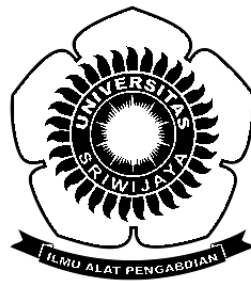


DETEKSI KATA KASAR DALAM LIRIK LAGU BERBAHASA INGGRIS MENGGUNAKAN MODEL BERT

*Diajukan Untuk Menyusun Skripsi
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh:

Bima Aryadinata

NIM : 09021282126035

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025

HALAMAN PENGESAHAN
SKRIPSI
DETEKSI KATA KASAR DALAM LIRIK LAGU BERBAHASA INGGRIS
MENGGUNAKAN MODEL BERT

Sebagai salah satu syarat untuk penyelesaian studi di
Program Studi S1 Teknik Informatika

Oleh:
BIMA ARYADINATA
09021282126035

Pembimbing 1 : Novi Yusliani, S.Kom., M.T
NIP. 198211082012122001
Pembimbing 2 : Muhammad Naufal Rachmatullah, S.Kom, M.Kom
NIP. 199212012022031008

Mengetahui
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D
198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jumat tanggal 09 Mei 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Bima Artadinata

NIM : 09021282126035

Judul : Deteksi Kata Kasar Dalam Lirik Lagu Berbahasa Inggris Menggunakan Model BERT

Dan dinyatakan **LULUS**.

1. Ketua Penguji

Samsuryadi, M.Kom., Ph.D.
NIP. 197102041997021003



2. Penguji

Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



3. Pembimbing I

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001



4. Pembimbing II

Muhammad Naufal Rachmatullah, S.Kom, M.Kom
NIP. 199212012022031008



Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Bima Aryadinata

NIM : 09021282126035

Program Studi : Teknik Informatika

Judul Skripsi : Deteksi Kata Kasar Dalam Lirik Lagu Berbahasa Inggris
Menggunakan Model BERT

Hasil Pengecekan *Software* Turnitin: 4%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 07 Mei 2025

Penulis,



Bima Aryadinata

NIM. 09021282126035

MOTTO DAN PERSEMBAHAN

Motto:

*“Bayangkan jika kita tidak menyerah, Tantangan apa pun dari Ayah atau dunia,
Kita hadapi kita lewati kita ikuti kita nikmati”*

- Bayangkan jika kita tidak menyerah (Hindia)

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Music plays a significant role through digital streaming platforms, while concerns about offensive language in lyrics, especially among young listeners, require attention. This study develops a BERT-based model to detect offensive language in English song lyrics. The dataset comprises 3,598 lyrics, divided into training 2,878 data, validation 360 data, and test 360 data. Experiments tested 12 parameter scenarios involving variations in learning rate 2e-5, 3e-6, 5e-5, batch size 16 and 32, and freeze layers 6 and 8. Evaluation used confusion matrix and metrics including accuracy, precision, recall, and F1-score. Results showed the model detected offensive language with 93.06% accuracy using learning rate 3e-6, batch size 16, and 8 freeze layers, demonstrating BERT's effectiveness for this task despite some classification errors due to sentence structure similarities.

Keywords: *Offensive Language Detection, Song Lyrics, BERT, NLP Model, Music Streaming*

ABSTRAK

Musik memiliki peran signifikan melalui platform streaming digital, namun kekhawatiran mengenai dampak kata kasar dalam lirik lagu, terutama pada kalangan muda, menjadi masalah penting. Penelitian ini mengembangkan model berbasis BERT untuk mendeteksi kata kasar dalam lirik lagu berbahasa Inggris. Dataset berjumlah 3.598 lirik, dibagi menjadi data latih sebanyak 2.878 data, data validasi 360 data, dan data uji 360 data. Eksperimen dilakukan dengan 12 skenario yang melibatkan variasi *learning rate* 2e-5, 3e-6, 5e-5, *batch size* 16 dan 32, dan jumlah *freeze layer* 6 dan 8. Evaluasi menggunakan *confusion matrix* dan metrik akurasi, presisi, *recall*, serta *F1-score*. Hasil terbaik menunjukkan model mampu mendeteksi kata kasar dengan akurasi 93,06% pada konfigurasi *learning rate* 3e-6, *batch size* 16, dan 8 *freeze layer*, membuktikan BERT efektif untuk tugas ini.

Kata Kunci: Deteksi Kata Kasar, Lirik Lagu, BERT, Model NLP, Streaming Musik

KATA PENGANTAR

Puji dan syukur kepada Allah SWT yang telah memberikan nikmat sehat, iman dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini dibuat sebagai salah satu syarat untuk menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam proses pembuatan skripsi ini penulis menerima bimbingan, bantuan, semangat, maupun petunjuk dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya sehingga, penulis dapat menyelesaikan skripsi ini dengan baik
2. Kedua orang tua dan keluarga yang telah mendoakan, memberi semangat dan motivasi, serta mendukung segala hal yang diperlukan untuk menyelesaikan skripsi ini.
3. Bapak Hadipurnawan Satria, Ph.D. selaku ketua jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Mastura Diana Marieska, S.T., M.T. selaku Dosen dan sekaligus pembimbing akademik yang telah memberikan banyak bantuan dan arahan kepada penulis selama perkuliahan.
5. Ibu Novi Yusliani, S.Kom., M.T. selaku Dosen Pembimbing skripsi I dan M. Naufal Rachmatullah, S.Kom, M.Kom selaku Dosen Pembimbing skripsi II yang telah memberi bimbingan, arahan serta semangat kepada penulis dalam menyelesaikan skripsi ini.

6. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Seluruh staf Administrasi dan Pegawai Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi.
8. Saudari Avinda Ghaniya yang memberikan semangat, bantuan, serta motivasi kepada penulis untuk menyelesaikan skripsi ini. Terimakasih atas segala bentuk *support* yang diberikan selama ini.
9. Organisasi tempat penulis berkembang, BEM KM FASILKOM Kabinet Askara Akasia dan Cakra Baswara serta BEM UNIVERSITAS SRIWIJAYA Kabinet Pijar Harmoni, yang membantu penulis membentuk karakter dan pribadi yang lebih baik selama perkuliahan ini.
10. Sahabat-sahabat penulis seperjuangan: Adit, Affandi, Anharul, Dzaky, Efan, Hafiz, Iman, Wahyu, Zidan, yang selalu menemani dan membantu serta mengingatkan penulis dalam menyelesaikan skripsi ini.
11. Teman-teman Invention, Teknik Informatika 2021 yang turut serta membersamai perjalanan ini.
12. Pihak-pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan karena kurangnya pengalaman dan pengetahuan penulis. Oleh karena

itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima Kasih.

Palembang, 07 Mei 2025

Penulis,

Bima Aryadinata

DAFTAR ISI

HALAMAN PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
<i>ABSTRACT</i>	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penulisan	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Penelitian	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Natural Language Processing (NLP)	II-1
2.2.2 Deteksi Kata Kasar.....	II-2
2.2.3 Pra-Pengolahan Teks	II-5
2.2.4 <i>Bidirectional Encoder Representations from Transformers</i> (BERT)...	II-8
2.2.5 <i>Confusion Matrix</i>	II-13
2.2.6 <i>Rational Unified Process</i> (RUP).....	II-15
2.3 Penelitian Lain yang Relevan	II-17
2.4 Kesimpulan	II-19

BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.2.1 Jenis dan Sumber Data	III-1
3.2.2 Metode Pengumpulan Data	III-1
3.3 Tahapan Penelitian	III-2
3.3.1 Mengumpulkan Data	III-4
3.3.2 Menentukan Kerangka Kerja Penelitian	III-5
3.3.3 Menentukan Kriteria Pengujian	III-8
3.3.4 Menentukan Format Data Pengujian.....	III-8
3.3.5 Menentukan Alat bantu Penelitian	III-10
3.3.6 Melakukan Pengujian Penelitian.....	III-11
3.3.7 Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-11
3.4 Metode Pengembangan Perangkat Lunak	III-12
3.5 Kesimpulan	III-13
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan Sistem	IV-1
4.2.3 Analisis dan Desain	IV-2
4.3 Fase Elaborasi	IV-24
4.3.1 Pemodelan Bisnis	IV-24
4.3.2 Kebutuhan Sistem	IV-28
4.3.3 Analisis dan Perancangan.....	IV-28
4.4 Fase Konstruksi.....	IV-35
4.4.1 Kebutuhan Sistem	IV-35
4.4.2 Implementasi	IV-37
4.5 Fase Transisi.....	IV-40
4.5.1 Pemodelan Bisnis	IV-40
4.5.2 Rencana Pengujian	IV-41
4.5.3 Implementasi	IV-42

4.6 Kesimpulan	IV-46
BAB V HASIL DAN ANALISIS	V-1
5.1 Pendahuluan	V-1
5.2 Hasil Penelitian	V-1
5.2.1 Konfigurasi Pengujian.....	V-1
5.2.2 Hasil Pengujian Skenario 1	V-3
5.2.3 Hasil Pengujian Skenario 2	V-5
5.2.4 Hasil Pengujian Skenario 3	V-7
5.2.5 Hasil Pengujian Skenario 4	V-9
5.2.6 Hasil Pengujian Skenario 5	V-11
5.2.7 Hasil Pengujian Skenario 6	V-13
5.2.8 Hasil Pengujian Skenario 7	V-15
5.2.9 Hasil Pengujian Skenario 8	V-17
5.2.10 Hasil Pengujian Skenario 9	V-19
5.2.11 Hasil Pengujian Skenario 10	V-21
5.2.12 Hasil Pengujian Skenario 11	V-23
5.2.13 Hasil Pengujian Skenario 12	V-25
5.3 Analisis Hasil Penelitian	V-27
5.4 Kesimpulan	V-42
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-1
DAFTAR PUSTAKA	xviii
LAMPIRAN.....	xxi

DAFTAR TABEL

Tabel II-1. Hasil Proses <i>Remove Punctuation</i>	II-6
Tabel II-2. Hasil Proses <i>Case Folding</i>	II-7
Tabel II-3. Hasil Proses <i>Stopword Removal</i>	II-7
Tabel II-4. Hasil Proses <i>Tokenization</i>	II-8
Tabel III-1. Contoh Data Lirik Lagu yang Digunakan	III-2
Tabel III-2. Skenario Konfigurasi Pengujian	III-9
Tabel III-3. Confusion Matrix	III-10
Tabel III-4. Alat Bantu Penelitian	III-10
Tabel III-5. Tabel Hasil Pengujian	III-11
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
Tabel IV-3. Contoh Lirik Lagu Berbahasa Inggris	IV-3
Tabel IV-4. Hasil <i>Case Folding</i>	IV-6
Tabel IV-5. Hasil <i>Remove Punctuation</i>	IV-8
Tabel IV-6. Hasil <i>Encoding</i>	IV-10
Tabel IV-7. Definisi Aktor sistem Pengujian	IV-15
Tabel IV-8. Definisi Aktor Sistem Pelatihan	IV-15
Tabel IV-9. Definisi <i>Use Case</i> Sistem Pengujian	IV-16
Tabel IV-9. Definisi <i>Use Case</i> Sistem Pengujian	IV-16
Tabel IV-10. Skenario Melakukan <i>Load Data</i>	IV-17
Tabel IV-11. Skenario Menampilkan Hasil Klasifikasi	IV-18
Tabel IV-12. Skenario Memprediksi Label Teks Lirik Lagu	IV-20
Tabel IV-13. Skenario Melakukan <i>Fine-tuning Pre-Trained Model BERT</i>	IV-22
Tabel IV-14. Implementasi Kelas Sistem Pengujian	IV-38
Tabel IV-15. Implementasi Kelas Sistem Pelatihan	IV-39
Tabel IV-16. Rencana Pengujian <i>Use Case</i> Menampilkan Hasil Klasifikasi..	IV-41
Tabel IV-17. Rencana Pengujian <i>Use Case</i> Melakukan <i>Load Data</i>	IV-41

Tabel IV-18. Rencana Pengujian <i>Use Case</i> Menampilkan Hasil Prediksi	IV-42
Tabel IV-19. Rencana Pengujian <i>Use Case</i> Pelatihan	IV-42
Tabel IV-20. Pengujian <i>Use Case</i> Menampilkan Hasil Klasifikasi.....	IV-43
Tabel IV-21. Pengujian <i>Use Case</i> Melakukan <i>Load Data</i>	IV-44
Tabel IV-22. Pengujian <i>Use Case</i> Menampilkan Hasil Prediksi.....	IV-45
Tabel V-1. Skenario <i>Fine-tuning</i> Model BERT.	V-1
Tabel V-2. Metrik Evaluasi Skenario 1.....	V-4
Tabel V-3. Metrik Evaluasi Skenario 2.....	V-6
Tabel V-4. Metrik Evaluasi Skenario 3.....	V-8
Tabel V-5. Metrik Evaluasi Skenario 4.....	V-10
Tabel V-6. Metrik Evaluasi Skenario 5.....	V-12
Tabel V-7. Metrik Evaluasi Skenario 6.....	V-14
Tabel V-8. Metrik Evaluasi Skenario 7.....	V-16
Tabel V-9. Metrik Evaluasi Skenario 8.....	V-18
Tabel V-10. Metrik Evaluasi Skenario 9.....	V-20
Tabel V-11. Metrik Evaluasi Skenario 10.....	V-22
Tabel V-12. Metrik Evaluasi Skenario 11	V-24
Tabel V-13. Metrik Evaluasi Skenario 12.....	V-26
Tabel V-14. Hasil Pengujian	V-27
Tabel V-15. <i>Id</i> Teks Lirik Lagu Berbahasa Inggris	V-29
Tabel V-16. Contoh Hasil Pengujian	V-36
Tabel V-17. Contoh Deteksi Kata Kasar Pada Lirik Lagu.....	V-37

DAFTAR GAMBAR

Gambar II-1. Arsitektur Deteksi Kata Kasar.....	II-2
Gambar II-2. Arsitektur BERT	II-9
Gambar II-3. Diagram siklus RUP.....	II-16
Gambar III-1. Rincian Tahapan Penelitian.	III-3
Gambar III-2. Distribusi Pembagian <i>Dataset</i>	III-4
Gambar III-3. Kerangka Kerja Penelitian.	III-5
Gambar IV-1. <i>Use Case</i> Diagram Sistem Pengujian.....	IV-14
Gambar IV-2. <i>Use Case</i> Diagram Sistem Pelatihan.	IV-15
Gambar IV-3. Desain Antarmuka Perangkat Lunak Klasifikasi.	IV-26
Gambar IV-4. Desain Antarmuka Perangkat Lunak Prediksi.....	IV-27
Gambar IV-5. Diagram Aktivitas <i>Load Data</i>	IV-29
Gambar IV-6. Diagram Aktivitas Menampilkan Hasil Klasifikasi.	IV-29
Gambar IV-7. Diagram Aktivitas Prediksi Kata Kasar Dalam Teks Lirik Lagu Berbahasa Inggris.....	IV-30
Gambar IV-8. Diagram Aktivitas Pelatihan.....	IV-31
Gambar IV-9. <i>Sequence</i> Diagram <i>Load Data</i>	IV-32
Gambar IV-10. <i>Sequence</i> Diagram Menampilkan Hasil Klasifikasi.....	IV-33
Gambar IV-11. <i>Sequence</i> Diagram Prediksi Kata Kasar Dalam Teks Lirik Lagu Berbahasa Inggris.....	IV-33
Gambar IV-12. <i>Sequence</i> Diagram Pelatihan.	IV-34
Gambar IV-13. Diagram Kelas Pengujian.....	IV-36
Gambar IV-14. Diagram Kelas Pelatihan.	IV-37
Gambar IV-15. Implementasi Antarmuka Halaman <i>Load Data</i>	IV-39
Gambar IV-16. Implementasi Antarmuka Halaman Hasil Klasifikasi.	IV-40
Gambar IV-17. Implementasi Antarmuka Halaman Hasil Prediksi.	IV-40
Gambar V-1. Grafik Akurasi dan <i>Loss</i> Skenario 1.	V-3
Gambar V-2. <i>Confusion Matrix</i> Skenario 1.....	V-3

Gambar V-3. Grafik Akurasi dan <i>Loss</i> Skenario 2.	V-5
Gambar V-4. <i>Confusion Matrix</i> Skenario 2.	V-5
Gambar V-5. Grafik Akurasi dan <i>Loss</i> Skenario 3.	V-7
Gambar V-6. <i>Confusion Matrix</i> Skenario 3.	V-7
Gambar V-7. Grafik Akurasi dan <i>Loss</i> Skenario 4.	V-9
Gambar V-8. <i>Confusion Matrix</i> Skenario 4.	V-9
Gambar V-9. Grafik Akurasi dan <i>Loss</i> Skenario 5.	V-11
Gambar V-10. <i>Confusion Matrix</i> Skenario 5.	V-11
Gambar V-11. Grafik Akurasi dan <i>Loss</i> Skenario 6.	V-13
Gambar V-12. <i>Confusion Matrix</i> Skenario 6.	V-13
Gambar V-13. Grafik Akurasi dan <i>Loss</i> Skenario 7.	V-15
Gambar V-14. <i>Confusion Matrix</i> Skenario 7.	V-15
Gambar V-15. Grafik Akurasi dan <i>Loss</i> Skenario 8.	V-17
Gambar V-16. <i>Confusion Matrix</i> Skenario 8.	V-17
Gambar V-17. Grafik Akurasi dan <i>Loss</i> Skenario 9.	V-19
Gambar V-18. <i>Confusion Matrix</i> Skenario 9.	V-19
Gambar V-19. Grafik Akurasi dan <i>Loss</i> Skenario 10.	V-21
Gambar V-20. <i>Confusion Matrix</i> Skenario 10.	V-21
Gambar V-21. Grafik Akurasi dan <i>Loss</i> Skenario 11.	V-23
Gambar V-22. <i>Confusion Matrix</i> Skenario 11.	V-23
Gambar V-23. Grafik Akurasi dan <i>Loss</i> Skenario 12.	V-25
Gambar V-24. <i>Confusion Matrix</i> Skenario 12.	V-25

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan menguraikan tentang landasan penelitian, yang mencakup latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika penulisan. Pembahasan mengenai hal-hal tersebut bertujuan untuk memberikan kerangka dasar yang mendasari proses penelitian.

1.2 Latar Belakang Masalah

Dalam era digital saat ini, musik menjadi salah satu pilihan medium komunikasi yang paling berpengaruh di dunia, terutama pada platform *streaming* seperti *Spotify*, *Apple Music*, dan *Youtube Music*. Berdasarkan data terbaru dari laporan *International Federation of the Phonographic Industry (IFPI) Engaging with Music 2023*¹, konsumsi musik digital melalui layanan streaming terus menunjukkan angka yang signifikan. Sekitar 74% orang dewasa mendengarkan musik melalui layanan audio streaming berlangganan atau yang didukung iklan, dengan rata-rata waktu yang dihabiskan untuk mendengarkan musik setiap minggu mencapai 20,7 jam. Platform streaming seperti *Spotify*, *Apple Music*, dan *YouTube Music* terus menjadi cara utama bagi pendengar di seluruh dunia untuk mengakses berbagai genre musik. Namun, di balik kepopuleran, terdapat kekhawatiran yang

¹ IFPI, *IFPI's global study finds we're listening to more music in more ways than ever*, <https://www.ifpi.org/ifpis-global-study-finds-were-listening-to-more-music-in-more-ways-than-ever/> diakses pada tanggal 10 Oktober 2024.

terus berkembang mengenai kata kasar dalam lirik lagu yang berdampak buruk, terutama pada kalangan muda.

Salah satu model NLP yang dapat diterapkan dalam deteksi kata kasar adalah *Bidirectional Encoder Representations from Transformers* (BERT). Model ini pertama kali diperkenalkan oleh (Devlin et al., 2018) dan telah digunakan secara luas dalam berbagai aplikasi, termasuk dalam deteksi kata kasar. Alasan penggunaan BERT dalam penelitian ini adalah karena pendekatan konvensional untuk deteksi kata kasar, seperti metode berbasis kamus (*lexicon-based*) atau pendekatan berbasis aturan, memiliki keterbatasan. Metode tersebut hanya mengandalkan pencocokan kata-kata kasar yang telah ada dalam *database*, sehingga kurang efektif dalam mengenali variasi ejaan atau makna kata yang dapat berubah sesuai dengan konteks penggunaannya. Sebaliknya, model berbasis *deep learning* seperti BERT dapat memahami hubungan antar kata dalam sebuah kalimat secara lebih akurat, sehingga mampu memberikan hasil prediksi yang lebih baik dalam mendeteksi kata kasar. Dalam konteks musik, teknologi BERT juga dapat disesuaikan agar mampu mendeteksi kata-kata kasar pada teks lirik lagu secara otomatis.

BERT (*Bidirectional Encoder Representations from Transformers*) memiliki keunggulan signifikan dalam pemrosesan bahasa alami, menjadikannya model yang sangat efektif untuk deteksi kata kasar dalam lirik lagu berbahasa Inggris. Salah satu keunggulan utama BERT adalah kemampuannya dalam memahami konteks dua arah, memungkinkan model ini untuk menangkap makna kata berdasarkan konteks sebelum dan sesudahnya. Pendekatan ini sangat penting

dalam menganalisis lirik lagu, yang sering kali memiliki struktur bahasa tidak baku, penggunaan metafora, serta permainan kata yang kompleks (Devlin et al., 2018).

Selain itu, BERT unggul dalam analisis sentimen dan deteksi ujaran kebencian yang merupakan aspek krusial dalam identifikasi kata kasar dalam lirik lagu. Pada studi sebelumnya (Amalia & Suyanto, 2024), menunjukkan bahwa model BERT mampu meningkatkan akurasi dalam klasifikasi ujaran kebencian hingga 89,11%, dengan *F1-score* mencapai 0,89, membuktikan efektivitasnya dalam memahami konteks suatu teks. Dalam konteks lirik lagu, dimana makna kata dapat bervariasi tergantung pada emosional yang diusung penyanyi, kemampuan BERT dalam memahami hal tersebut menjadi keunggulan model BERT itu sendiri.

Dalam penelitian lain, menunjukkan bahwa model ini mampu mengatasi tantangan dalam klasifikasi teks yang kompleks, termasuk deteksi kata kasar dalam berbagai konteks. González-Carvajal & Garrido-Merchán (2020) menemukan bahwa penggunaan BERT dalam analisis teks menunjukkan peningkatan akurasi dibandingkan metode lain seperti *Support Vector Machine* (SVM) dan *Naïve Bayes*, terutama dalam tugas klasifikasi teks yang membutuhkan pemahaman konteks yang lebih dalam. Dengan kemampuannya untuk menangkap konteks secara mendalam dan fleksibilitasnya dalam berbagai tugas NLP, BERT menjadi pilihan ideal dalam deteksi kata kasar pada lirik lagu berbahasa Inggris.

1.3 Rumusan Masalah

Penelitian ini bertujuan untuk menjawab pertanyaan-pertanyaan berikut:

1. Bagaimana membangun model *Bidirectional Encoder Representations from Transformers* (BERT) untuk mendeteksi kata kasar dalam teks lirik lagu berbahasa Inggris ?
2. Bagaimana kinerja model *Bidirectional Encoder Representations from Transformers* (BERT) untuk mendeteksi kata kasar dalam teks lirik lagu berbahasa Inggris berdasarkan akurasi ?

1.4 Tujuan Penulisan

Penelitian ini memiliki beberapa tujuan utama, yaitu :

1. Menghasilkan model BERT untuk mendeteksi kata kasar dalam teks lirik lagu berbahasa Inggris.
2. Mengetahui kinerja model BERT dalam mendeteksi kata kasar dalam teks lirik lagu berbahasa Inggris berdasarkan hasil akurasi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Akademis: Dapat menjadi kontribusi ilmiah berupa pemahaman baru tentang penerapan BERT untuk analisis teks lirik lagu dan bisa menjadi rujukan dalam penelitian terbaik.
2. Industri: Mendukung platform musik dalam menyaring konten yang lebih ramah pengguna sesuai regulasi.

3. Sosial: Melindungi audiens muda dari paparan konten kata kasar melalui teknologi deteksi otomatis.

1.6 Batasan Penelitian

Agar penelitian ini terfokus, beberapa batasan berikut diterapkan:

1. Penelitian hanya menggunakan dataset lirik lagu berbahasa Inggris².
2. Model yang digunakan adalah *BERT_{base}* dengan parameter yang disesuaikan untuk tugas klasifikasi teks.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar yang telah ditetapkan oleh Fakultas Ilmu Komputer Universitas Sriwijaya, yaitu:

BAB I. PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam penyusunan laporan akhir ini.

BAB II. KAJIAN LITERATUR

Bab ini menjelaskan mengenai landasan teori yang digunakan dalam penelitian. Dalam bab ini membahas mengenai kajian literatur, seperti penjelasan mengenai *Natural Language Processing* (NLP), pra-pengolahan teks, model BERTbase, serta penelitian sebelumnya yang berkaitan dengan penelitian ini.

² <https://github.com/heckzz/Explicit-Lyrics-Classfier/blob/master/data.json> diakses pada tanggal 13 Maret 2025

BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan proses yang dilakukan selama penelitian seperti, metode pengumpulan data, pelatihan model, hingga metode dalam perancangan perangkat lunak. Setiap tahapan penelitian dijelaskan secara rinci sesuai dengan kerangka kerja yang ditetapkan.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini menjelaskan tentang proses perancangan perangkat lunak yang akan dibangun, mulai dari analisa kebutuhan hingga tahap pengujian untuk menilai hasil yang diinginkan dalam pengembangan perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil dan analisis dari penelitian yang telah dilaksanakan sesuai dengan langkah serta metode yang direncanakan sebelumnya. Analisa ini digunakan sebagai landasan dalam menarik kesimpulan dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan yang diambil berdasarkan uraian dalam bab sebelumnya, serta saran yang diberikan berdasarkan penelitian yang telah dilakukan.

1.8 Kesimpulan

Dengan cepatnya perkembangan musik digital, kebutuhan akan penyaringan konten eksplisit menjadi semakin penting. Penelitian ini memberikan solusi berbasis BERT yang dapat diandalkan untuk mendeteksi kata kasar dalam lirik lagu Berbahasa Inggris.

DAFTAR PUSTAKA

- Aljawazeri, J. A., & Jasim, M. N. (2024). Addressing Challenges in Hate Speech Detection Using BERT-Based Models: A Review. *Iraqi Journal for Computer Science and Mathematics*, 5(2), 1–20. <https://doi.org/10.52866/ijcsm.2024.05.02.001>
- Amalia, F. S., & Suyanto, Y. (2024). Offensive Language and Hate Speech Detection using BERT Model. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 18(4). <https://doi.org/10.22146/ijccs.99841>
- Anwar, A. (2014). A Review of RUP (Rational Unified Process). In *Ashraf Anwar International Journal of Software Engineering (IJSE)* (Issue 5).
- Chen, X., Aljrees, T., Umer, M., Karamti, H., Tahir, S., Abuzinadah, N., Alnowaiser, K., Eshmawi, A. A., Mohamed, A., & Ashraf, I. (2023). A novel approach for explicit song lyrics detection using machine and deep ensemble learning models. *PeerJ Computer Science*, 9. <https://doi.org/10.7717/peerj-cs.1469>
- Chin, H., Kim, J., Kim, Y., Shin, J., & Yi, M. Y. (2018). Explicit Content Detection in Music Lyrics Using Machine Learning. *Proceedings - 2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018*, 517–521. <https://doi.org/10.1109/BigComp.2018.00085>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <http://arxiv.org/abs/1810.04805>
- Dwiyani, L. K. D., Suarjaya, I. M. A. D., & Rusjyanthi, N. K. D. (2023). Classification of Explicit Songs Based on Lyrics Using Random Forest Algorithm. *Journal of Information Systems and Informatics*, 5(2), 550–567. <https://doi.org/10.51519/journalisi.v5i2.491>
- Fariz, M., & Noval Riswandha, M. (2024). IMPLEMENTASI METODE RATIONAL UNIFIED PROCESS(RUP) UNTUK RANCANG BANGUN SISTEM INFORMASI PENGAJUAN CUTI DI PT. BETTS INDONESIA BERBASIS WEBSITE MENGGUNAKAN FRAMEWORK LARAVEL. *Jurnal Sistem Informasi Aplikasi Teknologi Informasi*, 1(2). <https://doi.org/10.53567/josiati.v1i2.7>
- González-Carvajal, S., & Garrido-Merchán, E. C. (2020). *Comparing BERT against traditional machine learning text classification*. <https://doi.org/10.47852/bonviewJCCE3202838>
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5). <https://doi.org/10.1371/journal.pone.0232525>

- Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification*. <http://arxiv.org/abs/1801.06146>
- Irawan Chandra, Y., Ruri Irawati, D., Riastuti, M., Jakarta STI, S., Jalan BRI No, K., Dalam, R., Baru, K., & Selatan, J. (n.d.). *Penerapan Model Rational Unified Process (RUP) Dalam Membangun Aplikasi Promosi dan Penjualan Berbasis Web (Studi Kasus: Toko Mie Ayam “Neng Kene Kiyeh”)*. <https://doi.org/10.37817/ikraith-informatika.v8i3>
- Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 887–894. <https://doi.org/10.25126/jtiik.1148315>
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Loshchilov, I., & Hutter, F. (2017). *Decoupled Weight Decay Regularization*. <http://arxiv.org/abs/1711.05101>
- Sathyanarayanan, S. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 4023–4031. <https://doi.org/10.53555/AJBR.v27i4S.4345>
- Schofield, A., Magnusson, M., & Mimno, D. (2017). *Pulling Out the Stops: Rethinking Stopword Removal for Topic Models*.
- Suarna, N., & Prihartono, W. (2024). PENERAPAN NLP (NATURAL LANGUAGE PROCESSING) DALAM ANALISIS SENTIMEN PENGGUNA TELEGRAM DI PLAYSTORE. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>
- Wang, C., Nulty, P., & Lillis, D. (2020). A Comparative Study on Word Embeddings in Deep Learning for Text Classification. *ACM International Conference Proceeding Series*, 37–46. <https://doi.org/10.1145/3443279.3443304>
- Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., & Kumar, R. (2019). *Predicting the Type and Target of Offensive Posts in Social Media*. <http://arxiv.org/abs/1902.09666>

Zhu, L., & Luo, D. (2023). A Novel Efficient and Effective Preprocessing Algorithm for Text Classification. *Journal of Computer and Communications*, 11(03), 1–14. <https://doi.org/10.4236/jcc.2023.113001>