

**KLASIFIKASI KONTEN EKSPLISIT PADA LIRIK LAGU  
BAHASA INDONESIA DAN BAHASA INGGRIS  
MENGUNAKAN METODE *MBERT***

**SKRIPSI**



**OLEH:**

**FARIDAH**

**NIM. 1810817220015**

**PROGRAM STUDI TEKNOLOGI INFORMASI  
FAKULTAS TEKNIK  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARMASIN, JUNI 2025**

**KLASIFIKASI KONTEN EKSPLISIT PADA LIRIK LAGU  
BAHASA INDONESIA DAN BAHASA INGGRIS  
MENGUNAKAN METODE *MBERT***

**Skripsi**

Diajukan Untuk Memenuhi Salah Satu Syarat  
Sarjana Strata-1 Teknologi Informasi



**OLEH:**

**FARIDAH**

**NIM. 1810817220015**

**PROGRAM STUDI TEKNOLOGI INFORMASI  
FAKULTAS TEKNIK  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARMASIN, JUNI 2025**

## PERNYATAAN

Saya yang bertanda tangan di bawah ini,

Nama : Faridah

NIM : 1810817220015

Fakultas : Teknik

Program Studi : Teknologi Informasi

Judul Tugas Akhir : Klasifikasi Konten Eksplisit Pada Lirik Lagu Bahasa Indonesia Dan Bahasa Inggris Menggunakan Metode mBERT

Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom.

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, 02 Mei 2025



FARIDAH

NIM. 1810817220015

**LEMBAR PENGESAHAN**

**SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI**

**Klasifikasi Konten Eksplisit Pada Lirik Lagu Bahasa Indonesia Dan Bahasa Inggris Menggunakan Metode MBERT**

**Oleh**

**Faridah (1810817220015)**

Telah dipertahankan di depan Tim Penguji pada 10 Juni 2025 dan dinyatakan

**LULUS**

**Komite Penguji :**  
**Ketua** : Dr. Ir. Yuslena Sari, S.Kom., M.Kom  
NIP. 198411202015042002  
**Anggota 1** : Muhammad Fajrian Noor, S.Kom., M.Kom  
NIP. 199611092023211009  
**Anggota 2** : Nurul Fathanah Mustamin, S.Pd., M.T.  
NIP. 199110252019032018  
**Pembimbing Utama** : Andreyan Rizky Baskara, S.Kom., M.Kom.  
NIP. 199307032019031011



.....  
.....  
.....  
.....


23 JUN 2025

Banjarbaru, .....

Diketahui dan disahkan oleh:

**Wakil Dekan Bidang Akademik**

**Fakultas Teknik ULM,**



**Dr. Ir. Mahmud S.T., M.T.**  
NIP. 197401071998021001

**Koordinator Program Studi**

**S-1 Teknologi Informasi,**



**Andreyan Rizky Baskara, S.Kom., M.Kom.**

NIP. 199307032019031011

**PERSETUJUAN TUGAS AKHIR**

**KLASIFIKASI KONTEN EKSPLISIT PADA LIRIK LAGU BAHASA INDONESIA DAN  
BAHASA INGGRIS MENGGUNAKAN METODE MBERT**

**OLEH  
FARIDAH  
NIM. 1810817220015**

**Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi, dan disetujui untuk  
dipertahankan di hadapan dewan penguji**

**Banjarmasin, 28 Mei 2025  
Pembimbing Utama,**



**Andreyan Rizky Baskara, S.Kom., M.Kom.  
NIP. 199307032019031011**

## ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi otomatis guna mendeteksi konten eksplisit pada lirik lagu berbahasa Indonesia dan Inggris. Lirik lagu sering kali memuat unsur eksplisit baik secara langsung maupun tersirat, namun klasifikasi otomatis terhadap jenis konten ini masih terbatas, khususnya untuk kasus yang melibatkan lebih dari satu bahasa. Penelitian sebelumnya umumnya hanya berfokus pada satu bahasa sehingga kurang relevan dalam konteks globalisasi konten digital. Dalam penelitian ini, penulis menggunakan model *Multilingual BERT (mBERT)* yang dirancang untuk memahami berbagai bahasa dalam satu arsitektur. Dataset yang digunakan terdiri dari 33.420 lirik lagu yang dikumpulkan dari penelitian terdahulu dan repositori publik di Kaggle. Tahapan *preprocessing* mencakup *case folding*, *text cleaning*, *undersampling* (untuk kondisi data seimbang), *label encoding*, dan *tokenisasi*. Data dibagi dengan rasio 80:20 dan 70:30, menggunakan dan tanpa stratified sampling. Model kemudian dievaluasi melalui delapan variasi pengujian berdasarkan kombinasi rasio pembagian, keseimbangan distribusi label, dan penggunaan stratifikasi. Model dilatih dengan pendekatan *transfer learning* berbasis mBERT dan dioptimalkan melalui *grid-search* untuk menentukan parameter terbaik, yaitu *batch size* 32, *epoch* 7–10, dan *dropout rate* 0,2–0,3. Hasil pengujian menunjukkan bahwa kondisi data seimbang dengan rasio 80:20 menghasilkan performa terbaik dengan *f1-score* sebesar 83,99%, sedangkan pada data tidak seimbang mencapai 66,53%. Analisis lebih lanjut menunjukkan bahwa meskipun model cukup efektif dalam mengenali konten eksplisit secara tekstual, kemampuannya masih terbatas dalam mendeteksi konten eksplisit yang bersifat implisit. Temuan ini menunjukkan perlunya pengayaan konteks semantik dalam pengembangan model lanjutan yang lebih adaptif terhadap nuansa bahasa.

*Kata Kunci: Klasifikasi Teks, Konten Eksplisit, mBERT, Multibahasa, NLP*

## **ABSTRACT**

*This study aims to develop an automatic classification system to detect explicit content in song lyrics written in Indonesian and English. Song lyrics often contain explicit elements either directly or implicitly, yet automatic classification for such content remains limited, especially in cases involving more than one language. Previous studies have typically focused on a single language, making them less applicable in the context of increasingly global digital content. This research employs the Multilingual BERT (mBERT) model, which is designed to handle various languages within a single architecture. The dataset consists of 33,420 song lyrics collected from previous research and public repositories on Kaggle. The preprocessing stages include case folding, text cleaning, undersampling (for balanced data), label encoding, and tokenization. The data was split using 80:20 and 70:30 ratios, both with and without stratified sampling. The model was evaluated through eight variations of testing based on combinations of data split ratios, label distribution balance, and the use of stratification. Training was performed using a transfer learning approach based on mBERT and optimized using grid search to determine the best parameters, namely batch size of 32, 7–10 epochs, and a dropout rate of 0.2–0.3. The best performance was achieved with balanced data using an 80:20 split, resulting in an F1-score of 83.99%, while the imbalanced data yielded 66.63%. Further analysis indicates that while the model is effective in detecting explicitly stated content, it remains limited in recognizing implicitly conveyed explicit content. These findings highlight the importance of semantic context enrichment in the future development of more adaptive models capable of handling nuanced language.*

*Keywords: Explicit Content, mBERT, Multilingual, NLP, Text Classification*

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, kesehatan, dan kekuatan yang telah diberikan, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Konten Eksplisit pada Lirik Lagu Bahasa Indonesia dan Bahasa Inggris Menggunakan Metode mBERT” dengan baik. Shalawat dan salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, suri teladan umat manusia, yang telah membawa kita dari zaman kegelapan menuju era penuh ilmu pengetahuan dan kemajuan teknologi.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 (S1) pada Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Lambung Mangkurat, Banjarmasin. Penelitian ini secara khusus diarahkan untuk menjawab salah satu tantangan di bidang Teknologi Informasi, yaitu kebutuhan akan sistem klasifikasi teks yang mampu bekerja secara efektif pada data multibahasa. Dalam hal ini, penulis menggunakan pendekatan *Multilingual BERT* (mBERT) sebagai model berbasis pembelajaran mendalam yang memungkinkan pemrosesan lirik dalam bahasa Indonesia dan Inggris secara simultan, tanpa perlu model terpisah untuk masing-masing bahasa.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem penyaringan konten berbasis teks, terutama dalam konteks moderasi konten musik digital yang kian berkembang. Meskipun dalam pelaksanaannya penulis menghadapi sejumlah tantangan, seperti pengelolaan data dalam jumlah besar, ketidakseimbangan kelas, serta keterbatasan perangkat saat proses pelatihan model, namun hal tersebut menjadi bagian dari proses pembelajaran yang sangat berharga.

Penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis menyampaikan penghargaan kepada:

1. Rektor Universitas Lambung Mangkurat, Bapak Prof. Dr. Ahmad, S.E., M.Si. atas kepemimpinan beliau di Universitas Lambung Mangkurat.
2. Bapak Prof. Dr. Ir. Iphan Fitriani Radam, S.T., M.T., IPU, selaku Dekan Fakultas Teknik Universitas Lambung Mangkurat, yang telah memberikan izin dan dukungan dalam proses penyusunan skripsi ini.

3. Bapak Andreyan Rizky Baskara, S.Kom., M.Kom., selaku Koordinator Program Studi Teknologi Informasi sekaligus pembimbing utama, yang dengan sabar telah memberikan arahan, bimbingan, dan motivasi kepada penulis.
4. Seluruh dosen dan staf Program Studi Teknologi Informasi yang telah memberikan ilmu, bimbingan akademik, dan bantuan teknis maupun administratif selama masa studi.
5. Diri penulis sendiri, Faridah, yang telah mencurahkan segala daya, upaya, dan tekad dalam penyelesaian skripsi ini.
6. Orang tua tercinta, Umi dan Abi, beserta kedua adik penulis, Raisya dan Ahmad Zain, serta seluruh keluarga besar yang tak henti memberikan doa, dukungan moral, dan materiil.
7. Teman-teman penulis: Alya Nurul Jannah, Tri Aryanto Ramadhana, Rizka Ardiyanti, Muhammad Tri Madya Lestiyanto, Ananda, Silvia Handayani, dan Taufik Nurhidayat, yang telah banyak membantu, memberikan semangat, dan bahkan meminjamkan perangkat untuk mendukung kelancaran penyusunan skripsi ini.
8. Teman-teman seperjuangan angkatan 2018 di Program Studi Teknologi Informasi, atas kebersamaan, semangat, dan kerja sama yang tidak ternilai selama masa perkuliahan.
9. Seluruh pihak yang telah memberikan dukungan dan bantuan, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu.

Skripsi ini terdiri dari beberapa bab dengan sistematika penulisan sebagai berikut:

- BAB I: Pendahuluan, yang mencakup latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian.
- BAB II: Tinjauan Pustaka, yang berisi landasan teori, penelitian terdahulu, dan kerangka pemikiran.
- BAB III: Metodologi Penelitian, yang menjelaskan metode yang digunakan dalam penelitian.

- BAB IV: Hasil dan Pembahasan, yang menyajikan hasil penelitian serta analisisnya.
- BAB V: Kesimpulan dan Saran, yang berisi ringkasan hasil penelitian dan rekomendasi untuk pengembangan selanjutnya.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang pemrosesan bahasa alami dan klasifikasi konten berbasis teks, serta menjadi referensi bagi pembaca yang membutuhkan.

Banjarmasin, 2 Mei 2025

Penulis,

FARIDAH

## DAFTAR ISI

SAMPUL DALAM.....	i
PERNYATAAN.....	ii
LEMBAR PENGESAHAN.....	iii
PERSETUJUAN TUGAS AKHIR.....	iv
ABSTRAK.....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xvii
DAFTAR LAMPIRAN.....	xx
DAFTAR RIWAYAT HIDUP.....	xxi
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait.....	7
2.1.1 <i>Explicit Content Detection in Music Lyrics Using Machine Learning</i> (2018).....	7
2.1.2 <i>Detecting explicit lyrics: a case study in Italian music</i> (2022).....	7
2.1.3 <i>Klasifikasi Konten Eksplisit Pada Lirik Lagu Berbahasa Indonesia</i> <i>Menggunakan Metode LSTM-CNN</i> (2024).....	8
2.1.4 <i>Improving Indonesian Text Classification Using Multilingual</i> <i>Language Model</i> (2020).....	9

2.1.5 <i>A Comparative Study of Transformer-based Models for Hate-Speech Detection in English-Kiswahili Code-Switched Social Media Text (2024)</i> .....	10
2.1.6 <i>Multilingual Sentiment Analysis in Tamil, Malayalam, and Kannada code-mixed social media posts using MBERT (2021)</i> .....	11
2.2 Landasan Teori.....	15
2.2.1 <i>Music Information Retrieval</i> .....	15
2.2.2 Lirik Lagu Eksplisit.....	15
2.2.3 Undang-Undang dan Konten Eksplisit pada Lirik Lagu Indonesia..	17
2.2.4 <i>Natural Language Processing</i> .....	19
2.2.5 Klasifikasi Teks.....	20
2.2.6 <i>Data Sampling</i> .....	21
2.2.7 <i>Stratified Sampling</i> .....	22
2.2.8 <i>Text Preprocessing</i> .....	24
2.2.9 <i>Word Embeddings</i> .....	25
2.2.10 <i>Transformer</i> .....	27
2.2.11 <i>Bidirectional Encoder Representations from Transformers (BERT)</i> .....	29
2.2.12 <i>Multilingual Bidirectional Encoder Representations from Transformers (mBERT)</i> .....	32
2.2.13 <i>Overfitting</i> .....	34
2.2.14 <i>Confusion Matrix</i> .....	35
2.3 Komparasi Metode.....	37
2.4 Kerangka Pemikiran.....	39
<b>BAB III METODOLOGI PENELITIAN</b>	
3.1 Alat dan Bahan.....	41
3.1.1 Alat Penelitian.....	41
3.1.2 Bahan Penelitian.....	42
3.2 Alur Penelitian.....	43

3.2.1 Identifikasi Masalah.....	43
3.2.2 Studi Literatur.....	43
3.2.3 Pengumpulan Data.....	44
3.2.4 Implementasi Model.....	45
3.2.5 Pengujian.....	52
3.2.6 Pengukuran Performansi.....	54
3.2.7 Analisis Hasil.....	54
3.2.8 Implementasi Sistem.....	54

#### BAB IV HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data.....	56
4.2 Pelabelan Data.....	56
4.3 <i>Text Preprocessing</i> .....	57
4.3.1 <i>Case Folding</i> .....	58
4.3.2 <i>Text Cleaning</i> .....	58
4.3.3 <i>Undersampling Dataset</i> .....	59
4.3.4 <i>Label Encoding</i> .....	60
4.3.5 <i>Data Splitting</i> .....	61
4.4 Implementasi Model mBERT.....	62
4.4.1 Persiapan dan Pemuatan Dataset.....	62
4.4.2 <i>Preprocessing</i> Data: Tokenisasi.....	63
4.4.3 Inisialisasi dan Konfigurasi Model mBERT.....	64
4.4.4 Proses Pelatihan ( <i>Fine-Tuning</i> ).....	65
4.4.5 Proses Pengujian <i>K-Fold Cross Validation</i> .....	69
4.4.6 Evaluasi Model pada Data <i>Testing</i> .....	70
4.5 Desain Eksperimen.....	72
4.6 Pengujian Model.....	74
4.6.1 Pengujian Pada Dataset Tidak Seimbang dengan Rasio Data 80:20.....	74
4.6.2 Pengujian Pada Dataset Tidak Seimbang dengan Rasio Data 70:30.....	78

4.6.3 Pengujian Pada Dataset Seimbang dengan Rasio Data 80:20.....	81
4.6.4 Pengujian Pada Dataset Seimbang dengan Rasio Data 70:30.....	85
4.6.5 Pengujian dengan <i>Stratified Sampling</i> .....	88
4.7 Analisis Hasil dan Evaluasi Model Terbaik.....	103
4.7.1 Perbandingan Antar Skenario.....	104
4.7.2 Pengaruh Distribusi Data (Seimbang vs Tidak Seimbang).....	105
4.7.3 Pengaruh <i>Stratified Sampling</i> .....	105
4.7.4 Pengaruh Rasio Data <i>Training</i> dan <i>Testing</i> .....	106
4.7.5 Model Terbaik dan Evaluasi Akhir.....	106
4.7.6 Analisis Kesalahan ( <i>Error Analysis</i> ).....	107
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	117
5.2 Saran.....	117
DAFTAR PUSTAKA.....	119
LAMPIRAN.....	127

## DAFTAR TABEL

Tabel 2.1 Ringkasan penelitian terkait dan keterkaitannya.....	13
Tabel 2.2 <i>Confusion matrix</i> pada klasifikasi biner.....	36
Tabel 2.3 Ringkasan hasil perbandingan penelitian ini dengan penelitian lain.....	38
Tabel 3.1 Alat penelitian.....	41
Tabel 3.2 Contoh data lirik lagu beserta labelnya.....	45
Tabel 3.3 Proses <i>case folding</i> .....	49
Tabel 3.4 Proses <i>text cleaning</i> .....	49
Tabel 3.5 Proses <i>stopwords removal</i> .....	49
Tabel 3.6 Proses <i>stemming</i> .....	50
Tabel 4.1 Contoh perubahan label pada dataset berbahasa Inggris.....	57
Tabel 4.2 Karakteristik label data [19].....	57
Tabel 4.3 Kombinasi <i>Hyperparameter</i> .....	73
Tabel 4.4 Skenario Pelatihan.....	73
Tabel 4.5 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 80:20.....	75
Tabel 4.6 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>imbalanced</i> 80:20.....	76
Tabel 4.7 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>imbalanced</i> 80:20... 77	77
Tabel 4.8 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 70:30.....	78
Tabel 4.9 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>imbalanced</i> 70:30.....	80
Tabel 4.10 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>imbalanced</i> 70:30. 80	80
Tabel 4.11 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 80:20.....	82
Tabel 4.12 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>balanced</i> 80:20.....	83
Tabel 4.13 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>balanced</i> 80:20.....	84

Tabel 4.14 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 70:30.....	85
Tabel 4.15 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>balanced</i> 70:30.....	87
Tabel 4.16 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>balanced</i> 70:30.....	88
Tabel 4.17 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	89
Tabel 4.18 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	91
Tabel 4.19 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	91
Tabel 4.20 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	93
Tabel 4.21 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	94
Tabel 4.22 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	95
Tabel 4.23 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	96
Tabel 4.24 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	98
Tabel 4.25 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	99
Tabel 4.26 Nilai metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 70:30 dengan <i>stratify</i> .....	100
Tabel 4.27 Hasil evaluasi model terbaik pada data <i>training</i> dan <i>testing</i> pada skenario <i>balanced</i> 70:30 dengan <i>stratify</i> .....	101
Tabel 4.28 Hasil evaluasi <i>k-fold</i> model terbaik pada skenario <i>balanced</i> 70:30 dengan <i>stratify</i> .....	102
Tabel 4.29 Ringkasan performa model dari seluruh skenario pengujian.....	103

Tabel 4.30 Rekap performa model pada metrik utama dengan <i>stratified sampling</i> .....	104
Tabel 4.31 Rekap performa model pada tiap metrik utama tanpa <i>stratified sampling</i> .....	104
Tabel 4.32 Distribusi prediksi model terhadap label sebenarnya.....	107
Tabel 4.33 Distribusi prediksi model yang tidak berhasil diklasifikasikan dengan benar.....	108
Tabel 4.34 Distribusi prediksi model yang berhasil diklasifikasikan dengan benar.....	109

## DAFTAR GAMBAR

Gambar 1.1 Gambaran label eksplisit pada lagu di platform Spotify.....	2
Gambar 2.1 Arsitektur model <i>transformer</i> [42].....	28
Gambar 2.2 (kiri) <i>Multi-Head Attention</i> yang terdiri dari beberapa lapisan <i>attention</i> yang berjalan secara paralel. (kanan) <i>Scaled Dot-Product Attention</i> [42].....	29
Gambar 2.3 Representasi <i>input</i> BERT. <i>Embedding input</i> adalah jumlah dari <i>token embeddings</i> , <i>segment embeddings</i> , dan <i>position embeddings</i> [12].....	30
Gambar 2.4 Keseluruhan prosedur <i>pre-training</i> dan <i>fine-tuning</i> untuk BERT [12].....	31
Gambar 2.5 Kerangka Pemikiran.....	39
Gambar 3.1 Alur penelitian.....	43
Gambar 3.2 <i>Workflow</i> implementasi model.....	47
Gambar 3.3 Penggunaan fungsi ‘DETECTLANGUAGE’ pada Google Spreadsheet [19].....	48
Gambar 3.4 Gambaran struktur model mBERT yang akan dibangun.....	52
Gambar 3.5 Rancangan prototipe sistem pengklasifikasian lirik lagu eksplisit.....	55
Gambar 4.1 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 80:20.....	76
Gambar 4.2 Visualisasi <i>confusion matrix</i> pada skenario <i>imbalanced</i> 80:20.....	77
Gambar 4.3 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>imbalanced</i> 80:20.....	78
Gambar 4.4 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 70:30.....	79
Gambar 4.5 Visualisasi <i>confusion matrix</i> pada skenario <i>imbalanced</i> 70:30.....	80
Gambar 4.6 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>imbalanced</i> 70:30.....	81
Gambar 4.7 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 80:20.....	83

Gambar 4.8 Visualisasi <i>confusion matrix</i> pada skenario <i>balanced</i> 80:20.....	84
Gambar 4.9 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>balanced</i> 80:20.....	85
Gambar 4.10 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 70:30.....	86
Gambar 4.11 Visualisasi <i>confusion matrix</i> pada skenario <i>balanced</i> 70:30.....	87
Gambar 4.12 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>balanced</i> 70:30.....	88
Gambar 4.13 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	90
Gambar 4.14 Visualisasi <i>confusion matrix</i> pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	91
Gambar 4.15 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>imbalanced</i> 80:20 dengan <i>stratify</i> .....	92
Gambar 4.16 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	94
Gambar 4.17 Visualisasi <i>confusion matrix</i> pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	95
Gambar 4.18 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>imbalanced</i> 70:30 dengan <i>stratify</i> .....	96
Gambar 4.19 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	98
Gambar 4.20 Visualisasi <i>confusion matrix</i> pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	99
Gambar 4.21 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	100
Gambar 4.22 Tren metrik berdasarkan kombinasi <i>hyperparameter</i> pada skenario <i>balanced</i> 70:30 dengan <i>stratify</i> .....	101
Gambar 4.23 Visualisasi <i>confusion matrix</i> pada skenario <i>balanced</i> 70:30 dengan <i>stratify</i> .....	102

Gambar 4.24 Visualisasi hasil per <i>fold</i> untuk model terbaik pada skenario <i>balanced</i> 80:20 dengan <i>stratify</i> .....	103
Gambar 4.25 Proporsi eksplisit tekstual vs eksplisit implisit.....	109
Gambar 4.26 Distribusi bahasa pada lirik eksplisit-implisit.....	110
Gambar 4.27 <i>WordCloud</i> lirik eksplisit-implisit (tanpa kata kasar langsung).....	111
Gambar 4.28 <i>WordCloud</i> dari semua data berlabel eksplisit.....	111
Gambar 4.29 <i>WordCloud</i> dari data berlabel noneksplisit.....	112
Gambar 4.30 Tampilan awal prototipe sistem <i>tab</i> 1.....	113
Gambar 4.31 Tampilan sistem saat mengklasifikasikan lirik lagu berlabel eksplisit pada <i>tab</i> 1.....	114
Gambar 4.32 Tampilan sistem saat mengklasifikasikan lirik lagu berlabel noneksplisit pada <i>tab</i> 1.....	114
Gambar 4.33 Tampilan awal prototipe sistem <i>tab</i> 2.....	115
Gambar 4.34 Tampilan sistem saat mengklasifikasikan lirik lagu berlabel eksplisit pada <i>tab</i> 2.....	116
Gambar 4.35 Tampilan sistem saat mengklasifikasikan lirik lagu berlabel noneksplisit pada <i>tab</i> 2.....	116

## DAFTAR LAMPIRAN

Lampiran 1. Hasil Evaluasi Per <i>Fold</i> Semua Skenario ( <i>10-Fold Cross Validation</i> ).....	127
Lampiran 2. Lembar Konsultasi.....	130

## DAFTAR RIWAYAT HIDUP



Nama : Faridah  
Tempat, Tanggal Lahir: Banjarmasin, 01 Januari 2001  
Alamat : Jl. Pramuka Komp. Semanda 3  
RT. 20 RW. 02, Banjarmasin,  
Kalimantan Selatan  
Agama : Islam  
Kewarganegaraan : Indonesia  
Anak Ke- : 1 dari 3 bersaudara  
Riwayat Pendidikan : SDN Sungai Lulut 8  
SMPN 7 Banjarmasin  
MAN 2 Banjarmasin