

**IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN  
KOMENTAR MEDIA SOSIAL MENGGUNAKAN  
*BIDIRECTIONAL LONG SHORT-TERM  
MEMORY* DAN *INDOBERT  
EMBEDDING***

**SKRIPSI**

**STEPHANI ULI BASA SILITONGA**

**201402068**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2025**

**IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN KOMENTAR  
MEDIA SOSIAL MENGGUNAKAN *BIDIRECTIONAL LONG SHORT-  
TERM MEMORY* DAN *INDOBERT EMBEDDING***

**SKRIPSI**

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

**STEPHANI ULI BASA SILITONGA  
201402068**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2025**

## PERSETUJUAN

Judul : IDENTIFIKASI PERNYATAAN MISOGINI  
BERDASARKAN KOMENTAR MEDIA SOSIAL  
MENGUNAKAN *BIDIRECTIONAL LONG SHORT-  
TERM MEMORY* DAN *INDOBERT EMBEDDING*

Kategori : SKRIPSI

Nama : STEPHANI ULI BASA SILITONGA

Nomor Induk Mahasiswa : 201402068

Program Studi : SARJANA (S-1) TEKNOLOGI INFORMASI

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA

Medan, 10 Januari 2025

Komisi Pembimbing:

Pembimbing 2,



Sarah Purnamawati, S.T., M.Sc.  
NIP. 198302262010122003

Pembimbing 1,



Umayah Ramadhani Putri Nasution,  
S.TI., M.Kom.  
NIP. 199104112024062001

Diketahui/disetujui oleh  
Program Studi S1 Teknologi Informasi

Ketua,



Dedy Arisandi, S.T., M.Kom.  
NIP. 197908312009121002



**PERNYATAAN**

IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN KOMENTAR  
MEDIA SOSIAL MENGGUNAKAN *BIDIRECTIONAL LONG SHORT-  
TERM MEMORY* DAN *INDOBERT EMBEDDING*

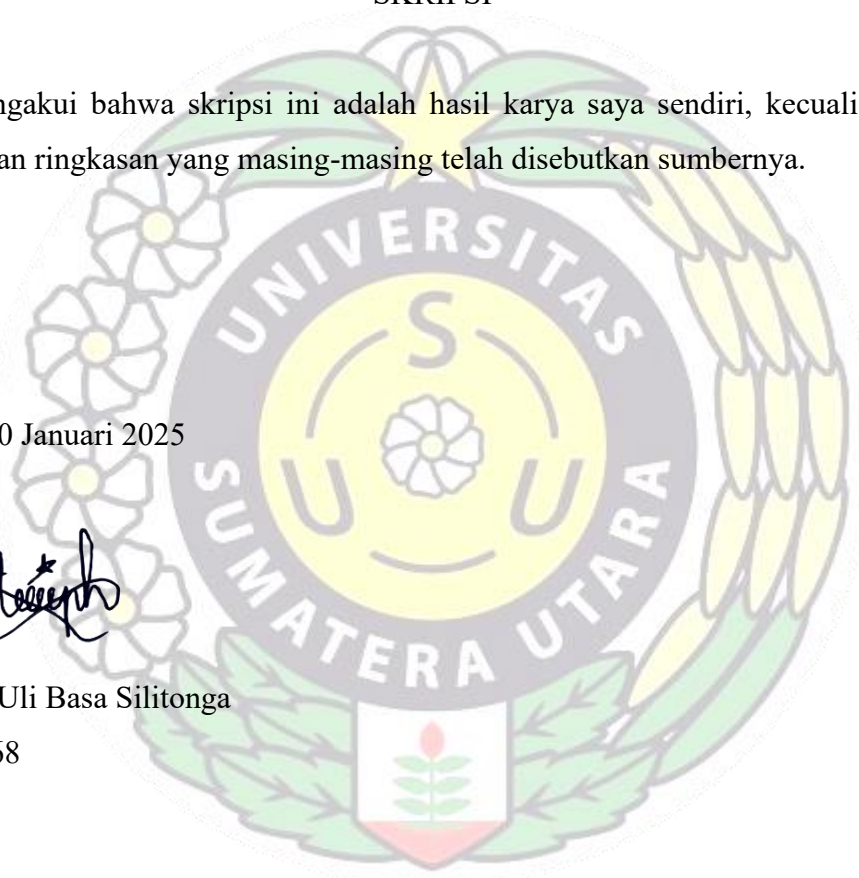
**SKRIPSI**

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 10 Januari 2025



Stephani Uli Basa Silitonga  
201402068



## UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan kepada Tuhan yang Maha Esa atas segala kasih dan karunia-Nya yang telah memampukan penulis dalam menuntaskan skripsi yang berjudul “**Identifikasi Pernyataan Misogini Berdasarkan Komentar Media Sosial Menggunakan *Bidirectional Long Short-Term Memory* dan IndoBERT *Embedding***” sebagai syarat kelulusan dalam meraih gelar sarjana dalam bidang Program Studi S-1 Teknologi Informasi di Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara. Adapun penuntasan skripsi ini tidak mungkin terwujud tanpa adanya kontribusi yang terlibat dari pihak-pihak di sekitar penulis, baik dalam hal bimbingan, dukungan maupun doa. Oleh karena itu, melalui kesempatan ini penulis hendak mengucapkan terima kasih kepada:

1. Keluarga penulis, Ayah Israel Silitonga dan Ibu Bertha Evelina Pasaribu yang selalu dengan senantiasa mendoakan, menyokong, dan menguatkan penulis selama menjalani kehidupan di perkuliahan hingga bisa sampai ditahap penyelesaian skripsi ini.
2. Kedua adik laki-laki penulis, Jeremi Togu Silitonga dan Hesekiel Silitonga yang selalu dengan senantiasa mendoakan, menyemangati, dan menghibur penulis.
3. Ibu Dr. Maya Silvi Lydia, M.Sc., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.
4. Bapak Dedy Arsandi S.T., M.Kom., selaku Ketua Program Studi S-1 Teknologi Informasi Universitas Sumatera Utara.
5. Ibu Umaya Ramadhani Putri Nasution, S.TI., M.Kom., selaku Dosen Pembimbing I dan Ibu Sarah Purnamawati, S.T., M.Sc., selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk memberikan bimbingan dan dukungan, serta kritik dan saran yang membangun, yang telah membantu penulis dalam menyelesaikan skripsi.
6. Bapak dan Ibu Dosen di Program Studi S-1 Teknologi Informasi, Universitas Sumatera Utara, yang telah memberikan ilmu dan pengetahuan kepada penulis sepanjang masa perkuliahan. Serta Bapak dan Ibu *Staff* di Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.
7. Teman-teman penulis semenjak SD, Welni Damanik, Nova Nainggolan, dan Vanessa Napitupulu yang telah memberikan dukungan dan doa kepada penulis.



Serta, tetap solid dalam mempertahankan dan menjalin hubungan pertemanan hingga saat ini.

8. Teman-teman penulis semenjak SMP, Hana Simamora, Septri Tamba, Naomi Napitupulu, Frisca Silaban, Louisa Sirait, dan Sheren Tan yang menyemangati, mendukung, dan menghibur penulis dengan segala agenda dan cerita yang ada.
9. Teman-teman penulis semenjak SMA, Elsa Sitompul, Dhea Sianturi, Fromas Simatupang, Hilda Simangunsong, dan Jessica Parapat yang membantu penulis untuk terus termotivasi dan semangat dalam menyelesaikan perkuliahan. Serta, dengan senantiasa mendoakan dan menghibur penulis dengan pembahasan dan keributan mengenai *K-POP* dan hal-hal lainnya yang tiada henti.
10. Teman-teman penulis semenjak awal perkuliahan, Kelvin, Kevin, Pretty, Frans, dan Yeftha yang telah saling berbagi suka duka menghadapi kehidupan baik di luar maupun dalam perkuliahan serta kepanitiaan, terlebih masa tugas akhir.
11. Teman-teman *Musketeer*, Christine Amanda, Wynne Jovita, dan Atikah Husna yang telah saling berbagi kebahagiaan, saling memberi dukungan dan bantuan selama menghadapi berbagai masalah, baik di luar maupun dalam perkuliahan.
12. Teman-teman penulis lainnya, Monica, Retno, Syavira, Fildzah, Jane, Dara, Bobby Manik, Felix, Kevin Bangun, dan Rifqi yang telah membersamai kehidupan huru-hara selama masa perkuliahan.
13. Teman-teman Angkatan 2020 Teknologi Informasi, Universitas Sumatera Utara, yang tidak dapat penulis ungkapkan satu-satu. Terima kasih telah saling memberi dukungan dan semangat, serta memori baik selama berkuliah.
14. Kakak/Abang Senior di Program Studi S-1 Teknologi Informasi yang telah berbagi ilmu dan pengalamannya dalam menjalani masa perkuliahan.
15. Seluruh keluarga dan teman-teman lain penulis yang tak memungkinkan untuk disebut satu per satu, yang juga telah memberi dorongan serta semangat kepada penulis hingga berhasil menyelesaikan skripsi ini.

Medan, 10 Januari 2025  
Penulis,



Stephani Uli Basa Silitonga  
201402068

## ABSTRAK

Dengan adanya anonimitas dalam penggunaan media sosial, para pengguna secara bebas mengunggah konten atau pun pendapat mereka melalui kolom komentar yang tersedia. Oleh karena hal tersebut, tidak jarang ditemukan akun-akun di kolom komentar media sosial yang dengan gamblang mengutarakan kebencian terhadap seseorang, terutama terhadap perempuan atau yang dapat dikatakan sebagai misogini. Akibatnya banyak perempuan merasa resah dan tidak nyaman karena telah menerima komentar yang merendahkan, meremehkan, dan melecehkan mereka. Untuk melakukan pengidentifikasian terhadap akun-akun yang berbuat demikian pun sukar dilakukan mengingat kemungkinan jumlah komentar yang terlalu banyak dan penafsiran makna kalimat pada komentar yang membutuhkan waktu apabila dilakukan secara manual. Sebagai tanggapan atas hal tersebut dibutuhkan suatu pendekatan dalam perancangan sistem yang memiliki kemampuan untuk mengidentifikasi komentar mengandung pernyataan misogini atau tidak misogini secara lebih efektif. Penelitian ini menerapkan kombinasi metode IndoBERT *Embedding* sebagai *word embedding* dan algoritma *Bidirectional Long Short-Term Memory* untuk mengidentifikasi komentar yang memuat pernyataan misogini dan non misogini pada *platform* media sosial. Dalam membangun model tersebut digunakan data komentar sebanyak 4000 data yang diambil dari media sosial Instagram, YouTube, dan X pada penelitian ini. Evaluasi terhadap model yang dibangun menggunakan *Confusion Matrix* memperlihatkan nilai akurasi sebesar 90%. Berdasarkan hal tersebut dapat dikatakan bahwa gabungan metode IndoBERT *Embedding* dan *Bidirectional Long Short-Term Memory* mampu secara efektif mengidentifikasi komentar yang memuat pernyataan misogini dan komentar yang tidak memuat pernyataan misogini.

**Kata Kunci:** Misogini, *Bidirectional Long-Short Term Memory*, IndoBERT *Embedding*.

**IDENTIFICATION OF MISOGYNY STATEMENTS BASED ON SOCIAL MEDIA  
COMMENTS USING BIDIRECTIONAL LONG SHORT-TERM MEMORY  
AND INDOBERT EMBEDDING**

**ABSTRACT**

*With the anonymity provided by social media platforms, users are free to upload content or share their opinions through available comment sections. Due to this, it is quite common to encounter accounts in the comment sections that clearly express hatred towards someone, especially towards women, which can be classified as misogyny. Consequently, many women feel distressed and uncomfortable due to receiving comments that belittle, demean, and harass them. Identifying accounts that engage in such behavior is challenging, considering the potential volume of comments and the need for time-consuming manual interpretation of the underlying meaning in comments. In view of this, an approach is needed in system design that has the ability to more effectively identify comments containing misogyny or non-misogyny statements. This study applies a combination of the IndoBERT Embedding method as word embedding and the Bidirectional Long Short-Term Memory algorithm to identify misogyny and non-misogyny comments within social media platforms. The model was developed using 4000 comment samples from social media platforms including Instagram, YouTube, and X. Evaluation of the model using the Confusion Matrix showed an accuracy value of 90%. Considering this, it is feasible to determine that the combination of IndoBERT Embedding and Bidirectional Long Short-Term Memory effectively identifies comments containing misogyny statements as well as those without such statements.*

**Keywords:** Misogyny, Bidirectional Long-Short Term Memory, IndoBERT Embedding.



## DAFTAR ISI

<b>PERSETUJUAN</b>	<b>iii</b>
<b>PERNYATAAN</b>	<b>iv</b>
<b>UCAPAN TERIMA KASIH</b>	<b>v</b>
<b>ABSTRAK</b>	<b>vii</b>
<b>ABSTRACT</b>	<b>viii</b>
<b>DAFTAR ISI</b>	<b>ix</b>
<b>DAFTAR TABEL</b>	<b>xii</b>
<b>DAFTAR GAMBAR</b>	<b>xiii</b>
<b>BAB 1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Metodologi Penelitian	5
1.7 Sistematika Penulisan	6
<b>BAB 2 LANDASAN TEORI</b>	<b>7</b>
2.1 Identifikasi	7
2.2 Pernyataan Misogini	7
2.3 Media Sosial	8
2.4 <i>Long Short-Term Memory</i>	9
2.5 <i>Bidirectional Long Short-Term Memory</i>	11
2.6 <i>Word Embedding</i>	12

2.7	<i>Confusion Matrix</i>	13
2.8	Penelitian Terdahulu	14
2.9	Perbedaan Penelitian	21
<b>BAB 3 ANALISIS DAN PERANCANGAN SISTEM</b>		<b>22</b>
3.1	Data yang Digunakan	22
3.2	Arsitektur Umum	24
3.2.1.	<i>Preprocessing</i>	25
3.2.2.	<i>Word Embedding</i>	31
3.2.3.	<i>Bidirectional Long Short-Term Memory</i>	34
3.2.4.	<i>Output</i>	36
3.3	Perancangan Sistem	38
3.3.1.	<i>Desain Halaman Beranda</i>	38
3.3.2.	<i>Desain Halaman Training</i>	39
3.3.3.	<i>Desain Halaman Testing</i>	39
3.3.4.	<i>Desain Halaman User Input berupa Teks</i>	40
3.3.5.	<i>Desain Halaman User Input berupa Tautan</i>	41
3.4	Metode Evaluasi	42
<b>BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM</b>		<b>44</b>
4.1.	Implementasi Sistem	44
4.1.1.	<i>Spesifikasi Perangkat Keras dan Perangkat Lunak</i>	44
4.1.2.	<i>Penerapan Desain Antarmuka Pengguna</i>	45
4.2.	Pelatihan Model	52
4.3.	Pengujian Kinerja Model	54
4.4.	Pengukuran Kinerja Model	58
<b>BAB 5 KESIMPULAN DAN SARAN</b>		<b>61</b>
5.1.	Kesimpulan	61
5.2.	Saran	61

**DAFTAR PUSTAKA****62****LAMPIRAN****65**

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	17
Tabel 3. 1 Dataset Komentar	23
Tabel 3. 2 Pembagian <i>Train Validation Test</i>	24
Tabel 3. 3 Implementasi <i>Cleaning</i>	26
Tabel 3. 4 Implementasi <i>Case Folding</i>	27
Tabel 3. 5 Implementasi <i>Punctuation Removal</i>	27
Tabel 3. 6 Implementasi <i>Normalization</i>	28
Tabel 3. 7 Implementasi <i>Stopword Removal</i>	29
Tabel 3. 8 Implementasi <i>Tokenization</i>	31
Tabel 3. 9 Konversi <i>Token-Index</i>	31
Tabel 3. 10 Ilustrasi <i>Confusion Matrix</i>	43
Tabel 4. 1 Tabel Percobaan Parameter Model	53
Tabel 4. 2 Hasil Pengujian Kinerja Model	54
Tabel 4. 3 Perhitungan Data Mengandung Pernyataan Misogini	59
Tabel 4. 4 Perhitungan Data Tidak Mengandung Pernyataan Misogini	59
Tabel 4. 5 Hasil Perhitungan Metrik Evaluasi	60

## DAFTAR GAMBAR

Gambar 2. 1 Komentar Pernyataan Misogini	8
Gambar 2. 2 Komentar Pernyataan Misogini	8
Gambar 2. 3 Arsitektur Umum <i>Long Short-Term Memory</i>	9
Gambar 2. 4 Arsitektur <i>Bidirectional Long Short-Term Memory</i>	11
Gambar 2. 5 Arsitektur BERT	13
Gambar 3. 1 Hasil <i>Embedding</i> terhadap Satu Contoh Kalimat	32
Gambar 3. 2 Vektor <i>Embedding</i> Kata ‘Murah’ pada Kalimat Misogini	33
Gambar 3. 3 Vektor <i>Embedding</i> Kata ‘Murah’ pada Kalimat Non Misogini	33
Gambar 3. 4 <i>Flowchart Bidirectional Long Short-Term Memory</i>	34
Gambar 3. 5 Arsitektur Umum	37
Gambar 3. 6 Desain Halaman Beranda	38
Gambar 3. 7 Desain Halaman <i>Training</i>	39
Gambar 3. 8 Desain Halaman <i>Testing</i>	40
Gambar 3. 9 Desain Halaman <i>Text Input</i>	41
Gambar 3. 10 Desain Halaman YouTube <i>Link Input</i>	42
Gambar 4. 1 Tata Muka Halaman Beranda	45
Gambar 4. 2 Tata Muka Halaman <i>Training</i>	46
Gambar 4. 3 Tata Muka Hasil <i>Training</i>	46
Gambar 4. 4 Tata Muka Grafik <i>Loss</i> dan <i>Accuracy</i>	47
Gambar 4. 5 Tata Muka Halaman <i>Testing</i>	47
Gambar 4. 6 Tata Muka Halaman Hasil Proses <i>Testing</i> (Keseluruhan)	48
Gambar 4. 7 Tata Muka Halaman Hasil Proses <i>Testing</i> (Salah Identifikasi)	48
Gambar 4. 8 Tata Muka Evaluasi Model	49
Gambar 4. 9 Tata Muka Halaman <i>Text Input</i>	49
Gambar 4. 10 Tata Muka Hasil Identifikasi <i>Text Input</i> (Non Misogini)	50
Gambar 4. 11 Tata Muka Hasil Identifikasi <i>Text Input</i> (Misogini)	50
Gambar 4. 12 Tata Muka Halaman <i>Link Input</i>	51
Gambar 4. 13 Tata Muka Hasil <i>Crawl</i> dan Identifikasi dari <i>Link Input</i>	51
Gambar 4. 14 Tata Muka <i>Bar Chart</i> dan Tombol Unduh	52



Gambar 4. 15 Tata Muka Isi Unduhan <i>File</i>	52
Gambar 4. 16 Tata Muka Grafik <i>Loss</i> dan <i>Accuracy Model</i>	54
Gambar 4. 17 <i>Confusion Matrix</i>	59



## **BAB 1**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Hingga saat ini, sering ditemukan masyarakat yang menganut paham patriarki di Indonesia. Hal ini membuat misogini masih terus berkembang secara luas seiring dengan perkembangan ideologi tersebut. Misogini sendiri memiliki arti kebencian terhadap perempuan dan pemahaman ini telah melahirkan seorang misogynis. Seorang misogynis memiliki sikap yang memandang bahwa perempuan merupakan pihak yang layak untuk disudutkan, ditindas, dan dieksploitasi, serta mendapat kebencian. Selain hal ini, terdapat juga kebiasaan yang meremehkan dan menomorduakan perempuan, menghina dan mendiskriminasi perempuan, serta menganggap bahwa perempuan layak dilecehkan dan diberi kekerasan (Lubis, 2021; Maharani, 2022). Tentunya, perilaku ini sudah termasuk kedalam tindakan kekerasan terhadap perempuan dan telah melanggar prinsip HAM terkait kesetaraan dan nondiskriminasi. Meskipun demikian, perilaku ini masih tersebar luas dan bahkan dapat ditemukan di media sosial.

Penggunaan media sosial yang identik dilakukan secara anonim mengakibatkan sering kali ditemukan pernyataan atau unggahan yang tidak berkenan terhadap sesama masyarakat terutama terhadap perempuan. Hal tersebut diperkuat pernyataannya berdasarkan laporan oleh *National Democratic Institute* (National Democratic Institute, 2019), dimana dikatakan bahwa perempuan-perempuan di Kolombia, Kenya, dan Indonesia menjadi yang paling umum mengalami kekerasan secara *online* berupa bahasa yang menghina. Dalam Catatan Tahunan (CATAHU) 2023 (Fuad et al., 2023), Komisi Nasional Anti Kekerasan terhadap Perempuan (Komnas Perempuan) mendapatkan pengaduan kasus Kekerasan Siber Berbasis Gender (KSBG) sebanyak 1.697 kasus yang dimana 869 kasus diantaranya terjadi di ranah publik. Di antara kasus-kasus tersebut, terdapat 383 pelaku merupakan teman korban di sosial media, sementara 304 pelaku merupakan orang asing bagi korban. Selain Komnas Perempuan, terdapat juga catatan pengaduan dari Lembaga Layanan yang menyatakan sebanyak 148 kasus KSBG diantara semua kasus yang diadukan. Banyaknya kasus yang terjadi ini dapat dan telah menimbulkan rasa tidak aman terhadap perempuan dalam menggunakan

teknologi seperti halnya media sosial. Bahkan, beberapa perempuan diantaranya menghentikan sementara aktivitas mereka di media sosial dan terdapat juga yang berhenti secara permanen menggunakan media sosial (National Democratic Institute, 2019).

Hal terkait kekerasan melalui jejaring sosial ini tentunya perlu diamati dan dicermati lebih lanjut karena telah merugikan perempuan, terutama terkait misogini yang menjadi salah satu bentuknya. Akan tetapi, untuk membuktikan suatu pernyataan merupakan misogini atau bukan, diperlukan ketelitian dan pemahaman terhadap setiap konteks kalimat yang ada pada komentar. Tentunya, hal tersebut akan memerlukan waktu yang tidak sedikit untuk membuktikan pernyataannya satu per satu. Sehingga diperlukan suatu pendekatan yang dapat mempermudah proses tersebut.

Pendekatan yang dapat dilakukan ialah dengan melakukan identifikasi pernyataan misogini menggunakan metode pembelajaran mesin. Sebelumnya telah dilakukan beberapa penelitian yang serupa, diantaranya studi dengan judul “*Identifikasi Pernyataan Misogini Berbahasa Indonesia Berdasarkan Komentar Youtube Menggunakan GloVe Embedding dan Random Forest Classifier*” (Damanik, 2021). Penelitian tersebut berupa skripsi yang melakukan identifikasi pernyataan misogini yang diinput dalam bentuk dokumen menggunakan algoritma *Random Forest (RF)* dan *GloVe Embedding* dengan hasil akurasi 92,5%. Penelitian lainnya berjudul “*Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings*” (Angeline et al., 2022). Studi tersebut melakukan pendeteksian cuitan Twitter terkait misogini dengan menggunakan *Logistic Regression (LR)*, *Convolutional Neural Network (CNN)*, dan *Long Short-Term Memory (LSTM)*, serta TF-IDF, Keras, dan BERT sebagai *embedding*. Hasil penelitian ini mendapatkan akurasi terbaik sebesar 86,15% melalui LSTM dengan bantuan BERT. Terdapat juga penelitian yang berjudul “*Misogynous Text Classification Using SVM and LSTM*” (Devi & Saharia, 2020). Penelitian ini melakukan klasifikasi pernyataan termasuk misogini atau bukan dalam dataset berbahasa India dan Inggris menggunakan *Support Vector Machine (SVM)* dengan ekstraksi TF-IDF, SVM dengan bantuan N-gram, *Long Short-Term Memory (LSTM)*, *Bidirectional LSTM (Bi-LSTM)*, dan LSTM dengan *GloVe Embedding*. Hasil dari penelitian ini memperlihatkan akurasi terbaik untuk dataset berbahasa India diraih oleh SVM dengan TF-IDF sebesar 87,1% dan Bi-LSTM sebesar 87%, sedangkan untuk dataset berbahasa Inggris diraih oleh Bi-LSTM sebesar 93,4% dan LSTM dengan *GloVe* sebesar 93,3%.

Selanjutnya, terdapat penelitian yang berjudul “*Identification of Misogyny on Social Media in Indonesia Using Bidirectional Encoder Representations From Transformers (BERT)*” (Tri Wibowo et al., 2023). Dalam penelitian ini, dilakukan identifikasi pernyataan misogini terhadap komentar berbahasa Indonesia yang diambil dari media Instagram dan YouTube. Penelitian ini menggunakan IndoBERT sebagai model dan menghasilkan akurasi sebesar 83.74%. Kemudian, terdapat juga penelitian yang berjudul “*Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets*” (Faisal & Mahendra, 2022). Penelitian dilakukan menggunakan *traditional machine learning* (NB, SVM, LR, DT, RF, dan XGB) dengan ekstraksi fitur (BoW, TF-IDF, Word2Vec, dan IndoBERT), serta penelitian juga dilakukan menggunakan *deep learning* (BERT, Bi-LSTM, DNN, dan CNN) dengan IndoBERT *Embedding*. Hasil penelitian menyatakan bahwa akurasi terbaik diraih oleh Bi-LSTM dengan bantuan IndoBERT sebesar 87,02%.

Ditemukan juga penelitian berjudul “*Sentiment Analysis using Bidirectional LSTM Network*” (Mahadevaswamy & Swathi, 2023). Penelitian ini melakukan analisis sentimen terhadap 102.975 ulasan dari produk *Mobile Electronic* pada *platform Amazon* dengan hanya menggunakan data teks ulasan dan *rating* bintang sebagai data. Pada penelitian ini, algoritma Bi-LSTM digunakan untuk melakukan analisis sentimen dan menghasilkan akurasi sebesar 91,4%. Selain itu, terdapat penelitian lain yang berjudul “*Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection*” (Isnain et al., 2020). Penelitian ini melakukan deteksi terhadap ujaran kebencian berbahasa Indonesia dengan menggunakan LSTM dan Bi-LSTM sebagai algoritma, serta *Continuous bag-of-word* (CBOW) dan *skipgram* sebagai ekstraksi fitur. Akurasi tertinggi yang dicapai dalam penelitian ini ialah senilai 96,93% dengan penerapan Bi-LSTM tiga *layer*.

Dengan mempertimbangkan latar belakang tersebut, penulis memutuskan untuk menggunakan algoritma *Bidirectional Long Short-Term Memory* dengan bantuan IndoBERT *Embedding* untuk mengembangkan sebuah sistem berbasis web yang dapat melakukan identifikasi terhadap pernyataan-pernyataan misogini yang tersebar di media sosial. Dengan demikian, penulis memberikan judul penelitian ini berupa “IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN KOMENTAR MEDIA SOSIAL MENGGUNAKAN BIDIRECTIONAL LONG SHORT-TERM MEMORY DAN INDOBERT EMBEDDING”.



## 1.2 Rumusan Masalah

Kebebasan yang diberikan terhadap pengguna media sosial melalui karakteristik anonimitas dan dalam mengunggah konten telah menjadi salah satu faktor maraknya unggahan yang berisikan hal-hal negatif, termasuk komentar misogini. Dan dalam menyikapi komentar-komentar tersebut, perempuan harus meluangkan banyak waktu untuk dapat memblokir atau melaporkan pengguna yang melakukan pelecehan karena harus memilah mana pernyataan pengguna yang termasuk misogini dan tidak misogini. Hal ini tentunya menyebabkan masyarakat khususnya perempuan merasa cemas dan tidak merasa aman berada di dunia maya maupun dunia nyata. Serta, dapat memberikan dampak negatif terhadap psikis perempuan yang menjadi korban. Dengan demikian, diperlukan pemodelan yang dapat mengidentifikasi suatu komentar mengandung pernyataan misogini atau non misogini tanpa intervensi manual.

## 1.3 Tujuan Penelitian

Adapun maksud dari penelitian ialah untuk melakukan identifikasi pernyataan yang termasuk kedalam misogini dan tidak termasuk misogini pada komentar di media sosial dengan memanfaatkan algoritma *Bidirectional Long Short-Term Memory* dan metode *IndoBERT Embedding*.

## 1.4 Batasan Penelitian

Sebagai upaya dalam menghindari terjadinya penyimpangan dan meluasnya lingkup permasalahan, maka penulis menetapkan beberapa hal yang dibatasi dalam penelitian sebagai berikut.

1. Data yang diterapkan berupa komentar yang menggunakan bahasa Indonesia.
2. Data *training* yang digunakan merupakan hasil *scrapping* dan *crawling* dari media sosial Instagram, YouTube, dan X.
3. Data yang digunakan berupa teks pada dokumen berekstensi .csv
4. Identifikasi komentar hanya dilakukan pada kata dan tidak melibatkan pemakaian *hashtag*, *username*, URL, dan emoji.
5. *Output* dari penelitian berupa hasil identifikasi komentar yang diekstraksi dari *link* postingan media sosial YouTube.



## 1.5 Manfaat Penelitian

Manfaat yang diharapkan melalui penelitian ini di antaranya:

1. Memberi bantuan dalam mengatasi isu terkait kebencian terhadap perempuan.
2. Memberikan informasi terhadap pengguna media sosial terkait komentar yang mengandung pernyataan misogini atau non misogini berdasarkan pengidentifikasian yang dilakukan.

## 1.6 Metodologi Penelitian

Beberapa tahapan yang dikerjakan ialah sebagaimana berikut ini.

### 1. Studi Literatur

Sebagai tahapan awal dalam penelitian, studi literatur melakukan proses pengumpulan data dan referensi dari berbagai artikel, buku, jurnal, skripsi, serta sumber-sumber lain yang terkait dengan identifikasi, *text processing*, *Bidirectional Long Short-Term Memory*, *IndoBERT Embedding*, media sosial, dan pernyataan yang bersifat misogini.

### 2. Analisis Permasalahan

Tahapan melakukan penelaahan terhadap literatur terdahulu untuk dapat memahami metode *Bidirectional Long Short-Term Memory* yang akan diintegrasikan dengan *IndoBERT Embedding* dalam mengidentifikasi pernyataan misogini.

### 3. Perancangan Sistem

Dalam tahapan ini, akan dilaksanakan rancangan dataset, desain antarmuka sistem, dan arsitektur umum yang akan diterapkan dalam penelitian berlandaskan studi literatur dan analisis permasalahan yang telah dilaksanakan sebelumnya.

### 4. Implementasi

Dalam tahapan ini, seluruh hasil perancangan sistem akan diterapkan untuk membangun sistem yang sesuai dengan tujuan penelitian.

### 5. Pengujian Sistem

Dalam tahapan ini, dilaksanakan uji coba sistem yang telah dikembangkan untuk memastikan rancangan sistem dan mengetahui tingkat akurasi yang dapat diraih oleh *Bidirectional Long Short-Term Memory* dan *IndoBERT Embedding*.

## 6. Penyusunan Laporan

Sebagai tahap akhir, penyusunan laporan bertujuan untuk menyusun dokumen terkait dengan keseluruhan proses dan hasil dari penelitian yang telah dilaksanakan.

### 1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini dimuat dalam lima bagian utama, antara lain:

#### **Bab 1: Pendahuluan**

Sebagai bagian pembuka atau pendahulu sebuah penelitian, bagian ini berisikan latar belakang penelitian yang dilakukan, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

#### **Bab 2: Landasan Teori**

Landasan teori menguraikan berbagai teori yang menjadi dasar dalam meningkatkan pemahaman terkait penelitian dan agar mampu menyelesaikan masalah, serta mencapai tujuan penelitian. Bagian ini akan menguraikan mengenai pernyataan misogini, media sosial, serta metode yang digunakan yaitu *Bidirectional Long Short-Term Memory* dan *IndoBERT Embedding*. Selain itu, akan diuraikan juga penelitian sebelumnya yang dijadikan sebagai pedoman dalam penelitian ini, serta teori terkait metode evaluasi *Confusion Matrix*.

#### **Bab 3: Analisis dan Perancangan Sistem**

Bagian ini menguraikan analisis masalah penelitian dan perancangan sistem identifikasi misogini berdasarkan komentar media sosial dengan menggunakan *Bidirectional Long Short-Term Memory* dan *IndoBERT Embedding*.

#### **Bab 4: Implementasi dan Pengujian Sistem**

Bagian ini menguraikan proses penerapan sistem yang telah didesain sebelumnya. Serta, menguraikan hasil pengujian sistem dan pengevaluasian terhadap sistem.

#### **Bab 5: Kesimpulan dan Saran**

Sebagai bagian terakhir dari penelitian, bagian ini menyajikan kesimpulan berdasarkan penelitian yang telah diselesaikan dan saran untuk penelitian selanjutnya yang dipaparkan oleh penulis setelah melihat hasil dalam penelitian ini sehingga penelitian di masa depan dapat lebih berkembang.

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 Identifikasi**

Identifikasi dalam KBBI berarti penentu atau penetapan identitas seseorang, benda, dan sebagainya. Secara umum, identifikasi dapat pula dipahami sebagai tindakan yang mengenali, meneliti, dan menemukan serta mencatat informasi guna menentukan atau membedakan suatu entitas dengan entitas yang lain (Nuryanto, n.d.). Dan dalam *machine learning*, identifikasi dapat diartikan sebagai tindakan dimana sistem yang dibangun menggunakan algoritma tertentu menghasilkan model yang mampu mengenali dan menetapkan sesuatu sesuai dengan pola atau kelas yang telah dipelajari dari data (MinTek, 2023).

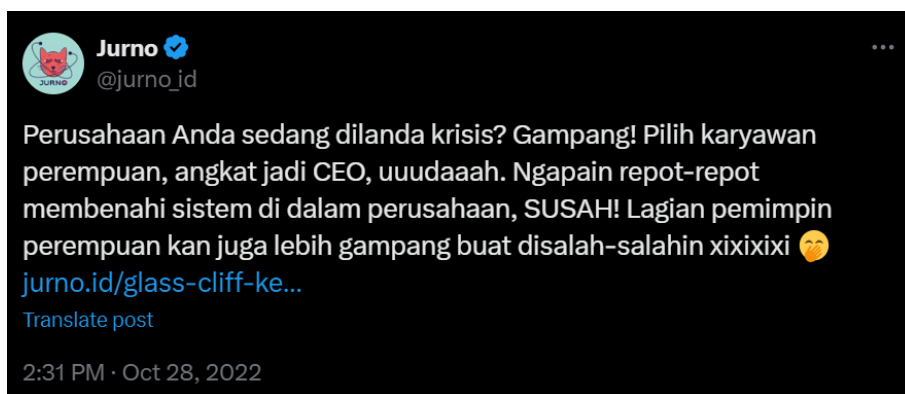
#### **2.2 Pernyataan Misogini**

Misogini terbentuk dari dua kata dalam bahasa Yunani, yaitu “*misein*” yang mengandung makna membenci dan “*gynē*” yang mengandung makna perempuan (Merriam-Webster, n.d.). Sehingga dapat diartikan bahwa misogini merupakan kebencian terhadap perempuan. Adapun hal yang dapat menggambarkan misogini ini ialah tindakan meremehkan perempuan, kekerasan terhadap perempuan, pelecehan seksual, objektifikasi seksual, diskriminasi perempuan, patriarki, dan *male privilege* (Srivastava et al., 2017). Dimana biasanya, tindakan-tindakan tersebut dialami oleh perempuan yang dianggap menentang norma, budaya, dan standar yang beredar di lingkungan masyarakat (Pertiwi, 2023). Tindakan misogini ini sendiri dapat ditemukan baik dalam kehidupan *virtual* atau pun dunia fisik.

Secara nyata atau fisik, tindakan misogini dapat berupa kekerasan fisik terhadap perempuan, melecehkan perempuan melalui tindakan fisik atau pun perkataan/*catcalling*, merendahkan dan mengucilkan perempuan di lingkungan bermasyarakat (*What Is Misogyny?*, n.d.). Bahkan, rasa kebencian yang berlebih tersebut dapat menjadi salah satu alasan terjadinya pembunuhan terhadap perempuan

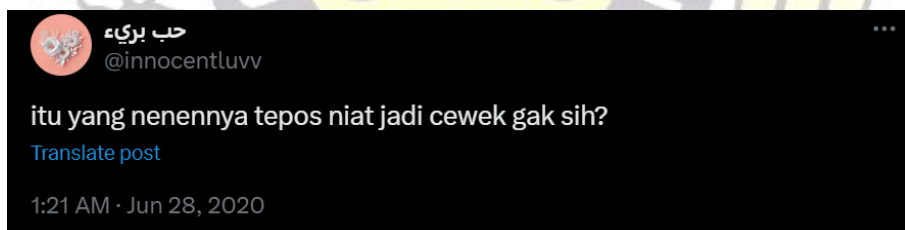
atau yang disebut juga dengan femisida (Hidayah, 2022). Secara *online* atau dunia maya, tindakan misogini dapat berupa perlisian konten video atau gambar tanpa persetujuan dengan maksud merendahkan dan menghina perempuan bersangkutan, serta adanya kolom komentar yang berisikan kalimat cercaan dan melecehkan perempuan di sosial media (*What Is Misogyny?*, n.d.). Kalimat pada komentar tersebut yang disebut dengan pernyataan misogini dan yang akan diteliti pada penelitian ini.

Berikut contoh komentar yang mengandung pernyataan misogini tercantum dalam Gambar 2.1 dan Gambar 2.2.



**Gambar 2. 1** Komentar Pernyataan Misogini

Komentar pada Gambar 2.1, dikatakan sebagai pernyataan misogini karena mengandung makna yang meremehkan perempuan.



**Gambar 2. 2** Komentar Pernyataan Misogini

Komentar pada Gambar 2.2, dikatakan sebagai pernyataan misogini karena mengandung makna yang melecehkan perempuan dan objektifikasi seksual.

## 2.3 Media Sosial

Media sosial adalah ruang *digital* yang memberikan kesempatan terhadap penggunaannya untuk melakukan interaksi sosial dan memberi serta menerima informasi atau pun konten hiburan dimana pun para pengguna tersebut selama masih ada jaringan internet. Kehadiran media sosial ini tentunya menjadi marak dikalangan masyarakat seiring perkembangan zaman, terlepas dari dampak positif maupun negatifnya. Terlebih lagi,

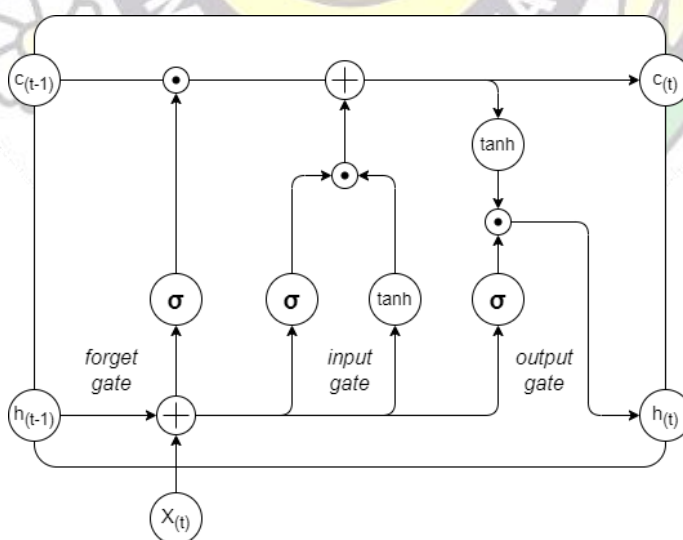


media sosial memberikan penggunanya kebebasan untuk mengunggah konten apa pun dan dengan bebas juga dapat menyampaikan pendapat atau komentar secara anonim di media sosial.

Di Indonesia sendiri per Januari 2023 ditemukan sebanyak 167 juta pengguna media sosial, dimana 46,8% diantaranya ialah perempuan dan 53,2% diantaranya ialah laki-laki. Dengan media sosial yang termasuk paling dominan digemari di Indonesia ialah Instagram dan Twitter (X). Kemudian, terdapat YouTube yang termasuk sebagai media sosial dengan jumlah pengguna dan waktu penggunaannya paling banyak di Indonesia (Kemp, 2023). Oleh karena hal tersebut, pada penelitian ini akan digunakan media sosial Instagram, X, dan YouTube sebagai sumber untuk mendapatkan data komentar.

## 2.4 Long Short-Term Memory

*Long Short-Term Memory* (LSTM) merupakan variasi dari algoritma *Recurrent Neural Network* (RNN) yang memiliki keunggulan dapat mempertahankan informasi dalam jangka lebih panjang tanpa mengalami *error backflow* (Angeline et al., 2022). Hal tersebut dapat dilakukan oleh LSTM karena menambahkan CEC (*Constant Error Carousels*) dengan *input* dan *output gate* yang membentuk *memory cell* (Van Houdt et al., 2020). Adapun arsitektur umum dari *Long Short-Term Memory* dapat dilihat melalui Gambar 2.3.



**Gambar 2. 3** Arsitektur Umum *Long Short-Term Memory*

Dalam algoritma *Long Short-Term Memory*, *memory cell* bertindak sebagai akumulator informasi status yang pengoperasiannya dikontrol oleh tiga *gate*, yaitu *input gate*, *forget gate*, dan *output gate*. *Forget gate* menjadi langkah pertama yang dilakukan



pada algoritma ini. Dimana, *gate* ini yang akan menetapkan informasi apa yang diteruskan ke dalam *cell state* dengan bantuan *sigmoid function*. Untuk perhitungan *forget gate* dapat dilihat melalui persamaan 2.1.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2.1)$$

Jika *output* yang dihasilkan dari persamaan ini adalah 0, maka informasi akan dilupakan. Namun, jika *output* yang dihasilkan 1 maka informasi berhasil masuk ke dalam *cell state*. Kemudian sebagai langkah kedua, *input gate* akan menentukan informasi mana yang akan disimpan ke dalam *cell state*. *Gate* ini melakukan beberapa perhitungan, diantaranya perhitungan untuk menentukan informasi yang akan diperbaharui dengan bantuan *sigmoid function* dan perhitungan untuk menampung kandidat informasi baru ( $\bar{C}_t$ ) yang nantinya dapat ditambahkan ke dalam *cell state* dengan menggunakan *tanh*. Pengukuran tersebut dapat dilihat melalui persamaan 2.2 dan 2.3.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2.2)$$

$$\bar{C}_t = \tanh(W_i[h_{t-1}, x_t] + b_i) \quad (2.3)$$

Kemudian, nilai inputan lama atau *cell state* lama ( $C_{t-1}$ ) diperbaharui dengan nilai *cell state* baru ( $C_t$ ) yang dapat dilihat melalui persamaan 2.4.

$$C_t = f_t \odot C_{t-1} + i_t \odot \bar{C}_t \quad (2.4)$$

Setelahnya, proses berlanjut pada *output gate* yang menentukan apa yang akan dijadikan *output*. *Gate* ini melakukan perhitungan untuk menentukan bagian *cell state* yang akan dikeluarkan sebagai *output* dengan bantuan *sigmoid* dan perhitungan untuk menentukan *output* yang akan diteruskan pada *time step* selanjutnya atau disebut *hidden state* ( $h_t$ ). Dimana, bagian tersebut dihitung bersama dengan *tanh* untuk menghasilkan *range value* antara -1 dan 1. Perhitungan tersebut dapat dilihat melalui persamaan 2.5 dan 2.6.

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.5)$$

$$h_t = O_t \odot \tanh(C_t) \quad (2.6)$$

Keterangan:

$f_t$  = Forget gate

$i_t$  = Input gate

$O_t$  = Output gate

$\sigma$  = Fungsi aktivasi *sigmoid*

$h_t$  = Hidden state

$h_{t-1}$  = Hidden state sebelumnya

$x_t$  = Inputan sekarang

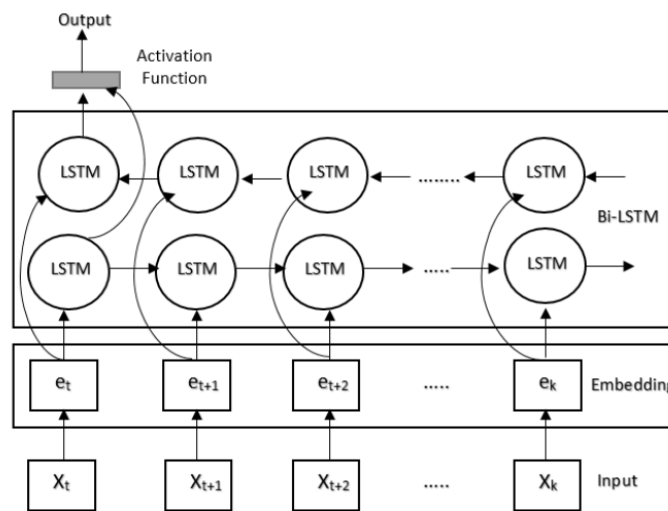
$W_f, W_i, W_o$  = parameter *weight*

$b_f, b_i, b_o$  = parameter *bias*

## 2.5 Bidirectional Long Short-Term Memory

Dalam penerapan *Long Short-Term Memory*, algoritma ini hanya dapat mengumpulkan informasi dari satu arah dengan urutan awal hingga akhir. Hal tersebut mengakibatkan mesin hanya mengetahui informasi dari masa lalu atau konteks sebelumnya. Sedangkan, untuk menganalisis teks, kata-kata lanjutan yang membentuk konteks keseluruhan kalimat mempengaruhi makna kata tersebut dan informasi tersebut tidak dapat dikelola dengan mengandalkan pemrosesan satu arah.

*Bidirectional Long-Short Term Memory* (Bi-LSTM) adalah algoritma yang mencakup dua *layer* algoritma *Long-Short Term Memory* yang memungkinkan proses dilakukan secara dua arah. Dimana, satu *layer* bertindak sebagai *forward layer* yang memodelkan konteks sebelumnya dan satu *layer* lainnya sebagai *backward layer* yang memodelkan konteks berikutnya. Kedua *layer* tersebut berkerja dengan menggunakan dua *hidden layer* terpisah yang *output*-nya akan digabungkan ke dalam *output layer* yang sama. Dimana hasil *output* tersebut menampung informasi dari masa lampau dan masa depan. Adapun arsitektur dari *Bidirectional Long Short-Term Memory* dapat dilihat melalui Gambar 2.4.



**Gambar 2. 4** Arsitektur *Bidirectional Long Short-Term Memory* (Isnain et al., 2020)

Berikut adalah perhitungan *hidden state forward* ( $\vec{h}_t$ ) dan *hidden state backward* ( $\overleftarrow{h}_t$ ), serta *output* akhirnya ( $y_t$ ) yang tercantum di persamaan 2.7, 2.8, dan 2.9 (Imrana et al., 2021).

$$\vec{h}_t = H(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}) \quad (2.7)$$

$$\vec{h}_t = H(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t+1} + b_{\vec{h}}) \quad (2.8)$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\vec{h}y}\vec{h}_t + b_y \quad (2.9)$$

Keterangan:

$W_{x\vec{h}}, W_{\vec{h}\vec{h}}, W_{\vec{h}y}$  = parameter *weight forward layer*

$W_{x\vec{h}}, W_{\vec{h}\vec{h}}, W_{\vec{h}y}$  = parameter *weight backward layer*

$b_{\vec{h}}$  = parameter *bias forward layer*

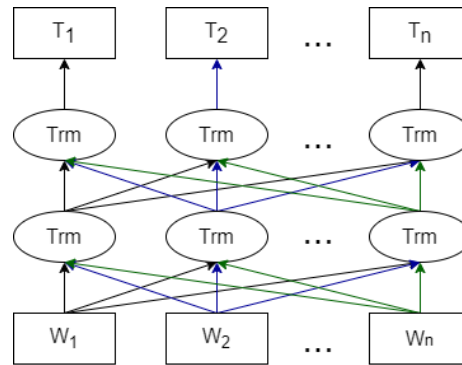
$b_{\vec{h}}$  = parameter *bias backward layer*

$b_y$  = parameter *bias output*

## 2.6 Word Embedding

*Word embedding* merupakan teknik yang merepresentasikan inputan teks menjadi vektor numerik. Teknik ini sangat diperlukan karena dapat meningkatkan pemahaman model atas data yang dikelola. *Word embedding* sendiri terdiri dari tiga kategori, yang pertama yaitu *traditional word embedding*, dimana representasi vektor dilakukan terhadap kata yang telah melalui pembobotan nilai kata berdasarkan frekuensi kemunculan kata tersebut. Kedua, *static word embedding*, dimana representasi vektor dilakukan terhadap setiap kata. Ketiga, *contextualized word embedding*, dimana representasi vektor dilakukan terhadap kata dengan memperhitungkan konteks yang ada pada kalimat (Birunda & Devi, 2020). Berdasarkan pengertian tersebut, dapat dilihat bahwa *contextualized word embedding* menghasilkan representasi vektor yang lebih baik dibandingkan lainnya karena bersifat dinamis dimana kata yang memiliki banyak arti juga turut memiliki representasi vektor yang berbeda (Angeline et al., 2022).

Salah satu yang termasuk ke dalam *contextualized word embedding* ialah BERT (*Bidirectional Encoder Representations from Transformers*). BERT dikenal sebagai teknik penyematan kata yang memanfaatkan pendekatan dua arah untuk memahami hubungan antar kata dalam dokumen sehingga dapat menangkap makna semantik yang lebih dalam secara efektif (Wang et al., 2019). Berikut arsitektur dari BERT yang tercantum di Gambar 2.5.



**Gambar 2. 5** Arsitektur BERT (Wang et al., 2019)

Berdasarkan survei, BERT juga dikatakan sebagai *word embedding* terbaik yang dapat melakukan berbagai tugas *natural language processing*, seperti pengklasifikasian teks, prediksi kalimat berikutnya, dan penemuan kesamaan antar teks (Birunda & Devi, 2020). Oleh karena itu, penelitian ini menggunakan BERT atau lebih spesifiknya IndoBERT sebagai *word embedding*. Dimana, IndoBERT merupakan model BERT yang dilatih menggunakan situs *online*, *blog*, artikel berita, dan teks media sosial yang berbahasa Indonesia. Serta, dapat dikatakan bahwa IndoBERT merupakan model BERT versi Bahasa Indonesia (Faisal & Mahendra, 2022).

## 2.7 Confusion Matrix

*Confusion Matrix* merupakan matriks dua dimensi yang setiap kolomnya berisi jumlah data yang diprediksi oleh model dan baris berisi jumlah data sebenarnya untuk setiap kelas. *Confusion Matrix* mampu menggambarkan dan menganalisis distribusi serta tumpang tindih antara label yang diprediksi salah dan benar dalam satu tampilan (Heydarian et al., 2022). Sehingga dapat dikatakan bahwa *Confusion Matrix* merupakan rangkuman terkait kinerja model dalam melakukan prediksi terhadap data uji yang labelnya sudah diketahui.

Dalam melakukan perhitungan kinerja klasifikasi model, nilai yang diambil berupa nilai *True Positive*, *False Positive*, *True Negative*, dan *False Negative* dari *Confusion Matrix*. Dimana selanjutnya, nilai tersebut digunakan pada perhitungan metrik *precision*, *recall*, *F-1 score*, dan *accuracy* untuk menghitung kinerja model. *Accuracy* adalah metrik yang menunjukkan seberapa banyak model melakukan klasifikasi yang tepat terhadap data. *Precision* adalah perhitungan yang bertujuan untuk menunjukkan seberapa akurat model dalam memberikan hasil yang relevan dari semua hasil yang diklasifikasi. *Recall* adalah metrik yang bertujuan untuk menunjukkan seberapa baik



model menemukan data yang benar dari semua data yang seharusnya diklasifikasi. *F-1 Score* adalah perbandingan yang dilakukan terhadap *precision* dan *recall* yang bertujuan untuk mengetahui rata-rata keduanya (Angeline et al., 2022). Berikut adalah persamaan sistematis untuk menentukan metrik *accuracy*, *precision*, *recall*, dan *f1-score* yang telah disebutkan.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (2.10)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.11)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.12)$$

$$F1\ Score = 2 \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (2.13)$$

## 2.8 Penelitian Terdahulu

Terdapat berbagai penelitian sebelumnya yang telah melakukan identifikasi pernyataan komentar misogini dan juga yang memanfaatkan *Bidirectional Long Short-Term Memory* sebagai *classifier*, serta pemanfaatan IndoBERT. Penelitian tersebut diantaranya ialah penelitian yang berjudul “*Misogynous Text Classification Using SVM and LSTM*” (Devi & Saharia, 2020). Penelitian ini melakukan klasifikasi pernyataan termasuk misogini atau bukan dalam dataset berbahasa India dan Inggris yang diambil dari komentar YouTube menggunakan *Support Vector Machine* (SVM) dengan ekstraksi *Term Frequency – Inverse Document Frequency* (TF-IDF), SVM dengan bantuan N-gram, *Long Short-Term Memory* (LSTM), *Bidirectional Long Short-Term Memory* (Bi-LSTM), dan LSTM dengan *GloVe Embedding*. Hasil dari penelitian ini memperlihatkan akurasi terbaik untuk dataset berbahasa India diraih oleh SVM dengan TF-IDF sebesar 87,1% dan Bi-LSTM sebesar 87%, sedangkan untuk dataset berbahasa Inggris diraih oleh Bi-LSTM sebesar 93,4% dan LSTM dengan *GloVe* sebesar 93,3%.

Selain itu, terdapat penelitian yang berjudul “*Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection*” (Isnain et al., 2020). Penelitian tersebut melakukan deteksi terhadap ujaran kebencian berbahasa Indonesia di Twitter yang datasetnya diambil dari penelitian (Alfina et al., 2017). Deteksi dilakukan menggunakan algoritma *Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short-Term Memory* (Bi-LSTM), serta *Continuous bag-of-word* (CBOW) dan *skipgram* sebagai ekstraksi fitur dari *Word2vec*. Penelitian ini mengelola data terlebih dahulu dengan ekstraksi fitur dan mendapatkan akurasi terbaik sebesar 93%



dari penggunaan CBOW. Kemudian, penelitian ini juga melakukan *testing* untuk menemukan jumlah *neuron* dan jumlah *epoch* yang terbaik untuk digunakan, serta melakukan *L2 regularization testing*. Setelahnya, penelitian ini melakukan identifikasi menggunakan LSTM, Bi-LSTM dan Bi-LSTM tiga *layer*. Akurasi tertinggi yang dicapai dalam penelitian ini ialah senilai 96,93% melalui penerapan Bi-LSTM tiga *layer*.

Kemudian, terdapat penelitian berupa skripsi yang memiliki judul “*Identifikasi Pernyataan Misogini Berbahasa Indonesia Berdasarkan Komentar Youtube Menggunakan GloVe Embedding dan Random Forest Classifier*” (Damanik, 2021). Penelitian ini melakukan identifikasi pernyataan misogini dalam komentar berbahasa Indonesia yang diambil dari media sosial YouTube menggunakan algoritma *Random Forest* (RF) dan *GloVe Embedding*. Dataset yang diterapkan pada penelitian ini mencakup 600 data, dimana 400 data menjadi *data train* dan 200 data menjadi *data test*. Sistem yang dibangun pada penelitian ini meminta pengguna untuk mengunggah file .csv yang berisikan data komentar yang ingin diidentifikasi. Melalui sistem tersebut, dilakukan identifikasi terhadap 200 *data test* yang telah dikumpulkan sebelumnya dan menghasilkan akurasi sebesar 92,5%.

Penelitian lainnya berjudul “*Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings*” (Angeline et al., 2022). Penelitian tersebut melakukan pendeteksian pernyataan misogini berbahasa Indonesia pada Twitter dengan menggunakan *Logistic Regression* (LR), *Convolutional Neural Network* (CNN), dan *Long Short-Term Memory* (LSTM), serta *Term Frequency – Inverse Document Frequency* (TF-IDF), Keras, dan *Bidirectional Encoder Representation from Transformers* (BERT) sebagai *embedding*. Penelitian ini hanya melakukan *case folding*, *data cleaning*, *normalization*, dan *tokenization* dalam tahap *preprocessing*-nya. Pada penelitian ini, ditemukan bahwa pemakaian BERT sebagai ekstraksi fitur untuk LR yang merupakan salah satu *traditional machine learning* menghasilkan akurasi jauh lebih rendah dibandingkan saat LR dikombinasikan dengan TF-IDF. Namun, penelitian ini membuktikan bahwa LSTM yang merupakan *deep learning* dengan bantuan BERT menghasilkan akurasi terbaik sebesar 86,15%.

Terdapat penelitian lain yang relevan berjudul “*Identification of Misogyny on Social Media in Indonesia Using Bidirectional Encoder Representations From Transformers (BERT)*” (Tri Wibowo et al., 2023). Dalam penelitian ini, dilakukan identifikasi

pernyataan misogini atau tidak misogini terhadap komentar berbahasa Indonesia yang diambil dari media Instagram dan YouTube. Jumlah data pada penelitian ini sebesar 1.223 dan memiliki distribusi yang tidak seimbang. Untuk tahapan *preprocessing* yang diterapkan pada penelitian ini ialah *data dropping*, *case folding*, *data cleaning*, *normalization*, dan *data stemming*. Kemudian penelitian ini menggunakan IndoBERT sebagai model dengan memanfaatkan model *pre-training*, yaitu *indoBert-base-p2* dari IndoNLU. Model dalam penelitian ini menghasilkan akurasi sebesar 83.74%.

Selain itu, terdapat juga penelitian berjudul “*Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets*” (Faisal & Mahendra, 2022). Penelitian ini melakukan dua tahapan klasifikasi secara *pipeline*, yaitu tahapan pertama menyaring *tweet* yang relevan dan selanjutnya tahapan kedua mendeteksi *tweet* misinformasi. Penelitian dilakukan menggunakan *traditional machine learning* berupa *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Decision Tree* (DT), *Random Forest* (RF), dan *XGBoost* (XGB) dengan bantuan ekstraksi fitur *Bag of Word* (BoW) untuk NB dan TF-IDF, Word2Vec, juga IndoBERT untuk SVM, LR, DT, RF, dan XGB. Selain itu, penelitian ini juga menggunakan *deep learning* berupa *Bidirectional Encoder Representation from Transformers* (BERT), *Bidirectional Long Short-Term Memory* (Bi-LSTM), *Deep Neural Network* (DNN), dan *Convolutional Neural Network* (CNN) dengan ekstraksi fitur IndoBERT *Embedding*. Hasil penelitian menyatakan bahwa akurasi terbaik diraih oleh Bi-LSTM dengan bantuan IndoBERT sebesar 87,02%.

Ditemukan penelitian lain berjudul “*Sentiment Analysis using Bidirectional LSTM Network*” (Mahadevaswamy & Swathi, 2023). Penelitian ini melakukan analisis sentimen terhadap 102.975 ulasan dari produk *Mobile Electronic* pada platform Amazon dengan distribusi data sebesar 99.975 untuk pelatihan dan 5.000 untuk pengujian. Analisis sentimen dilakukan dengan hanya menggunakan data teks ulasan dan *rating* bintang. Algoritma *Bidirectional Long Short-Term Memory* digunakan dalam penelitian untuk menganalisis sentimen terhadap ulasan produk yang positif dan negatif dengan bantuan *text encoder tfds.deprecated.text.TokenTextEncoder* dari TensorFlow sebagai *word encoding*. Model dalam penelitian ini melakukan analisis terhadap 128 ulasan per *batch*-nya. Dan hasil penelitian ini menyatakan bahwa model yang dirancang memiliki akurasi sebesar 91,4%.

Tabel 2. 1 Penelitian Terdahulu

No.	Penulis	Judul	Tahun	Keterangan
1.	Maibam Debina Devi dan Navanath Saharia	<i>Misogynous Text Classification Using SVM and LSTM</i>	2020	Penelitian ini melakukan klasifikasi pernyataan misogini berbahasa Inggris dan India dengan membandingkan beberapa model algoritma beserta kombinasi ekstrasi fiturnya. Hasil penelitian ini menunjukkan bahwa akurasi terbaik untuk dataset berbahasa India diraih oleh SVM dengan TF-IDF sebesar 87,1% dan Bi-LSTM sebesar 87%, sedangkan untuk dataset berbahasa Inggris diraih oleh Bi-LSTM sebesar 93,4% dan LSTM dengan <i>GloVe</i> sebesar 93,3%.
2.	Arnesa Julia Damanik	Identifikasi Pernyataan Misogini Berbahasa Indonesia Berdasarkan Komentar Youtube Menggunakan <i>GloVe Embedding</i> dan <i>Random Forest Classifier</i>	2021	Penelitian ini melakukan identifikasi pernyataan misogini berbahasa Indonesia dengan menggunakan <i>Random Forest</i> dan <i>GloVe Embedding</i> . Penelitian ini menggunakan 600 dataset, dimana 400 untuk <i>data train</i> dan 200

Tabel 2. 1 Penelitian Terdahulu (Lanjutan)

No.	Penulis	Judul	Tahun	Keterangan
				untuk <i>data test</i> . Hasil penelitian ini menunjukkan akurasi sebesar 92,5%.
3.	Rizkyta Shainy Angeline, Dade Nurjanah, dan Hani Nurrahmi	<i>Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings</i>	2022	Penelitian ini melakukan identifikasi pernyataan misogini berbahasa Indonesia dengan membandingkan beberapa kombinasi model. Penelitian ini hanya melakukan beberapa tahap berikut <i>case folding, data cleaning, normalization</i> , dan <i>tokenization</i> dalam proses <i>preprocessing</i> . Hasil penelitian ini menunjukkan bahwa akurasi terbaik diraih oleh <i>Long Short-Term Memory</i> dengan bantuan BERT sebesar 86,15%.
4.	Bagas Tri Wibowo, Dade Nurjanah, dan Hani Nurrahmi	<i>Identification of Misogyny on Social Media in Indonesian Using Bidirectional Encoder Representations From Transformers (BERT)</i>	2023	Penelitian ini melakukan identifikasi terhadap komentar berbahasa Indonesia terkait pernyataan misogini dengan menggunakan model IndoBERT yang telah melalui tahapan



Tabel 2. 1 Penelitian Terdahulu (Lanjutan)

No.	Penulis	Judul	Tahun	Keterangan
				<i>pre-training</i> , yaitu model <i>indoBert-base-p2</i> . Hasil penelitian ini menunjukkan bahwa akurasi model sebesar 83.74%.
5.	Auliya Rahman Isnain, Agus Sihabuddin, dan Yohanes Suyanto	<i>Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection</i>	2020	Penelitian ini melakukan deteksi ujaran kebencian berbahasa Indonesia. Penelitian ini dilakukan dengan membandingkan LSTM, Bi-LSTM, dan Bi-LSTM tiga <i>layer</i> . Hasil penelitian ini menunjukkan bahwa akurasi terbaik diraih oleh <i>Bidirectional Long Short-Term Memory</i> tiga <i>layer</i> sebesar 96,93%.
6.	Douglas Raevan Faisal dan Rahmad Mahendra	<i>Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets</i>	2022	Penelitian ini melakukan deteksi misinformasi terkait COVID-19 yang dilakukan dua tahap secara <i>pipeline</i> . Penelitian ini dilakukan dengan membandingkan <i>traditional machine learning</i> beserta ekstrasi fiturnya dan <i>deep learning</i> beserta kombinasinya dengan

Tabel 2. 1 Penelitian Terdahulu (Lanjutan)

No.	Penulis	Judul	Tahun	Keterangan
				IndoBERT. Hasil penelitian ini mengindikasikan bahwa akurasi terbaik diraih oleh <i>Bidirectional Long Short-Term Memory</i> dengan bantuan IndoBERT sebesar 87,02%.
7.	U. B. Mahadevaswamy dan Swathi P.	<i>Sentiment Analysis using Bidirectional LSTM Network</i>	2023	Penelitian ini melakukan sentimen analisis dari ulasan produk kategori <i>Mobile Electronic</i> pada platform <i>Amazon</i> . Sentimen analisis dilakukan untuk mengklasifikasikan ulasan positif dan negatif dengan memanfaatkan algoritma Bi-LSTM. Berdasarkan temuan penelitian ini, <i>Bidirectional Long Short-Term Memory</i> terbukti sebagai model yang efektif dan memberikan akurasi yang baik, yakni 91.4%.

## 2.9 Perbedaan Penelitian

Adapun perbedaan yang ada antara penelitian ini dengan yang dikerjakan oleh Damanik (2021) ialah penelitian tersebut melakukan identifikasi misogini terhadap *file .csv* berisi komentar yang diunggah ke sistem, sedangkan penelitian ini akan melakukan identifikasi misogini melalui komentar-komentar yang diekstraksi dari *link* postingan yang diunggah ke dalam sistem. Selain itu, ditemukan pembeda dari penelitian yang dikerjakan oleh Devi & Saharia (2020), Angeline et al. (2022), dan (Tri Wibowo et al., 2023) yang juga membahas identifikasi misogini, yaitu penelitian ini mengintegrasikan *Bidirectional Long Short-Term Memory* sebagai model untuk mengidentifikasi dan *IndoBERT Embedding* sebagai ekstra fitur yang membantu model menganalisis dan mempelajari hubungan kontekstual antar kata dengan lebih baik.



## BAB 3

### ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Data yang Digunakan

Penelitian ini menggunakan data berupa komentar-komentar yang diambil dari media sosial Instagram, YouTube, dan X. Adapun proses pengambilan komentar melalui laman Instagram ialah dengan menggunakan *library Python* yaitu Selenium, dimana *library* ini dapat melakukan otomatisasi pemakaian sebuah situs sehingga cocok untuk melakukan *scrapping* komentar. Pada proses pengambilan komentar melalui YouTube, penelitian ini menggunakan YouTube Data API v3 yang memungkinkan dalam mengakses video, daftar putar, dan akun *channel* sehingga hal ini dimanfaatkan untuk mengambil komentar. Proses pengambilan data komentar pada Instagram dan YouTube ini dilakukan terhadap akun selebriti, *video gamer*, dan jurnalis berjenis kelamin perempuan di Indonesia, serta akun-akun yang menyinggung perempuan. Berbeda halnya dengan proses pengambilan komentar melalui Instagram dan YouTube, pengambilan komentar pada X dilakukan menggunakan beberapa kata kunci tertentu saat mengambil data dengan memanfaatkan *library tweet-harvest*. Kemudian, komentar-komentar tersebut dipilah dan dikumpulkan ke dalam satu file berformat *.csv* berdasarkan keterkaitannya dengan topik penelitian ini. Total komentar yang dipilah dan yang hendak dipakai pada penelitian ini ialah sebanyak 4.000.

Dalam dataset yang digunakan pada penelitian ini terdapat kolom *Comment* yang berisikan komentar dari media sosial dan kolom *Label* yang diisi oleh angka 0 untuk menyatakan setiap komentar non misogini dan angka 1 untuk menyatakan setiap komentar misogini. Proses *labelling* ini dilakukan oleh penulis berdasarkan kesesuaian dengan definisi yang telah disampaikan dalam Bab 2. Data yang telah diberikan label inilah yang dijadikan sebagai acuan dalam melatih sistem. Data tersebut dapat diamati melalui Tabel 3.1.



**Tabel 3. 1** Dataset Komentar

No.	Comment	Label
1	RT @tirz_fabz: Jadi intinya laki2 memang lbh dipercaya utk jd pemimpin dibandingkan perempuan	1
2	Kalau saja dulu ibumu terpapar virus pemikiran feminis Child Free niscaya kamu sekarang tak 23unggal23 Gosah sok²an Child Free padahal aslinya males ngurus anak Awowkwowkwo #feminisindonesia #childfree #ngakakkocak	1
3	Pinter bacot ngide doang nih cewe kerjaannya jajan mulu keluyuran kek lonte	1
4	@AiraNtieReal Artis yg sedikit job nya kalo gak jdi lonte pejabat gak bakal 23ungga apartemen dan mobil.. Pejabat kalo gak ngelonte gak afdol	1
5	@RadioElshinta Supporter bola club Indonesia yg begini emg lahir dari 23ungg lonte yg memeknya dipake ribuan kontrol Pas lahir dulu daripada dibesarin mending kasih ke lele	1
6	Pengen ku kenytot boleh kah??	1
7	ewe lah cok, ngapain direkam rekam	1
8	Apa lagi... bungkus lalu kentod... siapa suruh bodoh jdi cwek	1
.	.	.
.	.	.
.	.	.
3993	Artis yang gak pernah di gosipin aneh²	0
3994	“Jadi buat para perempuan. Puas puasin jalani hidup, jangan buru buru nikah. Penuhi dulu mimpi kalian, kata mama gue Menikah itu saat kamu sudah bahagia sendiri jadi saat menikah kalian menikah bukan mencari kebahagiaan tetapi untuk melengkapi.”	0
3995	Duh gw kalo lagi kek org stress gini suka pengen kosple semi lonte gitu	0
3996	hidup gue kalo abis nolak cowo kalo ga di ancem ya di katain lonte <a href="https://t.co/CETZTMrpO5">https://t.co/CETZTMrpO5</a>	0

**Tabel 3. 1** Dataset Komentar (Lanjutan)

No.	Comment	Label
3997	<p>“Semua hal didunia ini bisa dilakukan oleh setiap individu tidak hanya perempuan saja atau laki laki saja</p> <p>Stereotype kaya gini harus di hapus</p> <p>Yang jadi pembeda adalah perempuan mempunyai tugas untuk melahirkan dan laki laki memproduksi nah ini ga bisa di wakikan oleh gender lain sekian...”</p>	0
3998	Yah saya tau, pertemanan laki-laki dan perempuan sulit sekali tanpa melibatkan perasaan lebih.	0
3999	6:53 kang sange tapi ekonomi nya kelas bawah, dasar abang” gaje	0
4000	Gimana sih rasanya nonton bokep langsung di depan pemainnya?	0

Dataset yang ditampilkan dalam Tabel 3.1 kemudian dibagi dan dijadikan sebagai data *training* dan data *testing* dengan rasio 80:20. Dimana 80% dari dataset merupakan bagian data *training* dan 20% merupakan data *testing*. Kemudian data *training* dipecahkan menjadi data *training* utama dan data *validation* dengan rasio 87.5:12.5. Dimana 87.5% dataset merupakan data *training* utama dan 12.5% merupakan data *validation*. Sehingga secara menyeluruh dapat dikatakan pembagian dataset ini memiliki rasio data *train*, data *validation*, dan data *test* sebesar 70:10:20. Adapun perincian pemecahan dataset ini tercantum di Tabel 3.2.

**Tabel 3. 2** Pembagian *Train Validation Test*

Data <i>Training</i>		Data <i>Testing</i>
Data <i>Training</i> Utama	Data <i>Validation</i>	
2800	400	800

### 3.2 Arsitektur Umum

Dalam penelitian ini penulis melakukan serangkaian proses yang di antaranya ialah pengumpulan data, dimana data yang terkumpul akan disimpan ke dalam file berekstensi *.csv* serta akan dimanfaatkan sebagai bahan untuk pengembangan sistem. Selanjutnya, pemecahan data, dimana dataset yang ada terbagi ke dalam tiga bagian yakni data *training*, data *validation*, dan data *testing*. Data *training* dilakukan untuk

mengasah model yang digunakan dalam menganalisis data. Data *validation* dilakukan untuk mengevaluasi kinerja model selama proses pelatihan. Data *testing* dilakukan untuk menguji model. Kemudian, terdapat tahapan *preprocessing* setelah data berhasil dikumpulkan ke dalam tiga bagian. Tahapan ini meliputi *cleaning*, *case folding*, *punctuation removal*, *normalization*, *stopword removal*, dan *tokenization*. Setelah tahapan *preprocessing* ini selesai, tahapan dilanjutkan dengan melakukan representasi kata menjadi vektor menggunakan IndoBERT *Embedding*. Lalu, hasil dari representasi vektor tersebut dijadikan sebagai *input-an* untuk *training* model *Bidirectional Long Short-Term Memory* yang akan menghasilkan suatu pemodelan yang nantinya akan dipakai saat proses *testing*. Hasil akhir dari penelitian ini merupakan pengidentifikasian terhadap pernyataan komentar yang akan diduga sebagai misogini dan non misogini. Berikut adalah penjelasan terkait tahapan-tahapan tersebut.

### 3.2.1. Preprocessing

*Preprocessing* merupakan tahapan penting dalam *Natural Language Processing* (NLP) yang melibatkan pembersihan *noise* dan inkonsistensi data, serta transformasi data tidak terstruktur menjadi lebih rapi dan terstruktur sehingga dapat mempermudah proses analisis dan meningkatkan akurasi model. Adapun dalam penelitian ini tahapan *text preprocessing* yang dilakukan diantaranya ialah *cleaning*, *case folding*, *punctuation removal*, *normalization*, *stop word removal*, dan *tokenization*.

#### 3.2.1.1. Cleaning

Pada tahapan ini, dilakukan penghapusan data komentar yang tidak relevan dalam proses identifikasi. Penghapusan tersebut akan dilakukan terhadap emoji, *hashtag*, URL, *username*, atau pun hal lainnya yang tidak diperlukan dalam penelitian. Sehingga dapat meningkatkan keberhasilan identifikasi terhadap suatu pernyataan komentar. Adapun tahapan *cleaning* yang dilakukan terhadap penelitian ini dapat dilihat berdasarkan *pseudocode* yang tertera di bagian bawah ini dengan hasil *cleaning* yang ditampilkan pada Tabel 3.3.

FOR EACH row IN dataframe:

# Menghapus *Username*

row['Comment'] = REPLACE\_REGEX(row['Comment'], '@[\w]+', '')

# Menghapus *Hashtag*

row['Comment'] = REPLACE\_REGEX(row['Comment'], '#[\w]+', '')

```

# Menghapus Non-ASCII
row['Comment'] = ENCODE_TO_ASCII_DECODE(row['Comment'])

# Menghapus Angka
row['Comment'] = REPLACE_REGEX(row['Comment'], '[\d]+', '')

# Menghapus Karakter Spesial (&,>,CR)
row['Comment'] = REPLACE_REGEX(row['Comment'], '&am|&gt;|\\r', '')

# Menghapus Spasi Ekstra
row['Comment'] = REPLACE_REGEX(row['Comment'], '\s+', ' ')

END FOR

```

Adapun penerapan *cleaning* dapat dilakukan dengan *regex (regular expression) function* yang berguna untuk mencari dan memanipulasi data komentar sesuai dengan informasi yang ditetapkan. Seperti yang tertera pada *pseudocode* di atas, untuk menghapus *hashtag* di setiap karakter yang dimulai dengan tagar '#' dan diikuti oleh karakter lain setelahnya akan dideteksi dan dihapuskan.

**Tabel 3. 3 Implementasi *Cleaning***

Sebelum Implementasi <i>Cleaning</i>	Setelah Implementasi <i>Cleaning</i>
<p>Perusahaan Anda sedang dilanda krisis? Gampang! Pilih karyawan perempuan angkat jadi CEO uuudaaah. Ngapain repot-repot membenahi sistem di dalam perusahaan SUSAH! Lagian pemimpin perempuan kan juga lebih gampang buat disalah-salahin xixixixi</p> <p><a href="https://t.co/czv3sNvgFc">https://t.co/czv3sNvgFc</a></p>	<p>Perusahaan Anda sedang dilanda krisis? Gampang! Pilih karyawan perempuan angkat jadi CEO uuudaaah. Ngapain repot-repot membenahi sistem di dalam perusahaan SUSAH! Lagian pemimpin perempuan kan juga lebih gampang buat disalah-salahin xixixixi</p>

### 3.2.1.2. Case Folding

Tahapan ini melakukan perubahan seluruh karakter huruf besar yang ada pada data menjadi karakter berhuruf kecil dengan tujuan menyederhanakan karakter pada teks. Dengan melalui tahapan ini pula dimensi data dapat berkurang, karena kata yang sama dengan penulisan huruf besar atau kecil yang berbeda telah dijadikan satu entitas yang sama dengan penyeragaman karakter ini. Adapun tahapan *case folding* ini dapat dilakukan dengan menggunakan *function String lower()*. Hasil proses ini tersedia dalam Tabel 3.4.



**Tabel 3. 4** Implementasi *Case Folding*

Sebelum Implementasi <i>Case Folding</i>	Setelah Implementasi <i>Case Folding</i>
Perusahaan Anda sedang dilanda krisis? <b>Gampang! Pilih</b> karyawan perempuan angkat jadi <b>CEO</b> uuudaaah. Ngapain repot-repot membenahi sistem di dalam perusahaan <b>SUSAH!</b> Lagian pemimpin perempuan kan juga lebih gampang buat disalah-salahin xixixixi	perusahaan anda sedang dilanda krisis? gampang! pilih karyawan perempuan angkat jadi ceo uuudaaah. ngapain repot-repot membenahi sistem di dalam perusahaan susah! lagian pemimpin perempuan kan juga lebih gampang buat disalah-salahin xixixixi

### 3.2.1.3. *Punctuation Removal*

Pada tahapan ini, dilakukan penghapusan terhadap seluruh tanda baca atau pun simbol yang terdapat dalam data dengan maksud menghilangkan bagian yang tidak diperlukan untuk efisiensi sistem dalam memahami data. Adapun tahapan *punctuation removal* yang diterapkan terhadap penelitian ini dapat diamati melalui *pseudocode* yang tertera di bagian bawah ini dengan hasil *punctuation removal* yang diperlihatkan di Tabel 3.5.

```
# Membuat tabel translasi punctuation
TRANSLATION_TABLE = MAKE_TRANSLATE(PUNCTUATION)
FOR EACH 'row IN' dataframe:
    # Menghapus Punctuation
    row['Comment'] = TRANSLATE(row['Comment'], TRANSLATION_TABLE)
END FOR
```

Adapun penerapan *punctuation removal* dapat dilakukan dengan menggunakan *String punctuation* untuk mendapatkan *list* tanda baca yang kemudian akan ditranslasi menjadi spasi dengan menggunakan fungsi *translate*. Dimana, penentuan pengubahan karakter tersebut ditampung dalam tabel translasi yang menggunakan fungsi *maketrans*.

**Tabel 3. 5** Implementasi *Punctuation Removal*

Sebelum Implementasi <i>Punctuation Removal</i>	Setelah Implementasi <i>Punctuation Removal</i>
perusahaan anda sedang dilanda krisis? gampang! pilih karyawan perempuan angkat jadi ceo uuudaaah. ngapain repot-	perusahaan anda sedang dilanda krisis? gampang pilih karyawan perempuan angkat jadi ceo uuudaaah ngapain repot

repot membenahi sistem di dalam perusahaan susah! lagian pemimpin perempuan kan juga lebih gampang buat disalah-salahin xixixixi	repot membenahi sistem di dalam perusahaan susah lagian pemimpin perempuan kan juga lebih gampang buat disalah salahin xixixixi
--	---

#### 3.2.1.4. Normalization

Tahapan ini melakukan proses normalisasi terhadap setiap kata dalam data. Dimana, kata yang tidak baku atau disingkat dan juga kata yang penulisannya salah akan diubah menjadi kata yang sesuai dengan kaidah bahasa yang lebih baik. Dengan demikian pengolahan data menjadi lebih sederhana dan data komentar dapat memiliki makna yang konsisten. Kamus baku yang dipakai pada penelitian ini disusun berdasarkan kata yang ditemukan tidak baku pada dataset. Adapun tahapan *normalization* ini dapat dilihat berdasarkan *pseudocode* berikut dengan hasil yang ditampilkan pada Tabel 3.6.

```

# Memuat file kamus kata baku dan slang
LOAD CSV FILE 'kamus_baku.csv' INTO 'normal_data'
# Membuat variabel yang menampung pengubahan kata slang menjadi normal
CREATE 'dict_normal' FROM 'normal_data' WHERE KEY='slang' AND
VALUE='normal'
# Memecah komentar menjadi kata
SPLIT 'comment' INTO 'words'
FOR EACH 'word' IN 'words'
    # Menormalisasi Pernyataan Komentar
    REPLACE 'word' WITH 'dict_normal'
JOIN 'words' INTO 'comment'
RETURN 'comment'

```

**Tabel 3. 6** Implementasi *Normalization*

Sebelum Implementasi <i>Normalization</i>	Setelah Implementasi <i>Normalization</i>
perusahaan anda sedang dilanda krisis gampang pilih karyawan perempuan angkat jadi ceo <b>uuudaaah ngapain</b> repot repot membenahi sistem di dalam perusahaan susah lagian pemimpin	perusahaan anda sedang dilanda krisis gampang pilih karyawan perempuan angkat jadi ceo <b>sudah mengapai</b> repot repot membenahi sistem di dalam perusahaan susah lagian pemimpin

perempuan kan juga lebih gampang buat disalah salahin xixixixi	perempuan kan juga lebih gampang buat disalah salahin xixixixi
--	--

### 3.2.1.5. Stopword Removal

Dalam tahap ini, diterapkan penghapusan terhadap kata-kata yang keberadaannya tidak mengandung arti khusus dan tidak mempengaruhi makna pada data. Dalam penerapan *stopword removal* ini, penulis menggunakan *library* Sastrawi yang menyediakan daftar *stopword* berbahasa Indonesia. Serta, terdapat tambahan kata *stopword* yang penulis sesuaikan dengan data penelitian, seperti ‘awowkwowkwo’, ‘njrr’, ‘slur’, ‘chuaksss’, ‘bjirrrr’ dan sebagainya yang disimpan dalam file *stopword.csv*. Adapun pelaksanaan *stopword removal* yang dilakukan terhadap penelitian ini dapat dilihat berdasarkan *pseudocode* yang tertera di bagian bawah ini dengan hasil *stopword removal* ditampilkan pada Tabel 3.7.

```
# Memuat stopword tambahan dan stopword dari Sastrawi
LOAD CSV FILE 'stopword.csv' INTO 'extra_stopword'
LOAD default stopwords FROM Sastrawi INTO 'ori_stopword'
# Menggabungkan kedua stopword
UNION 'ori_stopword' WITH 'extra_stopword' INTO 'data_stopword'
# Mempertahankan kata-kata tertentu pada komentar
REMOVE 'boleh', 'begitu', 'tidak', 'dalam', 'sampai', 'ada', 'harus', 'daripada',
'pasti', 'sudah', 'kenapa' FROM 'data_stopword'
# Membuat Stopword Remover
CREATE StopwordRemover USING 'data_stopword'
# Menghapus Stopword
APPLY StopwordRemover TO 'comment'
RETURN 'comment'
```

**Tabel 3. 7** Implementasi *Stopword Removal*

Sebelum Implementasi <i>Stopword Removal</i>	Setelah Implementasi <i>Stopword Removal</i>
perusahaan anda sedang dilanda krisis gampang pilih karyawan perempuan angkat jadi ceo sudah mengapai repot	perusahaan sedang dilanda krisis gampang pilih karyawan perempuan angkat jadi ceo mengapai repot repot

repot membenahi sistem <b>di</b> dalam perusahaan susah lagiin pemimpin perempuan kan juga lebih gampang buat disalah salahin <b>xixixixi</b>	membenahi sistem dalam perusahaan susah lagiin pemimpin perempuan kan lebih gampang buat disalah salahin
---	--

### 3.2.1.6. Tokenization

Tahapan ini melakukan pemisahan kalimat menjadi kata tunggal yang diumpamakan sebagai token. Dalam penerapan *tokenization* ini, penulis menggunakan *Hugging Face Transformers Library* untuk menginisiasi *tokenizer* dan juga model IndoBERT. Inisiasi *tokenizer* dalam tahapan ini ialah dengan menggunakan *tokenizer* yang telah dilatih sebelumnya untuk model IndoBERT. Dan sebelum tahapan tokenisasi dimulai, di setiap awal kalimat akan ditambahkan token khusus berupa [CLS] dan di akhir kalimat ditambahkan token [SEP]. Setelahnya ditambahkan juga token *padding* untuk menyamakan panjang kalimat. Hal tersebut dilakukan untuk membantu model IndoBERT. Adapun pelaksanaan lebih lanjut *tokenization* dalam penelitian ini dapat dilihat pada *pseudocode* yang tertera.

```

LOAD 'AutoTokenizer' FROM PRETRAINED 'indobert-base-p1' INTO 'tokenizer'
DEFINE FUNCTION 'tokenization(comment: string, max_length: integer = 512)':
  # Menambahkan Token Khusus
  SET 'add_token' = CONCAT '[CLS]', 'comment', '[SEP]'
  # Melakukan Tokenization
  SET 'tokenization' = APPLY 'tokenizer.tokenize' INTO 'add_token'
  # Memuat Segment
  SET 'segments' = LIST OF 1 WITH LENGTH OF 'tokenization'
  # Konversi Token ke Index
  SET 'index_token' = APPLY 'tokenizer.convert_tokens_to_ids' INTO
'tokenization'
  # Menambahkan Padding
  IF LENGTH OF 'index_token' < 'max_length':
    SET 'padding_length' = 'max_length' - LENGTH OF 'index_token'
    APPEND 'padding_length' ZEROS TO 'index_token'
    APPEND 'padding_length' ZEROS TO 'segments'
  RETURN 'index_token'

```





Representasi input yang dibangun oleh BERT merupakan penjumlahan dari *token embedding*, *segmentation embedding*, dan *position embedding* (Angeline et al., 2022). *Token embedding* merupakan input yang dihasilkan pada tahapan tokenisasi yang telah dilakukan sebelumnya menggunakan *AutoTokenizer* dan diubah ke dalam bentuk indeks. Setelahnya, melalui *segmentation embedding* dilakukan konversi nomor kalimat menjadi vektor. Pada penelitian ini dalam upaya membedakan bagian kalimat, konversi dilakukan dengan memberikan nilai tetap pada *segment* berupa 1 karena hanya terdapat satu kalimat dalam inputan. Kemudian melalui *position embedding* posisi kata dalam kalimat diubah menjadi vektor (Kurniasih & Parningotan Manik, 2022). Dalam penelitian ini, tahapan IndoBERT juga menerapkan *attention mask* untuk membantu efisiensi model dengan mengelola token yang valid bernilai 1 dan mengabaikan token *padding* bernilai 0 (Angeline et al., 2022). Selanjutnya, IndoBERT menghitung nilai rata-rata terhadap seluruh token dalam kalimat dan menghasilkan representasi vektor numerik berdimensi 768 yang nantinya akan digunakan sebagai data input terhadap algoritma *Bidirectional Long Short-Term Memory*. Hasil *embedding* tersebut dapat dilihat sebagai berikut melalui contoh kalimat “murahan sekali bro harga diri wanita ternyata lebih murah daripada harga motor kasihan besok bakal calon suami nya” yang tertera dalam Gambar 3.1.



```
[0.2944360375404358, 0.4688284397125244, -0.05580390244722366, 0.338185578584671,
-0.25409772992134094, 0.8519417643547058, 0.2691669464111328, 0.2938915491104126,
0.1559605598449707, 0.6361159682273865, -0.8182188868522644, 0.5896438956260681,
0.47623029351234436, 0.3246352970600128, 0.14122048020362854, 0.4566410183906555,
0.16740861535072327, 0.27142173051834106, -1.0100479125976562, -0.14406591653823853,
0.39222031831741333, 0.5082265734672546, 0.37301573157310486, -0.3773360550403595,
0.6127313375473022, 0.583882212638855, 0.08457474410533905, 0.1713087260723114,
-0.21430227160453796, -1.2280431985855103, -0.659153938293457, -0.46332836151123047,
0.43901491165161133, 0.2789587080478668, -2.1601510047912598, 0.25885874032974243,
-0.7963627576828003, 0.8520631790161133, 0.6669076085090637, -0.2651973366737366,
-1.2538118362426758, 0.7125016450881958, -0.5165520906448364, 0.06922135502099991,
-0.37331587076187134, 0.7433840036392212, -0.8228452801704407, 0.6061660051345825,
0.41132235527038574, 0.23818083107471466, -0.3141500949859619, -0.9750879406929016,
0.16433370113372803, 1.128466248512268, -0.34635835886001587, 0.7692850232124329,
0.8917250633239746, -0.863318920135498, -0.519882082939148, 1.602418303489685,
0.05829734727740288, 0.1390044391155243, -0.047548551112413406, 0.6367058753967285,
0.2353179007768631, 0.10348467528820038, -0.09874095022678375, -0.35219818353652954,
0.25149649381637573, -0.4140620827674866, -0.3327016830444336, -0.5710608959197998,
0.6011409163475037, -0.33472728729248047, 0.06635256111621857, -0.8168107271194458,
0.002257132437080145, 1.583011269569397, -0.29906269907951355, -0.3781459927558899,
-0.7389527559280396, -0.4007105827331543, -0.7054590582847595, 0.5525267124176025,
0.9248673915863037, -0.035639531910419464, 0.20873522758483887, -0.11962155997753143,
-0.7165272831916809, 0.23863165080547333, -0.3308861255645752, -0.48075103759765625,
1.1340590715408325, 0.8192136883735657, -0.5848682522773743, 0.30772364139556885,
0.6851749420166016, -0.4618600904941559, 0.1758824586868286, -0.12718982994556427, ...]
```

**Gambar 3. 1** Hasil *Embedding* terhadap Satu Contoh Kalimat

IndoBERT termasuk ke dalam *contextualized word embedding* yang melakukan pengubahan kata menjadi vektor berdasarkan konteks kalimat yang ada (Birunda &

Devi, 2020). Hal tersebut dapat diperlihatkan melalui hasil representasi vektor untuk kata “murah” pada kalimat “murahan sekali bro harga diri wanita ternyata lebih murah daripada harga motor kasihan besok bakal calon suami nya” dan “walaupun harganya murah kualitasnya tidak murahan inilah hp infinix termurah kamera belakang beresolusi kak esatuid” yang tampak dalam Gambar 3.2 dan 3.3.

```
[tensor([-4.8876e-01,  2.7092e-01, -9.2827e-01,  1.0512e+00, -2.1783e+00,
  7.3354e-01,  2.3975e+00, -9.2796e-01,  3.5817e-01,  1.3164e+00,
 -1.6814e+00, -8.6718e-02,  2.0932e-01,  4.4802e-02,  3.1519e-01,
  6.5024e-01,  1.3963e+00,  1.9486e+00, -1.1527e+00,  2.1474e-01,
 -3.6429e-01, -5.8407e-01,  1.7333e-01, -8.7297e-01,  9.0318e-01,
  7.6464e-01,  1.5441e+00, -5.9453e-01, -3.4707e-01, -1.5495e+00,
 -1.2670e+00, -9.2917e-01,  1.1989e+00,  7.3391e-01, -2.2236e+00,
  2.4406e+00, -9.2028e-01,  7.6476e-01,  8.4563e-01, -8.5914e-01,
 -1.5821e+00,  7.1499e-01, -2.0600e+00,  3.9057e-01, -8.1467e-01,
  1.6390e+00, -1.9338e+00, -9.3244e-01,  1.1024e+00,  5.4348e-01,
 -1.0498e+00, -1.4920e-01,  3.0059e-01,  5.2794e-01,  5.1765e-01,
  1.1473e+00,  1.5779e+00, -1.8911e+00,  1.0213e-01,  7.8070e-01,
  3.1527e-01,  8.8148e-01,  8.1711e-01,  4.9748e-01, -2.9057e-01,
  1.8988e-01, -3.7899e-01, -6.0557e-01,  5.5608e-01,  2.7951e-01,
  5.7637e-01,  5.9013e-01,  2.2762e+00, -1.1395e+00,  3.1593e-02,
 -2.0327e+00,  7.6067e-01,  1.5825e+00,  1.4998e+00,  1.1997e-01,
 -3.7542e-01, -1.9424e-01, -1.0033e+00, -4.3423e-02,  2.9406e+00,
  4.7927e-01, -8.2923e-02, -2.3056e-01, -4.4179e-01, -4.2893e-01,
 -6.9468e-01, -7.1767e-01,  2.7889e-01,  1.1141e-01, -1.3736e+00,
  1.7952e-01,  6.6201e-01, -9.6008e-01, -6.3763e-02, -1.9908e+00,
 -2.1157e+00,  3.6606e-01, -9.1795e-02, -2.2942e-02,  9.1095e-01,
  6.7036e-01,  1.1966e-02,  6.6400e-01,  1.2142e+00, -2.5951e+00,
 -4.2435e-01, -6.1602e-01, -3.1955e-01,  8.1684e-01,  9.3263e-01,
  3.1873e-02, -4.3677e-02, -2.0828e+00,  1.9747e+00,  3.4669e-01,
 ...,
  2.8624e-01,  5.0471e-01, -1.0661e+00, -1.4436e+00, -5.2910e-01,
  3.8696e-01, -1.6901e+00, -1.5649e-01, -3.9971e-01,  9.9105e-01,
 -1.3336e+00,  1.2581e+00,  6.1416e-01,  1.5169e+00,  1.7181e+00,
  1.5147e+00, -4.8369e-01, -4.9794e-01]])]
```

**Gambar 3.2** Vektor *Embedding* Kata ‘Murah’ pada Kalimat Misogini

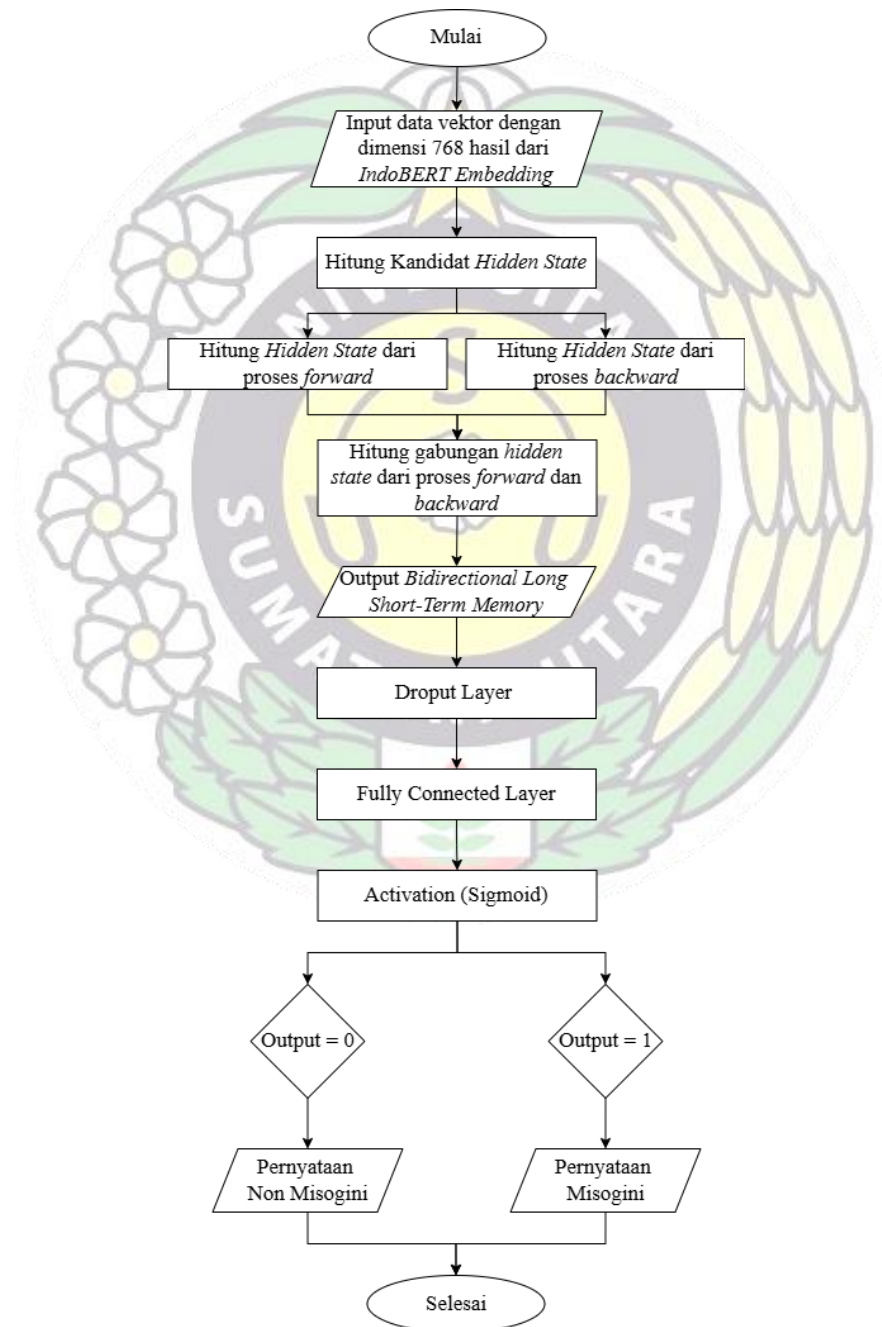
```
[tensor([-5.9621e-01,  4.2042e-02, -6.1148e-01, -2.6801e-01, -1.1934e+00,
  2.3161e-01,  5.5501e-01, -1.1685e+00, -1.9060e-02,  5.5049e-01,
 -1.0944e+00, -6.8246e-01, -3.2710e-01, -5.2699e-01, -3.5872e-01,
 -6.6991e-01,  7.9063e-01,  1.8935e+00,  6.6968e-01,  6.0482e-01,
  4.5196e-01, -1.3049e-01,  9.9969e-01,  2.7715e-01,  7.9852e-01,
 -1.3744e+00, -7.4482e-01,  8.7516e-01,  9.8930e-01, -9.4017e-01,
 -1.0502e+00,  1.0241e+00,  8.4876e-01,  1.1154e+00, -4.4016e-01,
  1.0839e+00,  1.3472e-01,  1.2712e+00,  9.5931e-01, -1.0181e-01,
 -7.9244e-01, -6.3734e-01, -9.5059e-01,  3.3535e-02, -1.2101e+00,
  1.7417e+00, -6.8041e-01,  5.7295e-01,  5.0020e-02, -1.0283e-01,
 -1.4628e+00,  1.0545e+00, -3.8015e-01, -2.0965e-01,  5.4881e-01,
  1.7280e+00,  2.1283e+00, -1.5318e+00,  8.6635e-01,  6.8161e-01,
 -6.6059e-01,  6.4740e-01,  4.7330e-01,  3.7560e-01,  8.2501e-01,
 -3.4436e-01,  3.9222e-01, -1.0197e+00, -9.3059e-01,  6.8268e-01,
  4.4038e-01,  3.6068e-01,  2.4566e+00, -1.2889e+00,  7.7574e-01,
 -8.8140e-01, -6.5688e-01,  1.8721e-01,  1.0723e+00, -3.8179e-02,
  2.9979e-01, -1.3317e+00, -1.4485e+00, -8.4573e-02,  3.1626e+00,
 -2.0872e+00, -1.3775e-01,  1.5209e-01, -3.8016e-01,  2.3047e-01,
  4.7830e-01,  1.0657e+00,  4.6074e-01, -7.8930e-01, -7.0228e-01,
  9.7844e-01,  3.6853e-01, -2.6667e-01, -8.9810e-01, -2.1237e+00,
 -1.1811e+00, -5.3067e-01,  2.5347e-01, -1.6415e-01,  8.6881e-01,
  6.3493e-01, -2.8991e-01, -5.4246e-02,  6.2343e-01, -1.0129e+00,
 -8.7071e-02, -3.5647e-01, -8.6130e-01, -1.8195e+00, -1.5590e-01,
  3.5267e-01,  1.3593e+00, -2.5828e+00,  1.3178e+00,  5.4815e-01,
 ...,
  8.7833e-01, -2.3482e-01,  3.4404e-01, -1.9300e+00, -5.4584e-01,
  4.1003e-01, -9.3105e-01,  1.0360e+00,  2.4198e-01,  5.8542e-01,
 -1.8745e+00, -1.6422e-01,  2.7287e-01,  4.8296e-01,  2.7479e+00,
  9.3037e-01, -6.4349e-01, -8.0407e-01]])]
```

**Gambar 3.3** Vektor *Embedding* Kata ‘Murah’ pada Kalimat Non Misogini

Berdasarkan hasil *embedding* di atas dapat dilihat representasi vektor dari kata yang sama dapat berbeda oleh karena pemaknaan kata tersebut di dalam kalimat. Hal ini tentunya dapat meningkatkan pemahaman model menjadi lebih baik terhadap data.

### 3.2.3. Bidirectional Long Short-Term Memory

Penelitian ini membangun model dengan menerapkan algoritma *Bidirectional Long Short-Term Memory*. Adapun proses pengembangannya tersedia pada *flowchart* yang dapat diperhatikan dalam Gambar 3.4.



**Gambar 3. 4** Flowchart Bidirectional Long Short-Term Memory



Alur kerja pada lapisan *Bidirectional Long Short-Term Memory* dalam Gambar 3.4 dapat diuraikan di bawah ini.

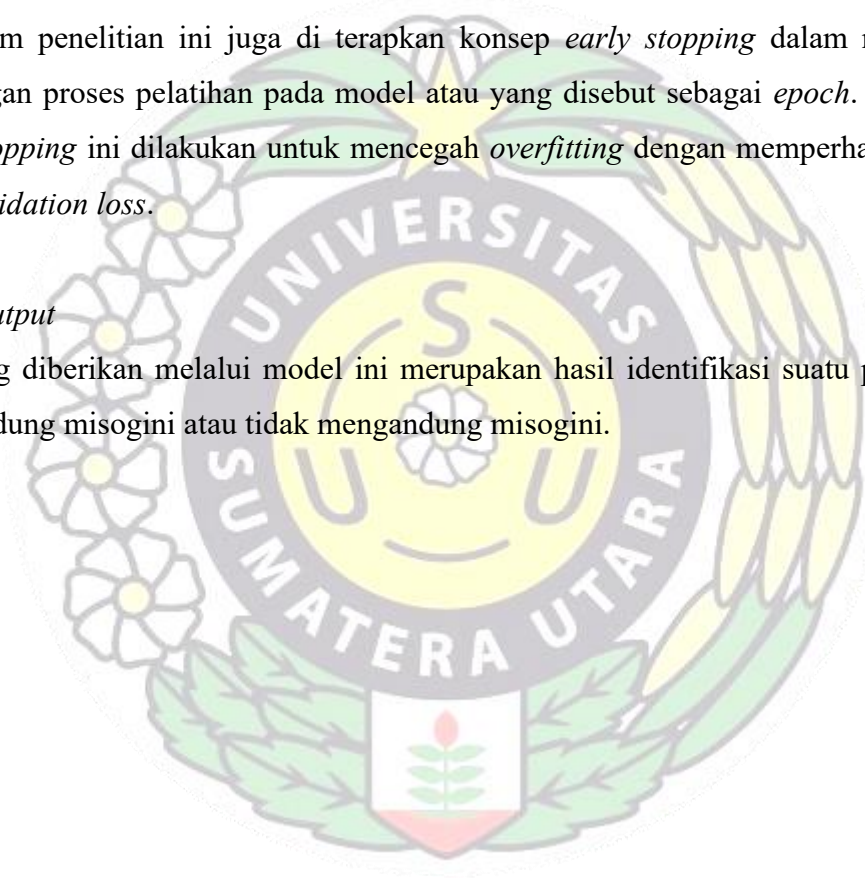
1. Pada langkah awal, proses ini dimulai dengan menggunakan hasil vektor IndoBERT *embedding* berdimensi 768 sebagai data input terhadap lapisan *Bidirectional Long Short-Term Memory*.
2. Kemudian dilakukan penentuan terhadap informasi yang dibuang, serta ditampung dan diteruskan ke *state* selanjutnya. Ini merupakan tahapan untuk menentukan kandidat *hidden state*.
3. Tahapan perhitungan *hidden state* ini dilakukan melalui dua proses, yaitu perhitungan *hidden state* dari kiri ke kanan (*forward pass*) dan perhitungan *hidden state* dari kanan ke kiri (*backward pass*). Perhitungan tersebut kemudian menghasilkan *output* yang diterima dari masa lalu dan yang diterima dari masa depan. Lalu, kedua hasil tersebut digabungkan untuk diteruskan ke tahapan selanjutnya.
4. Selanjutnya, hasil akhir *hidden state* pada lapisan *Bidirectional Long Short-Term Memory* diaplikasikan ke dalam *dropout layer* yang secara temporer menonaktifkan beberapa *neuron* pada *hidden layer* untuk mengurangi *overfitting* (Ainur Rohman et al., 2023).
5. Proses berikutnya menerapkan hasil ke dalam *fully connected layer*. Setiap *neuron* pada lapisan ini terintegrasi sepenuhnya ke lapisan sebelum dan setelahnya yang memungkinkan kombinasi untuk menentukan *output* terakhir. Seperti pada penelitian ini, model diharapkan untuk mengubah hasil ke dalam representasi yang sesuai dengan kelas target.
6. Kemudian dilakukan aktivasi *sigmoid* untuk membantu menghasilkan nilai probabilitas dalam rentang 0-1 yang sesuai dengan tugas klasifikasi biner pada penelitian ini.
7. Setelah *output* dihasilkan, model menentukan hasil identifikasi yang sesuai dengan rentang pada *sigmoid*. Dimana jika keluaran kalimat bernilai 1 maka kalimat tersebut diidentifikasi sebagai komentar yang memuat pernyataan misogini. Sedangkan untuk keluaran kalimat bernilai 0 maka kalimat tersebut diidentifikasi sebagai komentar yang tidak memuat pernyataan misogini alias non misogini.

Setelah semua tahapan di atas selesai, *optimizer* diterapkan pada model untuk mengoptimalkan kinerja model dan meminimalkan nilai *loss function* saat pelatihan model dilakukan. Penelitian ini menggunakan *optimizer* yang terbaik bernama Adam (*Adaptive Moment Estimation*) untuk melakukan optimasi. Hal itu dikarenakan Adam mengintegrasikan konsep *SGD* dan *AdaGrad* dengan memperbarui bobot melalui penggunaan rata-rata gradien dan menyesuaikan secara spesifik *learning rate* untuk setiap parameter (Ainur Rohman et al., 2023). Serta, penelitian ini memanfaatkan BCE (*Binary Cross Entropy*) sebagai *loss function* yang sesuai dengan target klasifikasi kelas, yaitu satu.

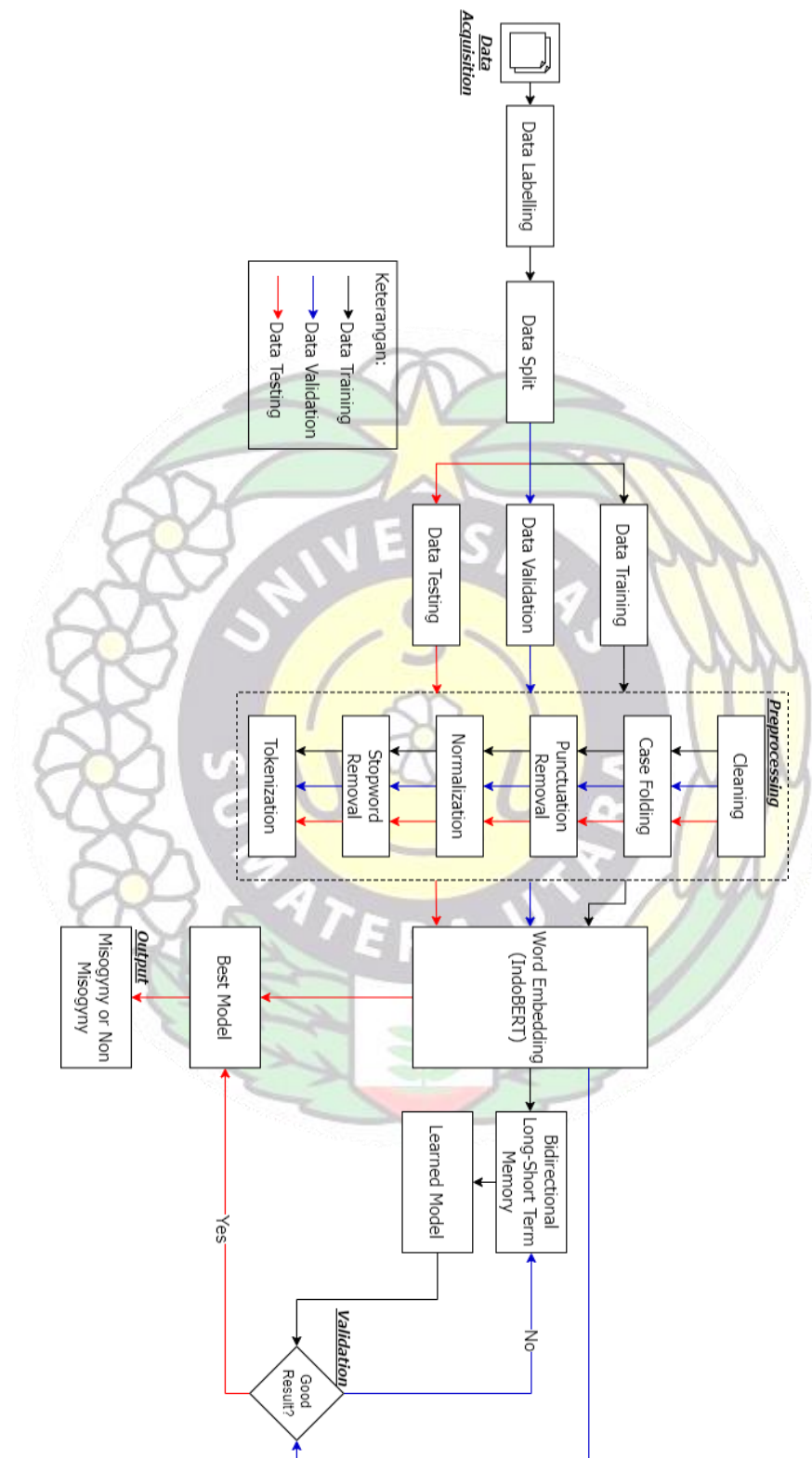
Dalam penelitian ini juga di terapkan konsep *early stopping* dalam melakukan perulangan proses pelatihan pada model atau yang disebut sebagai *epoch*. Penerapan *early stopping* ini dilakukan untuk mencegah *overfitting* dengan memperhatikan nilai pada *validation loss*.

#### 3.2.4. Output

Hal yang diberikan melalui model ini merupakan hasil identifikasi suatu pernyataan mengandung misogini atau tidak mengandung misogini.



Adapun keseluruhan proses yang telah dipaparkan terdapat dalam arsitektur umum yang tercantum dalam Gambar 3.5.



**Gambar 3. 5** Arsitektur Umum

### 3.3 Perancangan Sistem

Penelitian ini kemudian dilanjutkan dengan merancang *interface* sistem pengidentifikasi pernyataan misogini dan non misogini dengan berbasis *web*. Perancangan sistem ini bertujuan untuk memberikan pemahaman dan memudahkan pengoperasian sistem oleh pengguna, serta bertujuan sebagai pedoman penulis dalam membangun sistem.

#### 3.3.1. Desain Halaman Beranda

Halaman beranda merupakan tampilan awal yang akan dilihat oleh pengguna saat sistem mulai dioperasikan. Halaman ini memuat informasi terkait tombol yang menuju halaman lain, judul penelitian, identitas penulis, dan identitas perguruan tinggi. Desain halaman beranda tampak dalam Gambar 3.6.



**Gambar 3. 6** Desain Halaman Beranda

Keterangan:

1. Label 1 berfungsi meneruskan pengguna ke halaman *home*.
2. Label 2 berfungsi meneruskan pengguna ke halaman *training*.
3. Label 3 berfungsi meneruskan pengguna ke halaman *testing*.
4. Label 4 berfungsi sebagai *dropdown* yang memberikan pilihan *texts* dan YouTube *links* untuk *user input*.
5. Label 5 berfungsi meneruskan pengguna ke halaman *training*.



### 3.3.2. Desain Halaman Training

Halaman *training* memuat penyajian yang mengizinkan pengguna untuk mengunggah *file* berupa dataset berekstensi *.csv* dan kemudian memulai proses *training* dari data yang berhasil diunggah melalui tombol ‘Mulai Pelatihan’. Pada halaman ini juga ditampilkan hasil dari proses yang berlangsung ke dalam bentuk tabel. Desain halaman *training* diperlihatkan dalam Gambar 3.7.

**TUGAS AKHIR** Home Training Testing User Input~

Home/Training

**TRAINING**

Unggah data berekstensi *.csv* untuk proses *training*

Unggah File 1 Mulai Pelatihan 2

**HASIL PREPROCESSING**

Komentar	Preprocessing	Label

3

**Gambar 3. 7** Desain Halaman *Training*

Keterangan:

1. Label 1 berfungsi sebagai tempat unggahan *file data train* bagi pengguna.
2. Label 2 berfungsi meneruskan pengguna memulai proses *training*.
3. Label 3 berfungsi menampilkan hasil *preprocessing*.

### 3.3.3. Desain Halaman Testing

Halaman *testing* adalah laman yang mengizinkan pengguna untuk menjalankan tahapan pengujian terhadap model dengan memberikan dataset berekstensi *.csv* ke tempat unggahan *file* yang telah disediakan. Proses pengujian tersebut dapat dimulai ketika pengguna berhasil mengunggah *file* dan kemudian memilih tombol ‘Mulai Pengujian’. Halaman ini juga akan menampilkan hasil *testing* ke dalam tabel yang memuat informasi kolom berupa komentar, *preprocessing*, label, dan hasil identifikasi. Desain halaman *testing* dapat diamati melalui Gambar 3.8.

TUGAS AKHIR

Home

Training

Testing

User Input~

Home/Testing

TESTING

Unggah data berekstensi .csv untuk proses *testing*

Unggah File

Mulai Pengujian

HASIL TESTING

Komentar	Preprocessing	Label	Hasil Identifikasi

**Gambar 3. 8** Desain Halaman *Testing*

Keterangan:

1. Label 1 berfungsi sebagai tempat unggahan *file data test* bagi pengguna.
2. Label 2 berfungsi meneruskan pengguna memulai proses *testing*.
3. Label 3 berfungsi menampilkan hasil *testing* berupa identifikasi.

#### 3.3.4. Desain Halaman *User Input* berupa Teks

Halaman *text input* memuat tampilan yang mengizinkan pengguna untuk mengetikkan pernyataan yang ingin diidentifikasi ke area yang telah disediakan. Setelah usai menginput teks, pengguna dapat memulai proses identifikasi melalui tombol 'Identifikasi' dan kemudian melihat hasil identifikasi terhadap teks yang diberikan setelah proses identifikasi selesai dilakukan. Desain halaman *user input* berupa teks ini diperlihatkan oleh Gambar 3.9.

**TUGAS AKHIR** Home Training Testing User Input

Home/Text Input Texts Youtube Links

**TEXT INPUT**

Berikan pernyataan berupa teks yang akan diidentifikasi

Ketik pernyataan 1

Identifikasi 2

**HASIL IDENTIFIKASI**

3

**Gambar 3. 9** Desain Halaman *Text Input*

Keterangan:

1. Label 1 berfungsi menampung inputan pengguna yang berupa teks.
2. Label 2 berfungsi meneruskan pengguna melakukan proses identifikasi.
3. Label 3 berfungsi menampilkan hasil identifikasi terhadap teks yang diberikan.

### 3.3.5. Desain Halaman *User Input* berupa Tautan

Halaman YouTube *link input* adalah laman yang mengizinkan pengguna untuk memberikan *input* berupa tautan unggahan dari YouTube untuk diidentifikasi sebagai pernyataan yang mengandung misogini atau bukan. Proses identifikasi tersebut dapat dimulai ketika pengguna memberikan inputan tautan yang sesuai dan memilih tombol 'Identifikasi'. Hasil dari proses ini kemudian akan ditampilkan ke dalam bentuk tabel dengan kolom berisikan *username*, komentar, *preprocessing*, dan hasil identifikasi. Desain halaman YouTube *link input* tersedia dalam Gambar 3.10.

**TUGAS AKHIR**
Home
Training
Testing
User Input~

Home/YouTube Link Input

Texts

YouTube Links

YOUTUBE LINK INPUT

Berikan tautan unggahan video dari Youtube yang akan diidentifikasi

Masukkan tautan YouTube

Identifikasi

HASIL IDENTIFIKASI

Username	Komentar	Preprocessing	Hasil Identifikasi

**Gambar 3. 10** Desain Halaman YouTube *Link Input*

Keterangan:

1. Label 1 berfungsi menampung inputan pengguna yang berupa tautan YouTube.
2. Label 2 berfungsi meneruskan pengguna melakukan proses identifikasi.
3. Label 3 berfungsi menampilkan hasil identifikasi terhadap tautan yang diberikan.

### 3.4 Metode Evaluasi

Metode pengukuran kinerja yang diterapkan dalam penelitian ini ialah *Confusion Matrix*. Proses pengevaluasian dilakukan untuk meninjau seberapa besar ketepatan sistem yang telah dikembangkan dalam mengidentifikasi komentar yang memuat pernyataan misogini dan komentar yang tidak memuat pernyataan misogini. Metrik yang digunakan dalam mengevaluasi sistem yang dikembangkan ialah *precision*, *recall*, *F-1 score*, dan *accuracy*. Dimana nilai-nilai tersebut diambil dari penerapan *Confusion Matrix* yang memiliki nilai *True Positive*, *False Positive*, *True Negative*, dan *False Negative* seperti yang diperlihatkan oleh Tabel 3.10.



**Tabel 3. 10** Ilustrasi *Confusion Matrix*

		Aktual	
		Positif (1)	Negatif(0)
Identifikasi	Positif (1)	TP ( <i>True Positive</i> )	FP ( <i>False Positive</i> )
	Negatif (0)	FN ( <i>False Negative</i> )	TN ( <i>True Negative</i> )

Keterangan:

TP = Banyaknya data teridentifikasi sebagai positif dan aktual juga bernilai positif

FP = Banyaknya data teridentifikasi sebagai positif namun aktual bernilai negatif

TN= Banyaknya data teridentifikasi sebagai negatif dan aktual juga bernilai negatif

FN = Banyaknya data teridentifikasi sebagai negatif namun aktual bernilai positif



## BAB 4

### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1. Implementasi Sistem

Dalam proses pengembangan sistem identifikasi komentar berbahasa Indonesia yang mengandung unsur pernyataan misogini ataupun bukan dengan penerapan *Bidirectional Long Short-Term Memory* dan *IndoBERT Embedding*, penulis memanfaatkan komponen peranti keras dan peranti lunak sebagaimana penjelasan berikut ini.

##### 4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Rincian informasi terkait perangkat keras yang diterapkan dalam pengembangan sistem ialah yang tercantum berikut ini.

1. Laptop : HP 14s-cf2xxx
2. Processor : Intell CoreI i7-10510U CPU @ 1.80GHz 2.30 GHz
3. RAM : 8,00 GB

Rincian informasi terkait perangkat lunak yang diterapkan dalam pengembangan sistem ialah yang tercantum berikut ini.

1. Sistem Operasi : Windows 11 Home Single Language
2. IDE : Visual Studio Code
3. Bahasa Pemrograman : Python 3.10.8
4. Library :
  - *Flask* versi 3.0.3
  - *google-api-python-client* versi 2.129.0
  - *langdetect* versi 1.0.9
  - *matplotlib* versi 3.9.1
  - *numpy* versi 1.26.4
  - *pandas* versi 2.2.2
  - *Sastrawi* versi 1.0.1
  - *scikit-learn* versi 1.5.1
  - *seaborn* versi 0.13.2
  - *torch* versi 2.4.0
  - *transformers* versi 4.44.2

#### 4.1.2. Penerapan Desain Antarmuka Pengguna

Melalui rancangan antarmuka yang telah dipaparkan dalam Bab 3, berikut adalah realisasi dari rancangan tersebut.

##### 1. Tata Muka Halaman Beranda

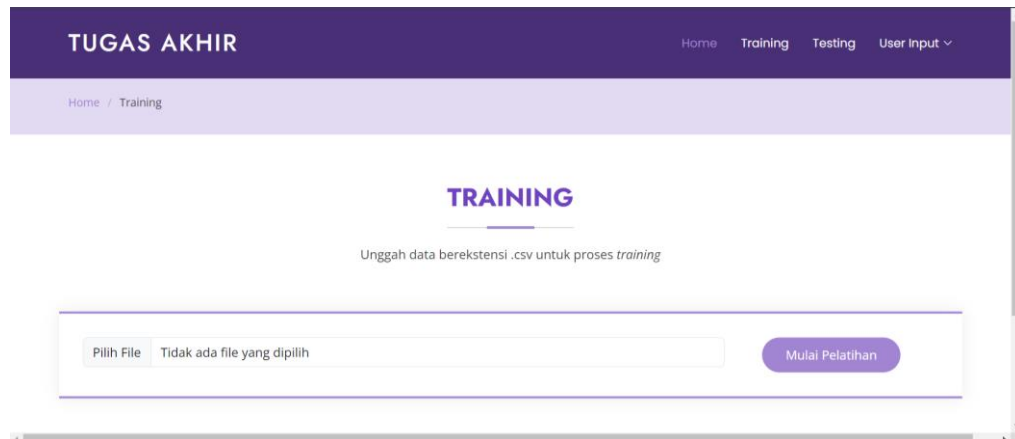
Beranda merujuk pada visualisasi halaman yang tersaji ketika pengguna pertama kalinya menjalankan sistem. Halaman ini menyediakan tombol dan navigasi yang dapat menghubungkan pengguna antara satu halaman dengan halaman lain. Di samping itu, melalui beranda ditampilkan informasi mengenai judul penelitian dan informasi mengenai penulis. Adapun tampilan halaman beranda terlihat dalam Gambar 4.1



**Gambar 4. 1** Tata Muka Halaman Beranda

##### 2. Tata Muka Halaman *Training*

Halaman *training* adalah *page* yang dapat dimanfaatkan oleh pengguna untuk menjalankan pelatihan terhadap model dengan mengunggah *file* dataset berformat *.csv* dan mengklik tombol “Mulai Pelatihan” untuk menjalankan prosesnya. Tata muka halaman *training* diperlihatkan dalam Gambar 4.2.



**Gambar 4. 2** Tata Muka Halaman *Training*

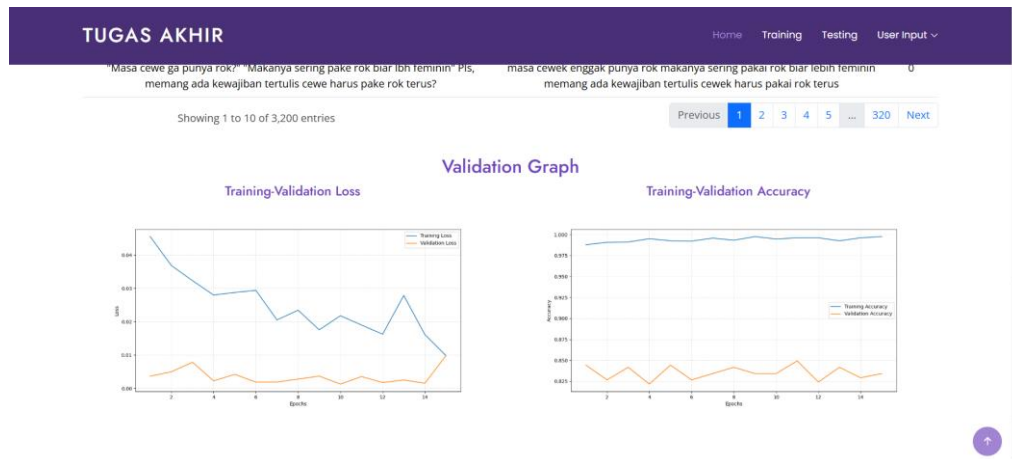
Pasca serangkaian tahap *training* usai, halaman akan menyajikan *output* proses dengan menggunakan tabel yang memuat data komentar, hasil *preprocessing*, dan label dari komentar tersebut. Adapun penyajian tabel ini dilakukan sebagai informasi terhadap pengguna agar dapat mengetahui perbandingan bentuk komentar hasil *scrapping* dengan komentar pasca *preprocessing*. Tata muka halaman hasil proses *training* diperlihatkan pada Gambar 4.3.

Komentar	Preprocessing	Label
Maaf ya padahal laki laki belum tentu perjaka tapi klo perempuan tidak perawan itu sangat terhina banget. Harusnya laki laki tidak perjaka juga harus juga berstatus sama dong. Itulah yang ga ada di negara ini	maaf iya padahal laki laki belum tentu perjaka kalo perempuan tidak perawan sangat hina banget harus laki laki tidak perjaka harus status sama dong yang enggak ada negara	0
dia nakal sdikit bodoh murah ??? pdhal cantik ??? nanggung bnget nakalx ngerugin mnding nakal kuat cuan ga siaA dosax sama ja ???	nakal sedikit bodoh murah padahal cantik nanggung banget nakal ngerugin mending nakal kuat cuan enggak sia dosa sama	1
jangan terlalu ambil hati komentar kotor dan jahat di luar sana	jangan terlalu ambil hati komentar kotor jahat luar sana	0
oh ibu rumah tangga ya? enak yaa nyantai huahahaha jujur cape jd ir drpd gue kerja lembur ampe jam 2 pagi dulu	ibu rumah tangga iya enak nyantai jujur capek jadi ibu rumah tangga daripada gue kerja lembur sampai jam pagi dulu	0
Saranku gk usah merasa paling pintar . Pilihlah pilihanmu tanpa harus menjatuhkan orang lain .	saran enggak usah rasa paling pintar pilih pilih harus jatuh orang	0
Sepertinya untuk menghilangkan stereotype adalah percaya dengan diri kita sendiri dan mencari pasangan (suami) yang mendukung juga. Plus jangan malu untuk bertanya dan minta di kritik.	seperti hilang stereotip percaya diri sendiri cari pasang suami dukung plus jangan malu tanya minta kritik	0
tau sendiri di masyarakat bagaimana, perempuan lebih tua lebih susah "laku"nya. Laki-laki juga banyak yang cari pasangan yang masih muda kan	tau sendiri masyarakat bagaimana perempuan lebih tua lebih susah laku nya laki laki banyak cari pasang masih muda kan	1

**Gambar 4. 3** Tata Muka Hasil *Training*

Selain menampilkan hasil *preprocessing*, di bawah tabel juga ditampilkan grafik laju *loss* dan *accuracy* antara *data train* dan *data validation*. Tujuan penyajian grafik ini dilakukan sebagai informasi tingkat *loss* dan *accuracy* tiap *epoch* pelatihan terhadap pengguna. Adapun tampilan grafik tersebut tersaji dalam Gambar 4.4.

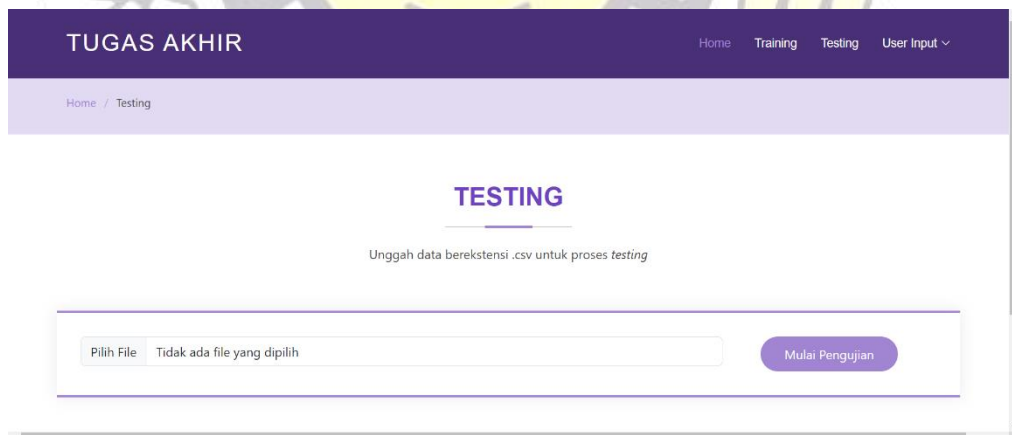




**Gambar 4. 4** Tata Muka Grafik *Loss* dan *Accuracy*

### 3. Tata Muka Halaman *Testing*

Halaman *testing* adalah *page* yang diperuntukkan bagi pengguna untuk menguji model yang telah melalui proses *training* sebelumnya. Mirip dengan halaman *training*, halaman *testing* juga meminta pengguna untuk mengunggah *file* dataset berformat *.csv* dan kemudian mengklik tombol “Mulai Pengujian” untuk memulai proses *testing*. Proses tersebut akan melakukan identifikasi terhadap data komentar dan menentukan label sesuai dengan pernyataan misogini atau non misogini yang ditemukan dalam komentar. Tampilan halaman *testing* terlihat dalam Gambar 4.5.



**Gambar 4. 5** Tata Muka Halaman *Testing*

Pasca proses *testing* usai, halaman akan menyajikan *output* proses dalam bentuk tabel yang memuat data komentar, hasil *preprocessing*, label seharusnya dari komentar tersebut, dan label identifikasi yang dilakukan oleh model. Selain itu, terdapat juga tombol ‘*Tampilkan Kesalahan Identifikasi*’ yang dapat melakukan penyaringan terhadap data yang salah identifikasi. Adapun penyajian tabel dan

tombol ini dilakukan sebagai informasi terhadap pengguna agar dapat mengetahui perbandingan hasil identifikasi komentar dengan label seharusnya. Hasil tersebut tercantum dalam Gambar 4.6 dan Gambar 4.7.

TUGAS AKHIR

HomeTrainingTestingUser Input

Hasil Identifikasi

Tampilkan Kesalahan Identifikasi

Show10entries

Search:

Komentar	Preprocessing	Label	Identifikasi
"Cewek kalo jadi wanita karir dapetin jodohnya susah, karena cowok2 bakal minder, tapi kalo cowok sebaliknya" Padahal aku cuma pengen jadi sukses & ga nyusahin banyak orang??	cewek kalau jadi wanita karir mendapatkan jodohnya susah cowok bakal minder kalau cowok sebaliknya padahal aku cuma pengen jadi sukses enggak nyusahin banyak orang	0	0
"jika anda mau dihargai, ya anda harus menghargai dong"	mau dihargai iya anda harus menghargai dong	0	0
"Kalian mau ngapain sih nyari ini tugas siapa itu tugas siapa? Nikah itu hidup bersama bukan tinggal bersama,apa salahnya saling bantu saling melengkapi,siapa yang sempat ya bantu kerjakan Suami klu lg kerja, istri yang kerjain rumah klu suami lg libur ya bantu kerjain rumah,ngga perlu tau ini kerjaan siapa"	kalian mau mengapai mencari tugas siapa tugas siapa nikah hidup bersama bukan tinggal bersama apa salahnya saling bantu saling melengkapi siapa sempat iya bantu kerjakan suami kalau kerja istri kerjain rumah kalau suami libur iya bantu kerjain rumah enggak perlu tahu pekerjaan siapa	0	0
"Kurus banget sih kek kurang gizi, emang lo pikir cowok? suka sama cewek yg kurus banget?" Udah jadi makanan setiap hari. Padahal tu orang tau sendiri gw kurus krn emang gini, sehari makan 4-5x aja gak bakal berubah	kurus banget kayak kurang gizi memang lo pikir cowok suka sama cewek kurus banget sudah jadi makanan hari padahal orang tahu sendiri gue kurus memang begini sehari makan kali enggak bakal berubah	0	0
"Tetap positif affirmation. Tidak perlu insecure dengan segala kekurangan puan. Bukankah dari dulu perempuan di nomor duakan, tapi kita tetap kuat."	tetap positif affirmation tidak perlu insecure segala kekurangan puan bukannya dulu perempuan nomor duakan kita tetap kuat	0	0

**Gambar 4. 6** Tata Muka Halaman Hasil Proses *Testing* (Keseluruhan)

TUGAS AKHIR

HomeTrainingTestingUser Input

Hasil Identifikasi

Tampilkan Keseluruhan Data Uji

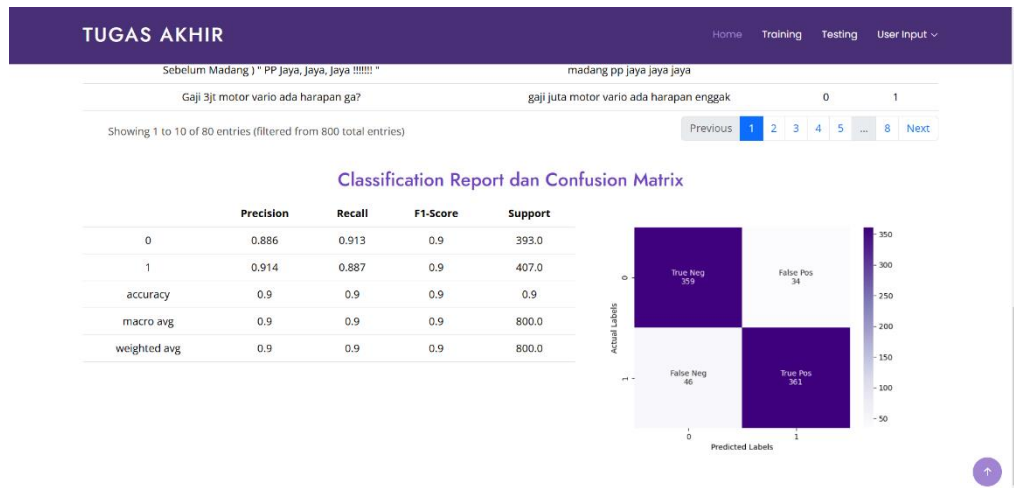
Show10entries

Search:

Komentar	Preprocessing	Label	Identifikasi
@Abi_Benedict @neohistoria_id mata lo pujaan. saya perempuan dan itu bentuk pelecehan pada wanita. lu coba ngomong ama mak lu kalau punya tobrut kalau nggak dipenggal lehermu ya dikubur hidup2.	mata lo pujaan perempuan itu bentuk pelecehan wanita lu coba mengomong sama mak lu kalau punya toket brutal kalau enggak dipenggal lehermu iya dikubur hidup	0	1
aku kok bisa nyasar ke sini?	aku kok nyasar sini	0	1
Aku nggak suka pake alas kaki. Buat apa aku pakai alas kaki kalau bukan kamu alasanku melangkah lagi?? dan juga bau amis ih,iya amis.amis you maksudnya?????	aku enggak suka pakai alas kaki buat apa aku pakai alas kaki kalau bukan kamu alasanku melangkah dan bau amis iya amis amis you maksudnya	0	1
Apa karna iri jadi pada banyak bacot	apa iri jadi banyak bacot	0	1
Apaan dah orang Indo masalah selangkangan orang aja dipermasalahkan, kaya gk pernah nonton video vokep aja. Mentang2 artis gitu biar rame ya.	apaan sudah orang Indonesia masalah selangkangan orang dipermasalahkan kaya enggak pernah menonton video bokep mentang artis begitu biar rame iya	0	1
Artis yang gak pernah di gosipin aneh	artis enggak pernah gosipin aneh	0	1
Bangun tdr pembahasannya berat bousss??	bangun tidur pembahasannya berat bos	0	1
Biasanya pake jaket, sekalinnya keluar pake jilbab cantik bgtt aaaaa	biasanya pakai jaket sekalinnya keluar pakai jilbab cantik banget	0	1

**Gambar 4. 7** Tata Muka Halaman Hasil Proses *Testing* (Salah Identifikasi)

Kemudian, di bawah tabel tersebut akan ditampilkan hasil evaluasi pemodelan menggunakan *Confusion Matrix* sebagaimana dijelaskan dalam Bab 3. Visualisasi yang dipakai untuk menampilkan hal tersebut ialah berbentuk *heatmap*. Dan di samping *heatmap*, halaman juga akan menampilkan *classification report* yang memuat nilai *precision*, *recall*, *F-1 score*, dan *accuracy*. Representatif dari penggambaran hasil tersebut diperlihatkan dalam Gambar 4.8.



**Gambar 4. 8** Tata Muka Evaluasi Model

4. Tata Muka Halaman *User Input* berupa Teks

Halaman *Text Input* adalah *page* yang dapat digunakan pengguna untuk menginput teks atau pernyataan yang ingin diidentifikasi ke dalam area *text* yang telah disediakan. Dan kemudian mengklik tombol “Identifikasi” untuk mulai mengidentifikasi apakah teks tersebut termasuk pernyataan misogini atau non misogini. Tata muka halaman untuk *text input* diperlihatkan oleh Gambar 4.9.

**TUGAS AKHIR** Home Training Testing User Input

Home / Text Input

### TEXT INPUT

Berikan pernyataan berupa teks yang akan diidentifikasi

Ketik pernyataan

Identifikasi

© Copyright Stephani Uli Basa Silitonga All Rights Reserved  
Designed by BootstrapMade

**Gambar 4. 9** Tata Muka Halaman *Text Input*

Ketika proses identifikasi selesai, halaman akan menampilkan hasil identifikasi seperti yang tercantum dalam Gambar 4.10 dan Gambar 4.11.

**TUGAS AKHIR** Home Training Testing User Input ▾

### TEXT INPUT

Berikan pernyataan berupa teks yang akan diidentifikasi

Nggak adil loh misal pihak cewek terus yang dipermasalahkan pulang malem meanwhile suaminya bisa nggak pulang sama sekali tanpa kabar

Ulang Identifikasi

#### Hasil Identifikasi terhadap Teks

Teks	Identifikasi
Nggak adil loh misal pihak cewek terus yang dipermasalahkan pulang malem meanwhile suaminya bisa nggak pulang sama sekali tanpa kabar. Itu yang bikin kesel.	Pernyataan Non Misogini

**Gambar 4. 10** Tata Muka Hasil Identifikasi *Text Input* (Non Misogini)

**TUGAS AKHIR** Home Training Testing User Input ▾

### TEXT INPUT

Berikan pernyataan berupa teks yang akan diidentifikasi

Tu bukan perempuan. Dia tu perempuan jalang. Kupu kupu malam

Ulang Identifikasi

#### Hasil Identifikasi terhadap Teks

Teks	Identifikasi
Tu bukan perempuan. Dia tu perempuan jalang. Kupu kupu malam	Pernyataan Misogini

**Gambar 4. 11** Tata Muka Hasil Identifikasi *Text Input* (Misogini)

5. Tata Muka Halaman *User Input* berupa Tautan

Halaman *Link Input* adalah *page* yang diperuntukkan bagi pengguna untuk melakukan identifikasi komentar terhadap sebuah postingan dari video YouTube secara langsung. Dimana pengguna dapat memberikan input berupa tautan *link* video YouTube dan kemudian mengklik tombol “Identifikasi” yang akan memulai proses *crawling* komentar sekaligus proses identifikasi. Adapun tampilan halaman *Link* YouTube tampak dalam Gambar 4.12.



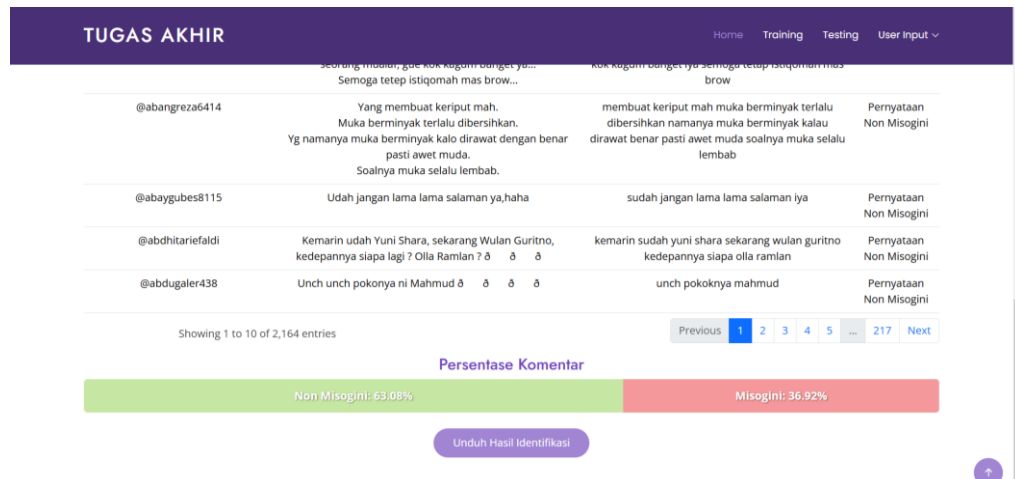
**Gambar 4. 12** Tata Muka Halaman *Link Input*

Saat proses selesai terlaksana, halaman akan menampilkan tabel yang berisikan *username* yang meninggalkan komentar, komentar dari *posting-an* YouTube, hasil *preprocessing*, dan hasil identifikasi seperti yang tercantum dalam Gambar 4.13.

Username	Komentar	Preprocessing	Identifikasi
@017_kadekthiraptraraditya8	UNDANG GOFAR HILMAN	undang gofar hilman	Pernyataan Non Misogini
@04.desakdemawarniindah.p36	eva lebih mirip dg papanya bukan dg mama nya..biarkan eva memilih jin sdri supaya lebih baik dg pasangan nya kelak (rita bali)	eva lebih mirip papanya bukan mama nya biarkan eva memilih jalan sdri lebih baik pasangan nya kelak rita bali	Pernyataan Non Misogini
@25_syahrezamoelim66	<a href="https://youtu.be/c5hb5tpjGSQ">https://youtu.be/c5hb5tpjGSQ</a>	teman bantu like dong buat tugas kuliah	Pernyataan Non Misogini
@_D_akbarharunbaihaqi	Teman2 bantu like dong buat tugas kuliah	undang komeng dong om	Pernyataan Non Misogini
@aanandiwan649	Undang chintya candra naya om Bantu up dong	undang chintya candra naya om bantu up dong	Pernyataan Non Misogini
@aansucipto7113	Gue sih seneng liat lo sexy, astaghfirillah... Kalau yang ngucap istighfar itu yg dari lahir dah muslim, mungkin udah gak aneh, tapi pas yg ngucapin adalah seorang mu'alaf, gue kok kagum banget ya... Semoga tetep istiqomah mas brow...	gue senang lihat lo seksi astaghfirillah kalau ngucap istighfar lahir sudah muslim mungkin sudah enggak aneh pas yang mengucapkan seorang mu'alaf gue kok kagum banget iya semoga tetap istiqomah mas brow	Pernyataan Non Misogini

**Gambar 4. 13** Tata Muka Hasil *Crawl* dan Identifikasi dari *Link Input*

Kemudian, di bawah tabel hasil identifikasi terdapat keterangan persentase jumlah pernyataan misogini dan pernyataan non misogini yang divisualisasikan dengan *bar chart*. Di samping itu, terdapat juga tombol 'Unduh Hasil Identifikasi' yang ketika ditekan akan mengunduh hasil identifikasi berupa kolom *username*, komentar, dan identifikasi ke dalam *file* berbentuk *.pdf*. Visualisasi terkait kedua hal tersebut dan isi *file pdf* diperlihatkan dalam Gambar 4.14 dan Gambar 4.15.



Gambar 4. 14 Tata Muka *Bar Chart* dan Tombol Unduh

The screenshot shows a web application interface with a dark theme. It displays a table of downloaded files. The table has columns for Username, Komentar, and Identifikasi. The comments are from users @017\_kadekdhirapatraraditya8, @04\_desakdemawarniindah.p36, @25\_syahrezamoslim66, @\_D\_akbarharunbaihaqi, @aanandiwan649, @aansucipto7113, @abangreza6414, @abaygubes8115, @abdhitariefaldi, @abdugaler438, and @AbdulHakim-ct4vi. All identifications are 'Pernyataan Non Misogini'.

Username	Komentar	Identifikasi
@017_kadekdhirapatraraditya8	UNDANG GOFAR HILMAN	Pernyataan Non Misogini
@04_desakdemawarniindah.p36	eva lebih mirip dg papanya bukan dg mama nya..biarkan eva memilih jin sdri supaya lebih baik dg pasangan nya kelak( rita bali)	Pernyataan Non Misogini
@25_syahrezamoslim66	<a href="https://www.youtube.com/watch?v=c5hbSlpJGSQ">https://youtu.be/c5hbSlpJGSQ</a>  Teman2 bantu like dong buat tugas kuliahôY™	Pernyataan Non Misogini
@_D_akbarharunbaihaqi	Undang Komeng dong om	Pernyataan Non Misogini
@aanandiwan649	Undang chintya candra naya om Bantu up dong	Pernyataan Non Misogini
@aansucipto7113	Gue sih seneng liat lo sexy, astaghfirloh... Kalau yang ngucap Istighfar itu yg dari lahir dah muslim, mungkin udah gak aneh, tapi pas yg ngucapin adalah seorang muafaf, gue kok kagum banget ya... Semoga tetep istiqomah mas brow...	Pernyataan Non Misogini
@abangreza6414	Yang membuat keriput mah. Muka berminyak terlalu dibersihkan. Yg namanya muka berminyak kalo dirawat dengan benar pasti awet muda. Soalnya muka selalu lembab.	Pernyataan Non Misogini
@abaygubes8115	Udah jangan lama lama salaman ya,haha	Pernyataan Non Misogini
@abdhitariefaldi	Kemarin udah Yuni Shara, sekarang Wulan Guritno, kedepannya siapa lagi ? Olla Ramlan ?	Pernyataan Non Misogini
@abdugaler438	Unch unch pokonya ni Mahmud	Pernyataan Non Misogini
@AbdulHakim-ct4vi	Editino dedy kasian skill editino nva oa keake full hahaha cuman cut and	Pernyataan Non Misogini

Gambar 4. 15 Tata Muka Isi Unduhan *File*

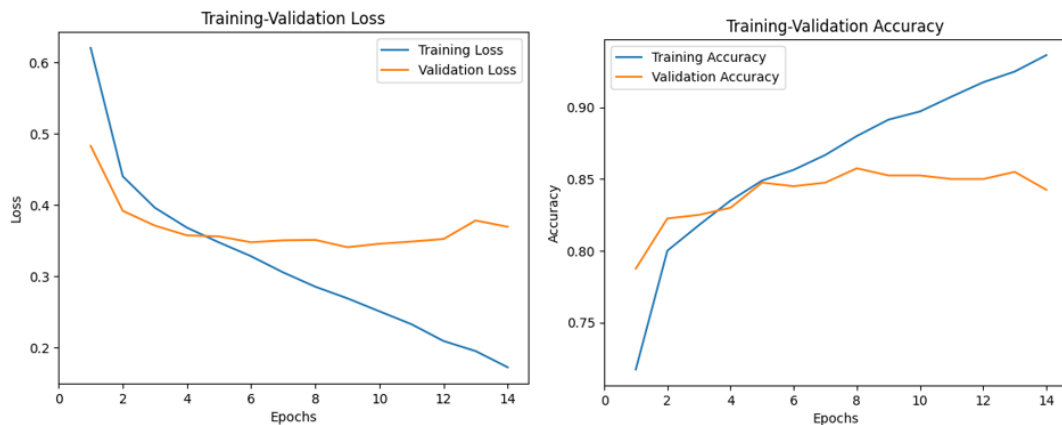
## 4.2. Pelatihan Model

Proses pelatihan model dilakukan dengan mengutak-atik parameter yang digunakan dalam model secara manual atau dapat disebutkan juga melalui proses *hit and trial* untuk menemukan potensi parameter terbaik. Proses tersebut dapat dikatakan sebagai *hyperparameter tuning* yang tujuannya ialah untuk menghasilkan model yang baik. *Hyperparameter tuning* pada model ini dilakukan pada nilai *learning rate*, jumlah *batch size*, dan jumlah *neuron*. Proses yang dimaksudkan dapat diamati menggunakan Tabel 4.1.

**Tabel 4. 1** Tabel Percobaan Parameter Model

No.	Learning Rate	Neuron	Batch Size	Validation Loss	Validation Accuracy
1.	0.00002	256	16	0.34065	0.85000
2.	0.00002	256	32	0.35469	0.85500
3.	0.00002	256	64	0.32436	0.84500
4.	0.00002	512	16	0.34382	0.85000
5.	0.00002	512	32	0.34476	0.85250
6.	0.00002	512	64	0.31432	0.85250
7.	0.00002	1014	16	0.38415	0.83000
8.	0.00002	1024	32	0.34814	0.85250
9.	0.00002	1024	64	0.33635	0.84250
10.	0.00003	256	16	0.35150	0.83000
11.	0.00003	256	32	0.35495	0.83500
12.	0.00003	256	64	0.32713	0.84500
13.	0.00003	512	16	0.40038	0.83250
14.	0.00003	512	32	0.34915	0.84000
15.	0.00003	512	64	0.32706	0.84500
16.	0.00003	1014	16	0.37911	0.82750
17.	0.00003	1024	32	0.36318	0.83250
18.	0.00003	1024	64	0.32270	0.84750
19.	0.00005	256	16	0.35300	0.86000
20.	0.00005	256	32	0.34602	0.84250
21.	0.00005	256	64	0.37064	0.81500
22.	0.00005	512	16	0.38317	0.84000
23.	0.00005	512	32	0.35664	0.84750
24.	0.00005	512	64	0.33016	0.85000
25.	0.00005	1014	16	0.37057	0.84250
26.	0.00005	1024	32	0.38628	0.83750
27.	0.00005	1024	64	0.34831	0.83750

Berdasarkan hasil percobaan *hyperparameter tuning* yang disajikan dalam Tabel 4.1, dapat ditarik sebuah kesimpulan dimana nilai parameter yang terbaik ialah *learning rate* senilai 0.00005, *neuron* senilai 256, dan *batch size* senilai 16. Hal tersebut dikarenakan parameter tersebut menghasilkan akurasi validasi yang terbaik yaitu sebesar 0.86 dan *loss* validasi sebesar 0.353. Adapun tampilan yang menyajikan gambar grafik *loss* dan akurasi validasi dalam Gambar 4.16.



**Gambar 4. 16** Tata Muka Grafik *Loss* dan *Accuracy* Model

### 4.3. Pengujian Kinerja Model

Pasca pelatihan model selesai, selanjutnya dilakukan pengujian terhadap kinerja model dengan melibatkan data baru yang tidak pernah dipelajari oleh model. Data tersebut merupakan data uji yang telah dibagi ditahap awal dan memiliki sebanyak 800 komentar atau 20% dari data yang dikumpulkan. Pengujian terhadap data ini tidak langsung dilakukan, melainkan data tersebut menjalani serangkaian tahapan yang sama seperti pada proses pelatihan model. Data uji akan melalui tahapan *preprocessing* yaitu *cleaning*, *case folding*, *punctuation removal*, *normalization*, *stop word removal*, dan *tokenization*. Berikutnya data uji juga melewati tahapan *embedding* dengan IndoBERT dan hasil representasi numerik yang dikeluarkan menjadi inputan untuk model *Bidirectional Long Short-Term Memory* yang telah dilatih sebelumnya. Hasil pengujian kinerja model diperlihatkan dalam Tabel 4.2.

**Tabel 4. 2** Hasil Pengujian Kinerja Model

No	Komentar	Hasil Preprocessing	Label	Identifikasi
1	Dah jelek tonggos	sudah jelek tonggos	1	1
2	Cantik bgt kak	cantik banget kak	0	0
3	Kaya bagus aja patudara	kaya bagus payudara	1	1
	nya ,, kendor gede kya hbs melahirakn	nya kendur gede kaya habis melahirakn		
4	Sorry kak , aku ketawa pas kakak bilang kecekek tete sendiri. Karena saya laki	sorry kak aku ketawa pas kakak bilang kecekek tetek sendiri	0	0



**Tabel 4. 2** Hasil Pengujian Kinerja Model (Lanjutan)

No.	Komentar	Preprocessing	Label	Identifikasi
	dan saya ga tau klo tete gede tu ternyata bikin tidak nyaman bahkan sampai terbawa tidur. Makasih ya narasi dan semua narasumber , hal ini memberiku pelajaran untuk bisa lebih menghargai semua aspek yang dimiliki seseorang entah dari segi fisik , kemampuan , dan lain-lain.	saya laki saya enggak tahu kalau tetek gede ternyata bikin tidak nyaman bahkan sampai terbawa tidur terima kasih iya narasi semua narasumber ini memberiku pelajaran bisa lebih menghargai semua aspek dimiliki seseorang entah segi fisik kemampuan , dan fisik kemampuan lain		
5	sering bgt dapet pertanyaan, baik sengaja mau nyindir atau ngga, banyak bgt yg nanya kpn nikah, seakan2 cewek yg telat nikah itu aib bgt,, kadang suka aneh jg banyak bahkan sesama perempuan melecehkan perempuan lain krna blm nikah ??	sering banget pertanyaan baik sengaja mau menyindir enggak banyak banget tanya kapan nikah seakan cewek telat nikah aib banget kadang suka aneh banyak bahkan sesama perempuan melecehkan perempuan karena nikah	0	0
6	@tuansinis @igtjsn @JNECare @JNE_ID Kenapa ya klo kayak gini BANYAK BANGET yang malah nyalahin korbannya BUKAN FOKUS SAMA KESALAHAN	kenapa iya kalau kayak begini banyak banget malah menyalahkan korbannya bukan fokus sama kesalahan pelakunya mulai hipnotis pelecehan	0	0

**Tabel 4. 2** Hasil Pengujian Kinerja Model (Lanjutan)

No.	Komentar	Preprocessing	Label	Identifikasi
	PELAKUNYA. Mulai dr hipnotis pelecehan BAHKAN PAKET ANCURPUN SALAH KORBAN WKWKW	bahkan paket hancur salah korban		
7	Gede beet enem nya jadi pengen pegang wkww	gede banget enem nya jadi pengen pegang	1	1
8	kurang lebar ngangkangnya	kurang lebar ngangkangnya	1	1
9	Makin cantik aja pacar saya ini:v	makin cantik pacar ini	0	0
10	Saya cowo yang belum pernah instal tiktok sama sekali dari dulu, di tiktok bukan cuma banyak konten joget joget gajelas ajh, tapi banyak juga konten brainwashing, dan penggiringan opini, banyak rumah tangga yang hancur karena istri kena brainwashing dari konten kretor cewe yang selalu ngeluarin statement dongo nya, dan akhirnya di serap juga oleh wanita yang pemikiran nya cetek	cowok belum pernah instal tiktok sama sekali dulu tiktok bukan cuma banyak konten joget joget enggak jelas tapi banyak konten brainwashing penggiringan opini banyak rumah tangga hancur istri kena brainwashing konten kreator cewek selalu mengeluari statement dongo nya akhirnya serap oleh wanita pemikiran nya cetek	1	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

**Tabel 4. 2** Hasil Pengujian Kinerja Model (Lanjutan)

No.	Komentar	Preprocessing	Label	Identifikasi
791	Masyallah teteteh mah cantik banget sumpah dahhhhhh	masyallah teteteh mah cantik banget sumpah	0	0
792	col* si biasanya	coli biasanya	1	1
793	Sayang nya kecil kalau gede wahh mantep bener?????	sayang nya kecil kalau gede mantap benar	1	1
794	Gabakal ada asap kalo gaada api	enggak bakal ada asap kalau enggak ada api	1	1
795	Klo banyak uang mah enak aja di liat nya ya ??	kalau banyak uang mah enak di lihat nya iya	1	1
796	Aku sih malah fokus di mimik wajahnya	aku malah fokus mimik wajahnya	0	0
797	iih knpa bajunya gitu iih tegang deh	kenapa bajunya begitu tegang deh	1	0
798	Sherly... kamu harus berubah... jangan suka menjatuhkan orang lain	sherly kamu harus berubah jangan suka menjatuhkan orang	0	0
799	Gaji 3jt motor vario ada harapan ga?	gaji juta motor vario ada harapan enggak	0	1
800	banyak makna keren	banyak makna keren	0	0

Dalam pengujian kinerja model, ditemukan sejumlah kesalahan prediksi yang dipengaruhi oleh beragam faktor. Penulis telah mengidentifikasi beberapa hal yang kemungkinan besar menjadi penyebab kesalahan identifikasi oleh model. Beberapa faktor tersebut ialah:

1. Terdapat komentar yang mengandung kata-kata berunsur negatif, sehingga model cenderung salah dalam melakukan identifikasi.

Contoh:

- @Abi\_Benedict @neohistoria\_id mata lo pujian. saya perempuan dan itu bentuk pelecehan pada wanita. lu coba ngomong ama mak lu kalau punya tobrut kalau nggak dipenggal lehermu ya dikubur hidup2.
- Jgn pernah jadikan wanita sebagai obyek halusinasi nafsu seks , sampai separah itu kah tidak dpt menahan sebuah hasrat?? Sungguh rendah sekali manusia yg melakukan pemerkosaan , apalagi kpd ank2 , gk ada otaknya !!!!

Komentar-komentar tersebut dianggap mengandung misogini oleh sistem dikarenakan terdapat kata yang menggambarkan unsur negatif mengarah misogini seperti “tobrut”, “dipenggal”, “nafsu”, “seks” dan sebagainya.

2. Adanya komentar yang tidak secara langsung menunjukkan unsur negatif.

Contoh:

- Baru tau nityior punya boba
- Mendadak tembus pandang mata gw
- Asupan susu yg bergizi bagi para lelaki ya ini....lanjutan..

Komentar-komentar di atas tidak dianggap mengandung misogini oleh sistem. Hal tersebut terjadi karena tidak terdapat kata kunci yang menyinggung misogini secara langsung dan juga karena adanya pemilihan bahasa bermakna kiasan.

3. Ditemukan kata-kata berbahasa asing dalam komentar.

Contoh:

- Pangkuanku ready kak??
- Km mau d unboxing...?

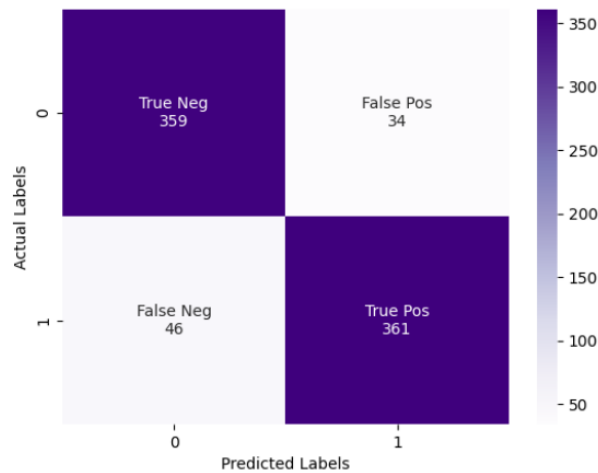
Komentar-komentar yang disebutkan di atas, tidak dianggap sebagai misogini karena terdapat penggunaan bahasa asing yang tidak dapat dipahami oleh sistem yang berbasis untuk Bahasa Indonesia.

#### 4.4. Pengukuran Kinerja Model

Dalam penelitian ini, evaluasi kinerja model dilakukan untuk mengetahui keakuratan IndoBERT *Embedding* dan *Bidirectional Long Short-Term Memory* dalam mengidentifikasi pernyataan yang termasuk misogini dan non misogini pada data uji yang berjumlah 800 data. Evaluasi tersebut diukur dengan menggunakan *Confusion Matrix* untuk menilai *precision*, *recall*, *F-1 score*, dan *accuracy*. Hasil pengukuran *confusion matrix* divisualisasikan dengan *heatmap* yang memudahkan untuk melihat



data yang teridentifikasi secara tidak tepat dan secara tepat, seperti yang diperlihatkan dalam Gambar 4.17.



**Gambar 4. 17** *Confusion Matrix*

Berdasarkan *heatmap confusion matrix*, maka nilai dari *True Positive Comments* (TPC), *True Negative Comments* (TNC), *False Positive Comments* (FPC), dan *False Negative Comments* (FNC) dapat dihitung dan secara rinci bisa diamati melalui Tabel 4.3 dan Tabel 4.4. Nilai- nilai tersebut selanjutnya dimanfaatkan untuk menentukan nilai *precision*, *recall*, *F1 score*, dan *accuracy* sebagai metrik evaluasi terhadap kinerja model.

**Tabel 4. 3** Perhitungan Data Mengandung Pernyataan Misogini

No.	Komentar Pernyataan Misogini	Jumlah
1.	<i>True Positive Comments</i>	361
2.	<i>True Negative Comments</i>	359
3.	<i>False Positive Comments</i>	34
4.	<i>False Negative Comments</i>	46

**Tabel 4. 4** Perhitungan Data Tidak Mengandung Pernyataan Misogini

No.	Komentar Pernyataan Non Misogini	Jumlah
1.	<i>True Positive Comments</i>	359
2.	<i>True Negative Comments</i>	361
3.	<i>False Positive Comments</i>	46
4.	<i>False Negative Comments</i>	34

Berikut adalah perhitungan metrik evaluasi dengan menggunakan persamaan yang sebelumnya telah dipaparkan pada Bab 2, yaitu persamaan 2.10, 2.11, 2.12, dan 2.13.

**Evaluasi terhadap komentar pernyataan misogini :**

$$Precision = \frac{TPC}{TPC + FPC} \times 100\% = \frac{361}{361 + 34} \times 100\% = \frac{361}{395} \times 100\% = 91.4\%$$

$$Recall = \frac{TPC}{TPC + FNC} \times 100\% = \frac{361}{361 + 46} \times 100\% = \frac{361}{407} \times 100\% = 88.6\%$$

$$F1\ Score = 2 \frac{Precision \times Recall}{Precision + Recall} \times 100\% = 2 \frac{0.914 \times 0.886}{0.914 + 0.886} \times 100\% = 90\%$$

**Evaluasi terhadap komentar pernyataan non misogini :**

$$Precision = \frac{TPC}{TPC + FPC} \times 100\% = \frac{359}{359 + 46} \times 100\% = \frac{359}{405} \times 100\% = 88.6\%$$

$$Recall = \frac{TPC}{TPC + FNC} \times 100\% = \frac{359}{359 + 34} \times 100\% = \frac{359}{393} \times 100\% = 91.4\%$$

$$F1\ Score = 2 \frac{Precision \times Recall}{Precision + Recall} \times 100\% = 2 \frac{0.886 \times 0.914}{0.886 + 0.914} \times 100\% = 90\%$$

**Akurasi terhadap keseluruhan data :**

$$Accuracy = \frac{TPC + TNC}{Total\ Data} \times 100\% = \frac{361 + 359}{800} \times 100\% = \frac{720}{800} \times 100\% = 90\%$$

Berikut adalah rangkuman hasil perhitungan metrik evaluasi terhadap data komentar yang mengandung misogini dan non misogini yang dimuat dalam Tabel 4.5.

**Tabel 4. 5** Hasil Perhitungan Metrik Evaluasi

	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
Pernyataan Misogini (1)	91.4%	88.6%	90%
Pernyataan Non Misogini (0)	88.6%	91.4%	90%

Berdasarkan Tabel 4.5, diperoleh temuan bahwa nilai *precision* komentar non misogini sebesar 88.6%, sedangkan pada komentar misogini nilai *precision* sebesar 91.4%. Nilai tersebut menunjukkan bahwa rasio yang digunakan untuk memprediksi rata-rata 90% akurat dalam mengidentifikasi *True Positive Comment* pada komentar misogini dan non misogini. Sementara itu, nilai *recall* untuk komentar non misogini dan misogini ialah sebesar 91.4% dan 88.6%. Untuk nilai *F1-Score* dari kedua jenis komentar bernilai sama yaitu sebesar 90%. Kedua kelas tersebut menunjukkan bahwa perhitungan rata-rata dari *precision* dan *recall* sudah tepat dan menghasilkan akurasi yang baik, yaitu sebesar 90%. Dengan mengacu pada nilai akurasi tersebut, dapat disimpulkan bahwa kombinasi IndoBERT sebagai *word embedding* dan *Bidirectional Long Short-Term Memory* sebagai metode untuk mengidentifikasikan komentar yang mengandung misogini dan yang tidak mengandung misogini adalah baik.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Dengan merujuk pada penelitian yang sudah diselesaikan, kesimpulan terkait integrasi IndoBERT *Embedding* dengan algoritma *Bidirectional Long Short-Term Memory* dalam mengidentifikasi komentar misogini dapat dijabarkan sebagai berikut:

1. Hasil kinerja IndoBERT sebagai *word embedding* dan *Bidirectional Long Short-Term Memory* sebagai algoritma menunjukkan performa yang baik dalam melakukan identifikasi terhadap komentar misogini dan tidak misogini dengan perolehan akurasi sebesar 90%.
2. Variasi ukuran unit *neuron*, *batch size*, dan *learning rate* berpengaruh pada kinerja model yang dibangun.

#### 5.2. Saran

Terdapat sejumlah poin yang bisa dioptimalkan dengan tujuan meningkatkan kinerja sistem dalam penelitian selanjutnya, yakni:

1. Dengan memperluas data komentar yang mencakup berbagai diksi, topik, dan gaya penulisan yang lebih beragam diharapkan dapat memberi dampak positif terhadap peningkatan kinerja model. Pendekatan ini diharapkan mampu menghasilkan model dengan tingkat akurasi yang lebih tinggi dibandingkan penemuan yang diperoleh dari penelitian yang dikerjakan oleh penulis.
2. Dengan mempertimbangkan pemakaian *emoticon* sebagai salah satu elemen utama dalam mengidentifikasi komentar. Pendekatan ini memungkinkan perbandingan hasil dengan penelitian yang hanya berfokus pada analisis teks.

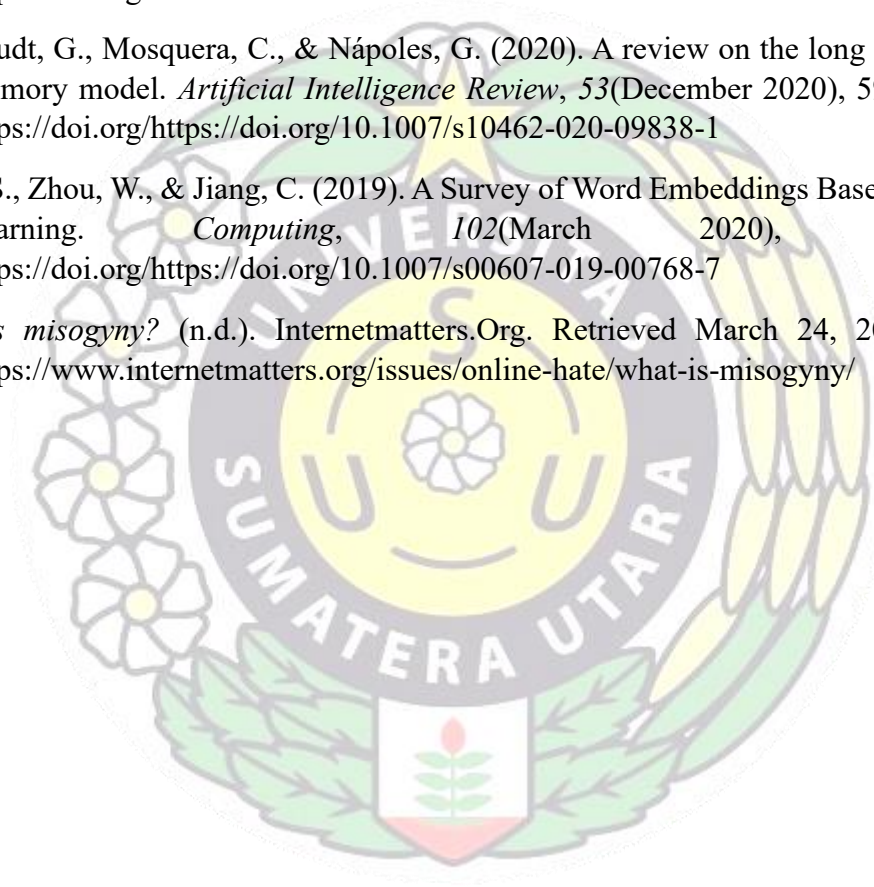
## DAFTAR PUSTAKA

- Ainur Rohman, M., Suhartono, & Chamidy, T. (2023). Bidirectional GRU dengan Attention Mechanism pada Analisis Sentimen PLN Mobile Bidirectional GRU with Attention Mechanism on Sentiment Analysis of PLN Mobile. *Techno.COM*, 22(2), 358–372. <https://doi.org/10.33633/tc.v22i2.7876>
- Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017). Hate speech detection in the Indonesian language: A dataset and preliminary study. *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 233–328. <https://doi.org/https://doi.org/10.1109/ICACSIS.2017.8355039>
- Angeline, R. S., Nurjanah, D., & Nurrahmi, H. (2022, August). Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings. *2022 5th International Conference on Information and Communications Technology (ICOIACT)*. <https://doi.org/https://doi.org/10.1109/ICOIACT55506.2022.9972171>
- Birunda, S. S., & Devi, R. K. (2020). A Review on Word Embedding Techniques for Text Classification. In J. S. Raj, A. M. Iliyasu, R. Bestak, & Z. A. Baig (Eds.), *Innovative Data Communication Technologies and Application* (pp. 267–281). Springer. [https://doi.org/https://doi.org/10.1007/978-981-15-9651-3\\_23](https://doi.org/https://doi.org/10.1007/978-981-15-9651-3_23)
- Damanik, A. J. (2021). IDENTIFIKASI PERNYATAAN MISOGINI BERBAHASA INDONESIA BERDASARKAN KOMENTAR YOUTUBE MENGGUNAKAN GLOVE EMBEDDING DAN RANDOM FOREST CLASSIFIER. In *Universitas Sumatera Utara*.
- Devi, M. D., & Saharia, N. (2020). Misogynous Text Classification Using SVM and LSTM. In D. Garg, K. Wong, J. Sarangapani, & S. K. Gupta (Eds.), *International Advanced Computing Conference 2020* (pp. 336–348). Springer Singapore. <https://doi.org/https://doi.org/10.1007/978-981-16-0401-0>
- Faisal, D. R., & Mahendra, R. (2022). *Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets*. <https://arxiv.org/abs/2206.15359>
- Fuad, B., Amiruddin, M., Yentriyani, A., Kanti, D., Qibtiyah, A., Adkiras, F., Nugroho, H. A., Anshor, M. U., Nahe'i, Novianti, Salampessy, O. C., Hutabarat, R. M., Ratnawati, R., Mashudi, S., Tardi, S. A., Iswarini, T. S. E., Wiandani, T., & Sitohang, V. (2023). *CATAHU 2023: Kekerasan terhadap Perempuan di Ranah Publik dan Negara: Minimnya Perlindungan dan Pemulihan*. <https://komnasperempuan.go.id/catatan-tahunan-detail/catahu2023-kekerasan-terhadap-perempuan-di-ranah-publik-dan-negara-minimnya-perlindungan-dan-pemulihan>
- Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, 10, 19083–19095. <https://doi.org/10.1109/ACCESS.2022.3151048>



- Hidayah, A. (2022, August 15). *Femisida Terjadi Karena Pelaku Merasa Superior dan Misogini Terhadap Perempuan*. Konde.Co. <https://www.konde.co/2022/08/femisida-terjadi-karena-pelaku-merasa-superior-dan-misogini-terhadap-perempuan/>
- Imrana, Y., Xiang, Y., Ali, L., & Abdul-Rauf, Z. (2021). A bidirectional LSTM deep learning approach for intrusion detection. *Expert Systems with Applications*, 185. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115524>
- Isnain, A. R., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14, 169–178. <https://doi.org/https://doi.org/10.22146/ijccs.51743>
- Kemp, S. (2023, February 9). *Digital 2023: Indonesia*. We Are Social. <https://datareportal.com/reports/digital-2023-indonesia>
- Kurniasih, A., & Parningotan Manik, L. (2022). On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(6). <https://doi.org/10.14569/IJACSA.2022.01306109>
- Lubis, F. (2021, October 28). *Seksisme dan Misogini dalam Perspektif HAM*. Komnas HAM. <https://www.komnasham.go.id/index.php/news/2021/10/28/1963/seksisme-dan-misogini-dalam-perspektif-ham.html>
- Mahadevaswamy, U. B., & Swathi, P. (2023). Sentiment Analysis using Bidirectional LSTM Network. *Procedia Computer Science*, 218, 45–56. <https://doi.org/10.1016/j.procs.2022.12.400>
- Maharani, A. P. (2022, November 10). *Misogini: Kebencian Ekstrim Hanya Karena Kita Perempuan*. Lembaga Pers Mahasiswa Gelora Sriwijaya. <https://gelorasriwijaya.co/blog/misogini-kebencian-ekstrim-hanya-karena-kita-perempuan/> -
- Merriam-Webster. (n.d.). *Misogyny*. Merriam-Webster.Com Dictionary. Retrieved February 16, 2024, from <https://www.merriam-webster.com/dictionary/misogyny>
- MinTek. (2023, November 27). *Mengenal Machine Learning: Pengertian dan Penerapannya di Kehidupan*. StadiTek. <https://www.staditek.id/insight/arti-machine-learning/>
- National Democratic Institute. (2019). *Tweets That Chill: Analyzing Online Violence Against Women in Politics*. <https://www.ndi.org/tweets-that-chill>
- Nuryanto, H. (n.d.). *Pengertian Identifikasi: Proses, Bentuk, dan Contohnya*. Gramedia. Retrieved March 22, 2024, from <https://www.gramedia.com/literasi/identifikasi/>

- Pertiwi, S. P. (2023, October 2). *Kamus Feminis: Apa Itu Misogini Atau Kebencian terhadap Perempuan*. Konde.Co. <https://www.konde.co/2023/10/kamus-feminis-apa-itu-misogini-atau-kebencian-terhadap-perempuan/>
- Srivastava, K., Chaudhury, S., Bhat, P., & Sahu, S. (2017). Misogyny, feminism, and sexual harassment. *Industrial Psychiatry Journal*, 26(2), 111–113. [https://doi.org/10.4103/ipj.ipj\\_32\\_18](https://doi.org/10.4103/ipj.ipj_32_18)
- Tri Wibowo, B., Nurjanah, D., & Nurrahmi, H. (2023). Identification of Misogyny on Social Media in Indonesian Using Bidirectional Encoder Representations from Transformers (BERT). *5th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2023*, 401–406. <https://doi.org/10.1109/ICAIIIC57133.2023.10067106>
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(December 2020), 5929–5955. <https://doi.org/https://doi.org/10.1007/s10462-020-09838-1>
- Wang, S., Zhou, W., & Jiang, C. (2019). A Survey of Word Embeddings Based on Deep Learning. *Computing*, 102(March 2020), 717–740. <https://doi.org/https://doi.org/10.1007/s00607-019-00768-7>
- What is misogyny?* (n.d.). Internetmatters.Org. Retrieved March 24, 2024, from <https://www.internetmatters.org/issues/online-hate/what-is-misogyny/>



## LAMPIRAN

No.	Komentar	Label	Hasil Identifikasi
1	Saya cowo yang belum pernah instal tiktok sama sekali dari dulu, di tiktok bukan cuma banyak konten joget joget gajelas ajh, tapi banyak juga konten brainwashing, dan penggiringan opini, banyak rumah tangga yang hancur karena istri kena brainwashing dari konten kreator cewe yang selalu ngeluarin statement dongo nya, dan akhirnya di serap juga oleh wanita yang pemikiran nya cetek	1	0
2	Bangun tdr pembahasannya berat bousss??	0	1
3	Makanya aurat ditutup... Dan ajarkan agama sedari dini... Klo udah kejadian kan banyak ahli2 dadakan tuh ngomong di media.. Klo perlu ubah syariat sekalian..biar gak ada itu kejahatan.	1	0
4	@TofaTofa_id Itulah kenapa perempuan ga boleh jd pemimpin mudah baperan apalagi betilafea ini	1	0
5	Kalo dia cantik tapi ngerasa kurang terus gw apaan anjing	0	1
6	Kunci sbnrnya dr perselingkuhan adalah dr pihak perempuan, krn kalau perempuan tegas menolak pasti tdk akan pnh ada yg namanya selingkuh segila apapun laki2	1	0
7	lebih ga enak ga ada duit daripada ga ada pasangan ????	0	1
8	Wajah ny Cantik bgt Pas memang tuk Pemeran Film Dewasa	1	0
9	neng awas kesapu ombak;((	0	1
10	Asupan susu yg bergizi bagi para lelaki ya ini....lanjutkan..	1	0
11	Roknya darimana itu mbak? Spill dong hehe	0	1

No.	Komentar	Label	Hasil Identifikasi
12	Terserah ellooo, emang loo siapa,loo bukan siapa siapa.	1	0
13	Waaaoooww pastinya suami bahagia n puas bnget tuh	1	0
14	Ngejar kamu tuh kek partai banteng merah?? Penuh perjuangan	0	1
15	Soal kekerasan seksual sebenarnya bisa dicegah hanya saja perempuan2 sekarang kebanyakan tidak mau menjaga diri .inginnya menonjolkan diri?? wajar dong . "ikan tertangkap karena ada umpan"	1	0
16	Jangan ketik 9 ya guys POKOKNYA JANGAN!!!	0	1
17	keren pertanyaannya terakhirnya saat ada kehidupan berikutnya dan Tuhan memberi kebebasan untuk memilih jadi perempuan atau laki laki... eh sih para perempuan milih jadi perempuan lagi, heheheheheh.... katanya enak jadi laki laki lha koq milih perempuan lagi . sudah terjawab semuanya knpa wanita tidak bisa menjadi pemimpin, krna pola pikirnya membagongkan ....	1	0
18	Baru tau nitvior punya boba	1	0
19	Ambil perjaka saya kakk aku ihlas kok	1	0
20	Biasanya pake jaket, sekalinya keluar pake jilbab cantik bgtt aaaaaa	0	1
21	Km mau d unboxing...?	1	0
22	segala bahas kutus kutus ????????	0	1
23	Pantes Buaya di sungai Pada sunyi Ternyata pada Ngumpul di sini	0	1
24	Nggak ada bakat jadi artis ny. Numpang cari cuan dari artis yg lagi naik daun. Sorry ya gue nggak follow you tp ntah knpa gue gerah liat posting elu yg numpang lewat beranda. Nggak berfaedah	1	0





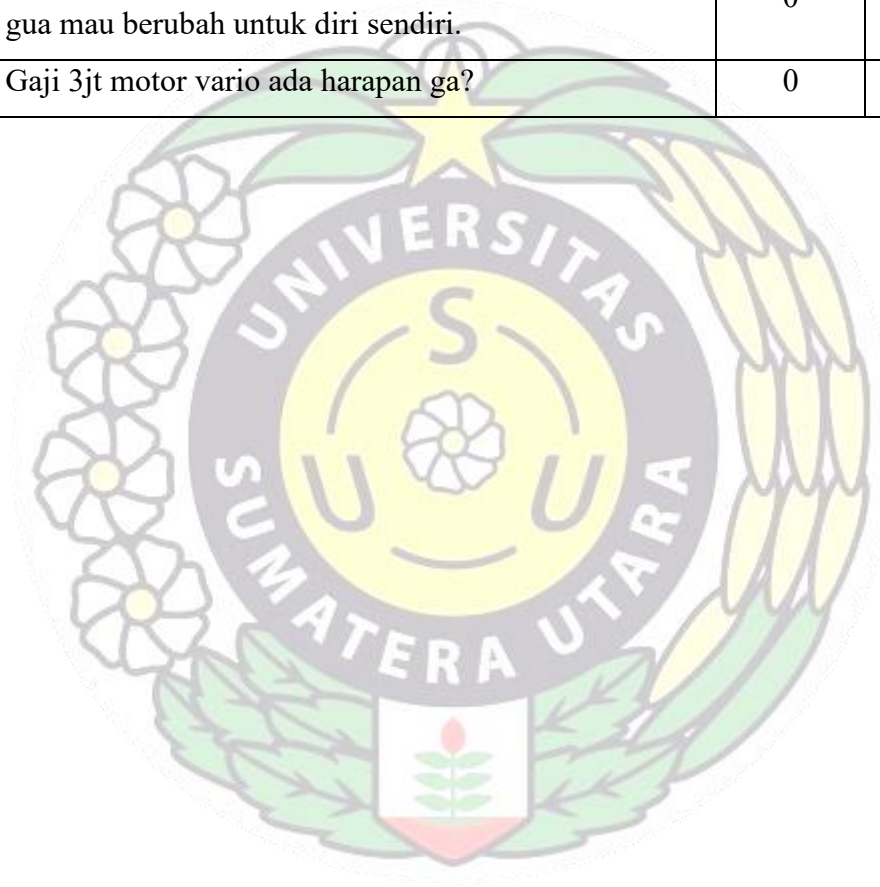
No.	Komentar	Label	Hasil Identifikasi
38	Bentuknya aja kek gitu, dibilang OpenBo aja ngga terima..?? Oyyy,,, Anak sd aja bisa menilai bodoh!!	1	0
39	omongannyu tinggi tapi memang fakta sih ^^	0	1
40	Seneng banget lihatnya hot ?????	1	0
41	Perempuan kalo ngomong ngasal anjir, jadi kalo ngegosip bisa ngarang cerita dari A ampe Z dan semuanya negatif semuanya. Biadab emang	1	0
42	Udahlah bro gausah basa basi kita meski bukan sodara atau temen deket pasti tau lah yakan ??	1	0
43	Kaget anjing nya	0	1
44	Gw kalo berhadapan ama perempuan, biasanya harus siap dengan pikiran liar dan asumsi di kepala mereka yg entah nggak tahu kenapa bisa ada dan dikeluarin begitu aja	1	0
45	Artis yang gak pernah di gosipin aneh	0	1
46	semoga kedepannya para wanita mudah di atur / di arahkan ke jalan yg benar.	1	0
47	Wajahnya tetap yg memukau,udelnnya biasa aja	1	0
48	Kuliah jangan tinggi tingi , cewe harus cepet nikah sebelum 25 thn astaga kenapa ya orang bisa bisanya mikir gitu ??	0	1
49	Punya bentuk payudara yang kecil selalu dapat hinaan, yang terlalu besar pun sama. Tetapi terimalah dirimu sendiri karena itu adalah karunia Tuhan.	0	1
50	kalo nafsu birahi cowok pada naik, yang disalahin malah ceweknya? cowok yang gabisa menundukkan pandangan dan cewek jadi korban pelecehan juga salah cewek?	0	1

No.	Komentar	Label	Hasil Identifikasi
51	Perlu gua email ketua nya gak sih? Bahaya banget lonte ada di organisasi	1	0
52	ewe lah cok, ngapain direkam rekam	1	0
53	Aku nggak suka pake alas kaki. Buat apa aku pakai alas kaki kalau bukan kamu alasanku melangkah lagi?? dan juga bau amis ih,iya amis.amis you maksudnya??????	0	1
54	Sebagai cowok aku gak relate sama mbak-mbaknya, tapi kok jadi ikutan kesel sama cowok cowok yang mulutnya lemes gitu ya hhh	0	1
55	Itu susu tjap nona ya...atau susu tjap dencow...??????????	1	0
56	Marvel sekarang mengecewakan, sekarang pahlawan ceweknya di tambah biar kelihatan dukung hak perempuan dan mengecewakan banyak orang	1	0
57	Mungkin kalo laki" terasa ingin lebih terlihat kuat itu hanya diluaran mbak, lihat saja di dalam keluarga, berapa banyak suami yang seakan gak ada harganya di hadapan istri, jd para laki" itu gak mau terlihat lemah begitu diluar,,, mungkin..	0	1
58	makasi kak aku udh keluar	1	0
59	wanita terkorup sri mulyani. yang tanda tangan BLBI cair. ?????? AHY jadi tersangka Kpk ,sri mulyani dibuang ke bank dunia Antasari ashar status ketua KPK difitnah kasus otak pembunuhan pemain golf,nazarudin bebas jadi buka suara yg ngaco ??????	1	0
60	Sok keren ngomong Inggris tp g paham audien org indo noob	1	0
61	Kita bukan teman bukan sodara tapi kita semua 1 pemikiran. Benar bukan:(v	1	0
62	Emang loh siapa sherly ??????????bodut loh?????????	1	0
63	Mendadak tembus pandang mata gw	1	0

No.	Komentar	Label	Hasil Identifikasi
64	Malam pertama meninggal sih ini vior	1	0
65	Apa karna iri jadi pada banyak bacot	0	1
66	Jdi cewek jgn bergaul sama banyak cowo,kek cewe gk bener aja.	1	0
67	Enaknya jadi perempuan ???? Punya tempat khusus dari fasilitas umum, bisa nolep di rumah saja, gk perlu nembak pasangan, gk perlu sukses dulu, gk perlu kaya untuk menghidupi keluarga, gk perlu jadi ketua, gk perlu ngapa-ngapain....	1	0
68	Kalau ditinggalin cowok karena ngerasa kita terlalu hebat atau mereka ngerasa tersaingi sama kita, seksism bukan sih?	0	1
69	aku kok bisa nyasar ke sini?	0	1
70	gak ada bagus2nya bilangin org lonte, gak sopan	0	1
71	Mbaaakk... Banyak orang jual semangka di luaraann tapi ini yg saya liat paling mengkell ????????	1	0
72	Bodo amat sherly ????	1	0
73	Syang ih dh dbilang ge jan post poto yg ini masihh yaa ngeyel dbilangin??	1	0
74	Puan2 paket komplit.	0	1
75	@Abi_Benedict @neohistoria_id mata lo pujian. saya perempuan dan itu bentuk pelecehan pada wanita. lu coba ngomong ama mak lu kalau punya tobrut kalau nggak dipenggal lehermu ya dikubur hidup2.	0	1
76	""Perempuan harus lengkap seperti apa yg telah di tentukan. Kalaupun tidak lengkap karna mungkin sering melawan kodrat ??"	1	0



No.	Komentar	Label	Hasil Identifikasi
77	istilah consent ga compatible buat masyarakat indonesia kata si gamal, yaiyalah wong wanita cuma dianggap objek meskipun istri jd budak penyalur nafsu nolak bersetubuh nanti dianggap wanita durhaka	0	1
78	Kalo bacotnya berisi kek gini sih aku sukaaa wkwkwk ???	0	1
79	Gua selalu di suruh berubah demi laki-laki padahal nih gua mau berubah untuk diri sendiri.	0	1
80	Gaji 3jt motor vario ada harapan ga?	0	1





**KEPUTUSAN  
DEKAN FAKULTAS ILMU KOMPUTER  
DAN TEKNOLOGI INFORMASI  
NOMOR : 160/UN5.2.14.D/SK/HK.07/2025  
DEKAN FAKULTAS ILMU KOMPUTER  
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA**

- Membaca : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 03 Januari 2025 perihal permohonan ujian skripsi:  
Nama : STEPHANI ULI BASA SILITONGA  
NIM : 201402068  
Program Studi : Sarjana (S-1) Teknologi Informasi  
Judul Skripsi : Identifikasi Pernyataan Misogini Berdasarkan Komentar Media Sosial Menggunakan Bidirectional Long Short-Term Memory dan Indobert Embedding
- 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 2024/2025.
- 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 : Ainul Hizriadi, S.Kom., M.Sc.  
NIP: 198510272017061001  
Sekretaris : Fahrurrozi Lubis B.IT., M.Sc.IT  
NIP: 198610122018052001  
Anggota Penguji : Umayya Ramadhani Putri Nasution, S.TI., M.Kom.  
NIP: 199104112024062001  
Anggota Penguji : Sarah Purnamawati, S.T., M.Sc.  
NIP: 198302262010122003  
Moderator : -  
Panitera : -
- Kedua : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2025.
- 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.

Medan, 10 Januari 2025  
Ditandatangani secara elektronik oleh:  
Dekan



Maya Silvi Lydia  
NIP 197401272002122001

Tembusan :

1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi
2. Yang bersangkutan