalanin putusan dpr dasarrrrr	No fallacy	No fallacy	True
2902	turun ke jalan	No fallacy	No fallacy	True
2903	halah gimmick...	Fallacy	Fallacy	True
2904	coba kita lihat, apakah omongannya berkualitas....	Fallacy	Fallacy	True
2905	adoh bisa apa nih orang....	Fallacy	Fallacy	True
2906	jancukkkkk ah	No fallacy	Fallacy	False
2907	menguntungkan anak nya hormati keputusan final maka , merugikan anak nya itu situasi konstitusional	No fallacy	No fallacy	True
2908	klaw team projo yang ngmong pengen kali letop jidatnya	Fallacy	Fallacy	True
2909	sehat selalu presiden ku	No fallacy	No fallacy	True
2910	cuci piring	No fallacy	No fallacy	True
2911	lawan , hancur kan politik dinasti	Fallacy	No fallacy	False
2912	juru bicara begal	Fallacy	Fallacy	True
2913	orang islam yang ikut demo jangan lupa sholat dan meminta doa kepada allah mohon kebaikan untuk negara kita.	No fallacy	No fallacy	True
2914	selamatkan demokrasi hukum dan keadilan negeri ini	No fallacy	No fallacy	True

2915	raja ya raja.	No fallacy	No fallacy	True
2916	hari ini bukti bahwa kekuasaan tertinggi ada pada rakyat, jangan biarkan mereka berpikir kekuasaan ada di kaki mereka..	Fallacy	No fallacy	False
2917	aku tidak ketagihan soalnya lebih ke reel sama short yutub	No fallacy	No fallacy	True
2918	klarifikasi.. terus meminta maaf. ini tuh takutnya malah membuat orang berpikir seenaknya saja bertindak, karena kan tinggal klarifikasi dan meminta maaf selesai	Fallacy	Fallacy	True
2919	sebagus apapun calonnya kalo yang ngusung pdip tidak akan gue pilih	Fallacy	Fallacy	True
2920	ntar kalo sudah dideportasi playing victim kayak paling di dzolimi	Fallacy	Fallacy	True
2921	semoga ibu bisa lanjut di pemerintahan	No fallacy	No fallacy	True
2922	hebat banget	No fallacy	No fallacy	True
2923	aku masih bingung masalah beginian	Fallacy	No fallacy	False
2924	kebanyakan siswa di kelas lebih suka minuman rasa anggur, jadi pasti anggur itu rasa yang paling disukai anak sma	Fallacy	Fallacy	True
2925	banyak pihak yang provaksin juga tidak akan punya pendapat yang terinformasi dan kenyataannya memang tidak	Fallacy	Fallacy	True
2926	pak heru mendukung pelarangan penggunaan hp dan dia itu goblok banget dah, jadi biarkan saja guruguru buat menyimpan hp	Fallacy	Fallacy	True
2927	banyak dari kami yang bertanggung jawab begitu dan tidak mungkin kami semua salah	Fallacy	Fallacy	True
2928	alkitab sudah jelas banget makanya ada 2 miliar orang kristen	Fallacy	Fallacy	True
2929	3 anggota dpr pernah berselingkuh. anggota dpr tukang selingkuh	Fallacy	Fallacy	True

2930	apa lu tidak tau kalo poster ini sangat populer di sub reddit oculus	Fallacy	Fallacy	True
2931	pendapat dia tidak bisa dipercaya, orangnya sombong banget.	Fallacy	Fallacy	True
2932	siapa juga yang mau mendengarkan sih ani mengomong soal lingkungan orang dia sendiri saja pakai mobil mewah tiap hari, sudah jelas omongannya bacot doang	Fallacy	Fallacy	True
2933	gue sudah dua kali ketemu bule yang sombong, kayaknya memang orang barat itu pada arogan	Fallacy	Fallacy	True
2934	lu kan cuma ibu rumah tangga, mana mengerti soal bisnis	Fallacy	Fallacy	True
2935	pak lurah bilang kita harus hidup sehat, tapi dia sendiri perokok berat. mana bisa kita mengikuti sarannya	Fallacy	Fallacy	True
2936	pengacara itu bilang kita harus taat hukum, tapi dulu pas kuliah sering tawuran. mana bisa dipercaya argumennya	Fallacy	Fallacy	True
2937	mending mana, nikah muda atau karir sukses	Fallacy	Fallacy	True
2938	lo tidak setuju sama gue berarti lo dukung koruptor	Fallacy	Fallacy	True
2939	kalo tidak setuju vaksin wajib, berarti lo antisains ya	Fallacy	Fallacy	True



Gambar 4. 10. *Confusion Matrix*

Adapun hasil dari *confusion matrix* divisualisasikan menggunakan *heatmaps* pada Gambar 4. 10. Diketahui bahwa sebanyak 1215 data fallacy bernilai *True Positive* dan 252 bernilai *False Negative*, dengan data no_fallacy terdiri dari sebanyak 1101 data bernilai *True Positive* dan 372 data bernilai *False Negative*. Sehingga dari nilai-nilai yang didapatkan tersebut dapat disimpulkan hasil akurasi dari model sebanyak 2940 dapat dijabarkan seperti berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \\
 &= \frac{1215+1101}{1215+252+372+1101} \times 100\% \\
 &= \frac{2316}{2940} \times 100\% \\
 &= 78,77\%
 \end{aligned}$$

Dari hasil evaluasi model didapatkan akurasi dengan nilai 78,77% di mana 21,23% adalah hasil evaluasi yang bernilai error, hal ini mungkin disebabkan oleh beberapa hal seperti:

1. Perbedaan struktur bahasa informal dan formal yang digunakan pada dataset penelitian terdahulu dan dataset yang diperoleh dari sosial media kemungkinan membuat model kesulitan dalam mengidentifikasi potensi *logical fallacy*.

2. Model kemungkinan rentan melakukan kesalahan dalam mengidentifikasi komentar yang mengandung sarkasme, hiperbola, ataupun kiasan.

Diketahui juga bahwa jumlah epoch dalam pelatihan model mempengaruhi akurasi dalam identifikasi, epoch yang terlalu tinggi menyebabkan *overfitting* yang menyebabkan model untuk menghafal data yang digunakan dalam pelatihan dan kesulitan ketika berhadapan dengan data evaluasi yang belum pernah dilihat.

4.5. Pengujian Sistem

Pengujian sistem dilakukan dengan 20 data testing yang didapatkan melalui ekstensi yang kemudian diidentifikasi potensi *logical fallacy*-nya dengan model yang telah dilatih. Hasil dari pengujian dapat dilihat pada Tabel 4. 2 dan pada Tabel 4. 3 dapat dilihat koreksi dari kesalahan identifikasi oleh model.

Tabel 4. 2. Hasil Pengujian Ekstensi

No	Komentar	Prediction	Actual	Match
1	Ketika bencana berbicara bencana	Fallacy	Fallacy	True
2	Apa cuma aku, yg setiap melihat berita soal bapak ini, yg muncul di otak cuma kata "kocak, konyol, gaptek" ya?	Fallacy	Fallacy	True
3	Kalo di Taiwan kalo ada peringatan gempa langsung masuk ke hp pemberitahuan nya mala sbm terjadi nya gempa berlangsung.dan kaya peringatan darurat lain nya misal kaya ada latihan militer dan akan terjadi nya gempa.muda2 han aplikasi nya seperti itu	No fallacy	No fallacy	True
4	Sumpah tampangnya kayak orang b**on wkwk	Fallacy	Fallacy	True
5	Keren, gini dong...	No fallacy	No fallacy	True
6	Haha org kyk dia dipercaya	Fallacy	Fallacy	True
7	Banyak bacot nih Mentri njirr	Fallacy	Fallacy	True
8	Temu adalah platform e-commerce yang mempertemukan langsung end user dengan produsen.	No fallacy	No fallacy	True

9	Hampir semua sektor kacau. Populasi membludak tpi ga terdidik	Fallacy	Fallacy	True
10	Min nanya dong, dari awal kan ideologi Indonesia Pancasila which is Demokrasi, tp kok bisa diizinin berdirinya partai berbasis komunis???	No fallacy	No fallacy	True
11	Lucu caplin, semua juga tau itu	No fallacy	Fallacy	False
12	Umur 70 tahun jalannya masih kayak umur 40 an	Fallacy	No fallacy	False
13	Kalau saya secara pribadi, memandang manusia kedalam 2 tipe: oportunistis dan polos.	Fallacy	Fallacy	True
14	Ayolah bukan saat nya buat satir dan sarkas. Ini momentum buat satu suara dgn hacker pro rakyat, suarakan dgn konten lantang dan jelas	No fallacy	No fallacy	True
15	ini bukan ajang pamer, ini cuma memberi contoh cara memberi respon yg baik & benar	No fallacy	No fallacy	True
16	Mon maaf, life hacks pake lipbalm ini selain buat lipbalm kalo aku dikulit yg super kering kalo abis pake sunscreen trs pake produk ini dibuat blush kece bat. Blushy glowingnya berasa pake balm korea wkwkww	No fallacy	No fallacy	True
17	Karena bukan lo yang ngalamin jadi lo gampang banget ngebacot ga jelas gt	Fallacy	Fallacy	True
18	Bagus ga sih di bibir gelap? Lip balm nya madam Gie aja pas aku coba malah ga bisa nyatu di bibir :(No fallacy	No fallacy	True
19	Buset.... Minuman di campur penenang kuda... Pantas g pd sadarkan diri	No fallacy	No fallacy	True
20	semoga semua korbannya bisa bangkit dan le0as dr trauma dan tidak melakukan hal yg sama pd orang lain	No fallacy	No fallacy	True

Tabel 4. 3. Koreksi Identifikasi

No	Komentar	Penjelasan	Koreksi
1	<p>Komentar:</p> <p>Lucu caplin, semua juga tau itu</p> <p>Identifikasi Model:</p> <p>No fallacy</p>	<p>Pada komentar ini konteks yang dibahas adalah komentar menyerang karakter lawan bicara, tetapi karena model tidak memiliki konteks sepenuhnya kemungkinan membuat model menganggap bahwa kata ‘lucu’ pada komentar tersebut bukanlah sebagai serangan kepada lawan bicara.</p>	Fallacy
2	<p>Komentar:</p> <p>Umur 70 tahun jalannya masih kayak umur 40 an</p> <p>Identifikasi Model:</p> <p>Fallacy</p>	<p>Konteks yang dibahas pada komentar ini adalah mengenai seorang selebriti, tetapi karena model tidak memiliki konteks sepenuhnya kemungkinan menyebabkan model menganggap komentar tersebut sebagai generalisasi tentang semua orang yang berumur 70 tahun.</p>	No Fallacy

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari penelitian identifikasi potensi *logical fallacy* di social media menggunakan IndoBERT ada beberapa kesimpulan yang dapat diambil seperti berikut:

1. Model mampu mengidentifikasi potensi logical fallacy pada komentar di sosial media menggunakan IndoBERT, sesuai dengan tujuan yang telah dituliskan dalam penelitian ini.
2. Hasil akurasi dalam mengidentifikasi potensi logical fallacy menggunakan IndoBERT mencapai angka sebesar 78,77%.
3. Performa dari IndoBERT cukup baik dalam mengidentifikasi potensi logical fallacy terhadap komentar media sosial.
4. Hasil dari Finetuning IndoBERT dengan epoch 16, batch size sebanyak 16, dan learning rate $2e-5$ mendapatkan hasil akurasi yang cukup.

5.2. Saran

Walaupun model yang telah dilatih mampu mengidentifikasi potensi *logical fallacy* pada komentar sosial media dengan cukup baik tetapi tingkat error pada model mencapai 21,23% yang kemungkinan diakibatkan oleh kurangnya dataset, adapun beberapa cara yang dapat dilakukan untuk meningkatkan dan mengembangkan topik penelitian yang serupa dengan penelitian ini, antara lain:

1. Penelitian selanjutnya disarankan untuk melakukan identifikasi dan klasifikasi multilabel *logical fallacy* pada penelitian selanjutnya.
2. Penelitian selanjutnya disarankan untuk memberikan penjelasan dari tiap hasil identifikasi.
3. Penelitian selanjutnya dapat menambahkan metode yang dapat mengkonfirmasi validitas identifikasi *logical fallacy*.

DAFTAR PUSTAKA

- Adams, E., & Rollings, A. (2006). *Fundamentals of game design*. Prentice Hall. ISBN 978-0131687479.
- Aditya, Y.R. (2021). *Penerapan metode finite state machine pada game sejarah "Legenda Depati Parbo" menggunakan RPG Maker MV*.
- Hidayat, D. N., Nurhalimah, Defianty, M., Kultsum, U., Zulkifli., & Sufyan, A. (2020). Logical fallacies in sosial media: A discourse analysis in political debate. Proceedings of the 8th International Conference on Cyber and IT Service Management (CITSM), 23-24.
- Situmorang, G., & Purba, R. (2024). Potensi Depresi dari Unggahan Media Sosial X Menggunakan Teknik NLP dan Model IndoBERT. Building of Informatics, Technology and Science (BITS), 649-661. <https://doi.org/10.47065/bits.v6i2.5496>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A, N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), 47–82. <https://doi.org/10.48550/arXiv.1706.03762>.
- 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.
- Koto, F., Rahimi, A., Lau, J, H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained
- Young, J. M. (2006). A Case Study of Hasty Generalization. Akita Prefectural University Comprehensive Science Research , 95-98.
- LaBoissiere, M. C. (2010). 42 Fallacies. Ohio: CreateSpace Independent Publishing Platform.
- Tarigan, Henry Guntur. (1995). Pengajaran Morfologi. Bandung: Angkasa.
- Copi, I. M., & Cohan, C. (1990). Introduction to Logic: Eighth Edition. New York: Macmilan Publishing Company.

- Mayfield, Marlys. (2014). *Thingking for Yourself: Developing Critical Thinking Skills Through Reading and Writing*. Wadsworth: Lyn Uhl Publisher.
- Solso, Robert L, dkk. (2008). *Psikologi Kognitif*. Jakarta: Erlangga.
- Ganesh, P., Chen, Y., Lou, X., Khan, M. A., Yang, Y., Sajjad, H., Nakov, P., Chen, D., & Winslett, M. (2021). Compressing Large-Scale Transformer-Based Models: A Case Study on BERT. *Transactions of the Association for Computational Linguistics*, 9, 1061–1080. <https://doi.org/10.48550/arXiv.2002.11985>
- Luo, L., & Wang, Y. (2019). EmotionX-HSU: Adopting Pre-trained BERT for Emotion Classification. <http://arxiv.org/abs/1907.09669>.
- Khan, M., Naeem, M. R., Al-Ammar, E. A., Ko, W., Vettikalladi, H., & Ahmad, I. (2022). Power Forecasting of Regional Wind Farms via Variational Auto-Encoder and Deep Hybrid Transfer Learning. *Electronics (Switzerland)*, 11(2). <https://doi.org/10.3390/electronics11020206>.
- Fimoza, D. (2019). Analisis Sentimen Terhadap Film Indoensia Dengan Pendekatan BERT. *Expert Systems with Applications*, 118, 272–299. Diakses dari <http://repositori.usu.ac.id/handle/123456789/30445%0Ahttps://linkinghub.elsevier.com/retrieve/pii/S0957417418306456>.
- Archie, L, C. (1997). *Ad Populum: Appeal to Popularity*. Diakses pada 20 Februari 2024, dari <https://philosophy.lander.edu/logic/popular.html>.
- Logical Fallacies. False Dilemma. Diakses pada 20 Februari 2024, dari <https://www.logicalfallacies.org/false-dilemma.html>.
- Logically Fallacious. Appeal to Common Belief. Diakses pada 02 Oktober 2024, dari <https://www.logicallyfallacious.com/logicalfallacies/Appeal-to-Common-Belief>.
- Alammar, J. (2018). *The Illustrated Transformer*. Diakses pada 30 Maret 2024, dari <https://jalammar.github.io/illustrated-transformer>.
- Riyanto, A, D. (2024). *Hootsuite (We are Sosial): Data Digital Indonesia 2024*. Diakses pada 1 Maret 2024, dari <https://andi.link/hootsuite-we-are-sosial-data-digital-indonesia-2024>.
- Wikipedia. Browser Extension. Wikimedia Foundation. Diakses pada 20 Juli 2024, dari https://en.wikipedia.org/wiki/Browser_extension.
- Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Roth schild, D., et al., 2018. The science of fake news. *Science* 359, 1094–1096.

- Anonim. (2024). COCOLOFA: News Comment Sections with Common Logical Fallacies. Anonymous ACL submission.
- Vorakitphan, V., Cabrio, E., & Villata, S. (2022). PROTECT: A Pipeline for Propaganda Detection and Classification. CLiC-it 2021 - Italian Conference on Computational Linguistics, 352-358. <https://doi.org/10.4000/books.aaccademia.10884>.
- Goffredo, P., Chaves, M., Villata, S., & Cabrio, E. (2023). Argument-based Detection and Classification of Fallacies in Political Debates. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 11101—11112. <https://doi.org/10.18653/v1/2023.emnlp-main.684>.
- Ruiz-Dolz, R. & Lawrence, J. (2023). Detecting Argumentative Fallacies in the Wild: Problems and Limitations of Large Language Models. Proceedings of the 10th Workshop on Argument Mining, 1-10. <https://doi.org/10.18653/v1/2023.argmining-1.1>.
- Salsabila, N, A., Winatmoko, Y, A., Septiandri, A, A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. IEEE 2018 International Conference on Asian Language Processing (IALP), 226-229. <https://doi.org/10.1109/IALP.2018.8629151>.
- Sourati, Z., Venkatesha, V, P, P., Deshpande, D., Rawlani, H., Illievski, F., Sandlin, H., & Mermoud, A. (2023). Robust and Explainable Identification of Logical Fallacies in Natural Language Arguments. <https://arxiv.org/abs/2212.07425>.
- Sahai, S, Y., Balalau, O., & Horincar, R. (2021). Breaking Down the Invisible Wall of Informal Fallacies in Online Discussions. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 644-657. <https://doi.org/10.18653/v1/2021.acl-long.53>.
- Jin, Z., Lalwani, A., Vaidhya, T., Shen, X., Ding, y., Lyu, Z., Sachan, M., Mihalcea, R., & Schölkopf, B. (2022). Logical Fallacy Detection. <https://doi.org/10.48550/arXiv.2202.13758>.
- Raiderbdev & Watchful1. Separate Dump Files For The Top 40k Subreddits, Through The End Of 2023. Diakses pada 28 Februari 2024, dari https://www.reddit.com/r/pushshift/comments/1akrhg3/separate_dump_files_for_the_top_40k_subreddits/.

Modelbit. Modelbit. Diakses pada 1 Agustus 2024, dari <https://doc.modelbit.com/getting-started>.

Habernal, I., Hannemann, R., Pollak, C., Klamm, C., Pauli, P., & Gurevych, I. (2017). Argotario: Computational Argumentation Meets Serious Games. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 7-12. <https://doi.org/10.18653/v1/D17-2002>.

Alzahrani, E., & Jololian, L. (2021). How Different Text-preprocessing Techniques Using The BERT Model Affect The Gender Profiling of Authors. Computing Research Repository. <https://arxiv.org/abs/2109.13890>.



KEPUTUSAN
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI
NOMOR : 3879/UN5.2.14.D/SK/TD.06/2024

DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

- Membaca : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 14 Oktober 2024 perihal permohonan ujian skripsi:
Nama : FIRDHA ISLY RAMADHANI
NIM : 181402029
Program Studi : Sarjana (S-1) Teknologi Informasi
Judul Skripsi : Identifikasi Potensi Logical Fallacy Pada Komentar Sosial Media Menggunakan Indobert
- Memperhatikan : Bahwa Mahasiswa tersebut telah memenuhi kewajiban untuk ikut dalam pelaksanaan Meja Hijau Skripsi Mahasiswa pada Program Studi Sarjana (S-1) Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara TA 2023/2024.
- Menimbang : Bahwa permohonan tersebut diatas dapat disetujui dan perlu ditetapkan dengan surat keputusan
- Mengingat : 1. Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah Nomor 17 tahun 2010 tentang pengelolaan dan penyelenggara pendidikan.
3. Keputusan Rektor USU Nomor 03/UN5.1.R/SK/SPB/2021 tentang Peraturan Akademik Program Sarjana Universitas Sumatera Utara.
4. Surat Keputusan Rektor USU Nomor 1876/UN5.1.R/SK/SDM/2021 tentang pengangkatan Dekan Fasilkom-TI USU Periode 2021-2026

MEMUTUSKAN

- Menetapkan :
Pertama : Membentuk dan mengangkat Tim Penguji Skripsi mahasiswa sebagai berikut:

Ketua : Dr. Erna Budhiarti Nababan M.IT
NIP: 196210262017042001

Sekretaris : Ade Sarah Huzaifah S.Kom., M.Kom
NIP: 198506302018032001

Anggota Penguji : Dr. Baihaqi Siregar S.Si., MT.
NIP: 197901082012121002

Anggota Penguji : Dedy Arisandi ST., M.Kom.
NIP: 197908312009121002

Moderator : -

Panitera : -
- Kedua : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2024.
- Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatunya akan diperbaiki sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan dalam surat keputusan ini.

- Tembusan :
- 1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi
 - 2. Yang bersangkutan
 - 3. Arsip

Medan, 14 Oktober 2024
Ditandatangani secara elektronik oleh:
Dekan



Maya Silvi Lydia
NIP 197401272002122001

