

**ANALISIS SENTIMEN LIRIK LAGU UNTUK PREDIKSI POPULARITAS
LAGU DENGAN ALGORITMA BERT**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Hana Agatha

00000035764

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**ANALISIS SENTIMEN LIRIK LAGU UNTUK PREDIKSI POPULARITAS
LAGU DENGAN ALGORITMA BERT**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Hana Agatha

00000035764

UMMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Hana Agatha
Nomor Induk Mahasiswa : 00000035764
Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Lirik Lagu untuk Prediksi Popularitas Lagu dengan Algoritma BERT

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 12 Juni 2023



(Hana Agatha)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN LIRIK LAGU UNTUK PREDIKSI POPULARITAS LAGU DENGAN ALGORITMA BERT

oleh

Nama : Hana Agatha
NIM : 00000035764
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 27 Juni 2023

Pukul 13.00 s/d 15.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Adhi Kusnadi, S.T, M.Si.)

NIDN: 0303037304

Penguji




(Dr. Maria Irmina Prasetiyowati, S.Kom.,

M.T.)

NIDN: 0725057201

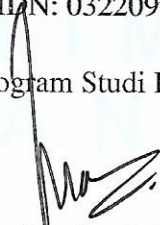
Pembimbing



(Alethea Suryadibrata, S.Kom., M.Eng.)

NIDN: 0322099201

Ketua Program Studi Informatika,



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Hana Agatha
NIM : 00000035764
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**ANALISIS SENTIMEN LIRIK LAGU UNTUK PREDIKSI POPULARITAS
LAGU DENGAN ALGORITMA BERT**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 3 Juli 2023
Yang menyatakan



Hana Agatha

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Halaman Persembahan / Motto

Do the best and let God do the rest.



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Lirik Lagu untuk Prediksi Popularitas Lagu dengan Algoritma BERT dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Alethea Suryadibrata, S.Kom., M.Eng., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang tua, keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Teman-teman saya yang terus-menerus memberikan dukungan serta semangat, dalam pengerjaan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 12 Juni 2023



Hana Agatha

ANALISIS SENTIMEN LIRIK LAGU UNTUK PREDIKSI POPULARITAS LAGU DENGAN ALGORITMA BERT

Hana Agatha

ABSTRAK

Persaingan di industri musik yang semakin ketat merugikan sebagian musisi. Para musisi perlu lebih memperhatikan faktor-faktor yang mempengaruhi popularitas suatu lagu, agar lagu mereka dapat menjadi terkenal dan mereka pun mendapat keuntungan. Salah satu faktor yang dapat mempengaruhi popularitas lagu adalah lirik lagu. Pengaruh lirik pada popularitas lagu dapat diteliti dengan analisis sentimen. Berdasarkan analisis sentimen pada lirik lagu, popularitas lagu dapat diprediksi. Dengan prediksi popularitas lagu, penulis lagu dapat mengevaluasi lagu buatannya, sehingga dapat menjadi terkenal. Adapun analisis sentimen merupakan studi komputasi yang mempelajari sentimen atau emosi dari suatu teks. Salah satu algoritma *deep learning* yang dapat digunakan untuk analisis sentimen adalah *Bidirectional Encoder Representations from Transformers* (BERT). Algoritma BERT merupakan algoritma yang unggul dalam hal analisis sentimen. Pada penelitian ini, suatu model BERT dikembangkan untuk keperluan prediksi popularitas lagu, berdasarkan analisis sentimen lirik lagu. Suatu lagu akan diprediksi kelas popularitasnya, berdasarkan hasil analisis sentimen pada lirik lagu tersebut. Model yang dikembangkan merupakan model yang sudah dilatih dengan data lagu berbahasa Inggris. Berdasarkan pengujian model, model yang menggunakan metode *oversampling* mencapai *accuracy* sebesar 87%, *precision* sebesar 88%, *recall* sebesar 87%, dan *f1-score* sebesar 87%. Model tersebut kemudian diimplementasikan dalam bentuk sebuah aplikasi web sederhana.

Kata kunci: analisis sentimen, *bidirectional encoder representations from transformers*, pembelajaran mesin, popularitas lagu

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Sentiment Analysis on Song Lyrics for Song Popularity Prediction Using BERT Algorithm

Hana Agatha

ABSTRACT

The increasingly competitiveness in music industry is giving some musicians disadvantages. Musicians need to pay more attention to the factors that influence the popularity of a song, so their song can be popular and they can gain a lot of profit. One of the various factors that can affect a song popularity is the lyrics. The influence of the lyrics can be explored through sentiment analysis. Sentiment analysis is a computing study that identify sentiments or emotions in a text. By conducting sentiment analysis on the lyrics, song popularity can be predicted. Based on the prediction result, songwriters can evaluate their lyrics, so their song can be popular. One of deep learning algorithms that can be used for sentiment analysis is Bidirectional Encoder Representations from Transformers (BERT). BERT is an excellent algorithm in terms of sentiment analysis. In this study, a BERT model was developed to predict the song popularity, based on the sentiment analysis of the song lyrics. The popularity class of a song will be predicted, based on the results of lyrics sentiment analysis. The developed model is a model that has been trained with English songs. Based on the model testing, the model that uses oversampling method achieved accuracy by 87%, precision by 88%, recall by 87%, and f1-score by 87%. The model then got implemented in the form of a simple web application.

Keywords: bidirectional encoder representations from transformers, machine learning, sentiment analysis, song popularity

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Lirik Lagu	6
2.2 Analisis Sentimen	6
2.3 Preprocessing	7
2.4 Bidirectional Encoder Representations From Transformers (BERT)	7
2.4.1 Representasi Input/Output Model BERT	8
2.4.2 Pemodelan BERT	8
2.5 Metrik Evaluasi	10
2.5.1 Accuracy	11
2.5.2 Precision	11
2.5.3 Recall	12
2.5.4 F1-Score	12
BAB 3 METODOLOGI PENELITIAN	13
3.1 Metodologi Penelitian	13
3.1.1 Telaah Literatur	13
3.1.2 Analisa Kebutuhan	13
3.1.3 Pengumpulan Data	13
3.1.4 Perancangan Model	14
3.1.5 Implementasi Model	14
3.1.6 Pengujian Model dan Evaluasi	14
3.1.7 Dokumentasi	14
3.2 Perancangan Model	14
3.2.1 Import Libraries	15
3.2.2 Data Labelling	15
3.2.3 Preprocessing	16
3.2.4 Split Dataset	17
3.2.5 Training Model	17
3.2.6 Evaluate Model	19
3.3 Perancangan Desain Antarmuka	20

BAB 4	HASIL DAN DISKUSI	22
4.1	Spesifikasi Sistem	22
4.2	Implementasi Aplikasi	23
4.3	Implementasi Model	24
4.3.1	Load Dataset	24
4.3.2	Data Labelling	25
4.3.3	Preprocessing	26
4.3.4	Split Dataset	31
4.3.5	Training Model	31
4.3.6	Evaluate Model	34
4.4	Uji Coba dan Evaluasi	35
4.4.1	Hasil Uji Coba	37
4.4.2	Evaluasi	49
BAB 5	SIMPULAN DAN SARAN	51
5.1	Simpulan	51
5.2	Saran	51
	DAFTAR PUSTAKA	52



DAFTAR GAMBAR

Gambar 2.1	Proses tokenisasi <i>input</i> model BERT	8
Gambar 2.2	Tahap-tahap pemodelan BERT	9
Gambar 2.3	<i>Confusion matrix</i>	11
Gambar 3.1	<i>Flowchart</i> utama	15
Gambar 3.2	<i>Flowchart preprocessing</i>	16
Gambar 3.3	Rancangan arsitektur model BERT	18
Gambar 3.4	<i>Flowchart training model</i>	19
Gambar 3.5	<i>Flowchart evaluate model</i>	20
Gambar 3.6	Desain tampilan <i>user interface</i>	21
Gambar 4.1	Tampilan bagian pertama dari <i>website</i>	23
Gambar 4.2	Tampilan bagian kedua dari <i>website</i>	24
Gambar 4.3	Cuplikan tabel <i>dataset</i>	25
Gambar 4.4	Arsitektur model	33
Gambar 4.5	Grafik distribusi lagu berdasarkan label popularitasnya . .	36
Gambar 4.6	Grafik <i>accuracy</i> dan <i>loss</i> sewaktu <i>training</i> model uji coba pertama	37
Gambar 4.7	<i>Confusion matrix</i> performa model tanpa <i>stopwords removal</i> dan <i>lemmatization</i>	38
Gambar 4.8	<i>Classification report</i> performa model tanpa <i>stopwords removal</i> dan <i>lemmatization</i>	39
Gambar 4.9	Grafik <i>accuracy</i> dan <i>loss</i> sewaktu <i>training</i> model uji coba kedua	40
Gambar 4.10	<i>Confusion matrix</i> performa model dengan <i>stopwords removal</i> dan <i>lemmatization</i>	41
Gambar 4.11	<i>Classification report</i> performa model dengan <i>stopwords removal</i> dan <i>lemmatization</i>	42
Gambar 4.12	Grafik <i>accuracy</i> dan <i>loss</i> sewaktu <i>training</i> model uji coba ketiga	43
Gambar 4.13	<i>Confusion matrix</i> performa model dengan <i>undersampling</i> .	44
Gambar 4.14	<i>Classification report</i> performa model dengan <i>undersampling</i>	46
Gambar 4.15	Grafik <i>accuracy</i> dan <i>loss</i> sewaktu <i>training</i> model uji coba keempat	47
Gambar 4.16	<i>Confusion matrix</i> performa model dengan <i>oversampling</i> . .	47
Gambar 4.17	<i>Classification report</i> performa model dengan <i>oversampling</i>	49

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 4.1	Perbandingan teks tanpa <i>stopwords removal</i> dan teks dengan <i>stopwords removal</i>	28
Tabel 4.2	Perbandingan teks tanpa <i>lemmatization</i> dan teks dengan <i>lemmatization</i>	29
Tabel 4.3	Hasil perhitungan nilai <i>confusion matrix</i> pemodelan tanpa <i>stopwords removal</i> dan <i>lemmatization</i>	38
Tabel 4.4	Hasil perhitungan nilai <i>confusion matrix</i> pemodelan dengan <i>stopwords removal</i> dan <i>lemmatization</i>	41
Tabel 4.5	Hasil perhitungan nilai <i>confusion matrix</i> pemodelan dengan <i>undersampling</i>	44
Tabel 4.6	Hasil perhitungan nilai <i>confusion matrix</i> pemodelan dengan <i>oversampling</i>	48
Tabel 4.7	Hasil uji coba	50



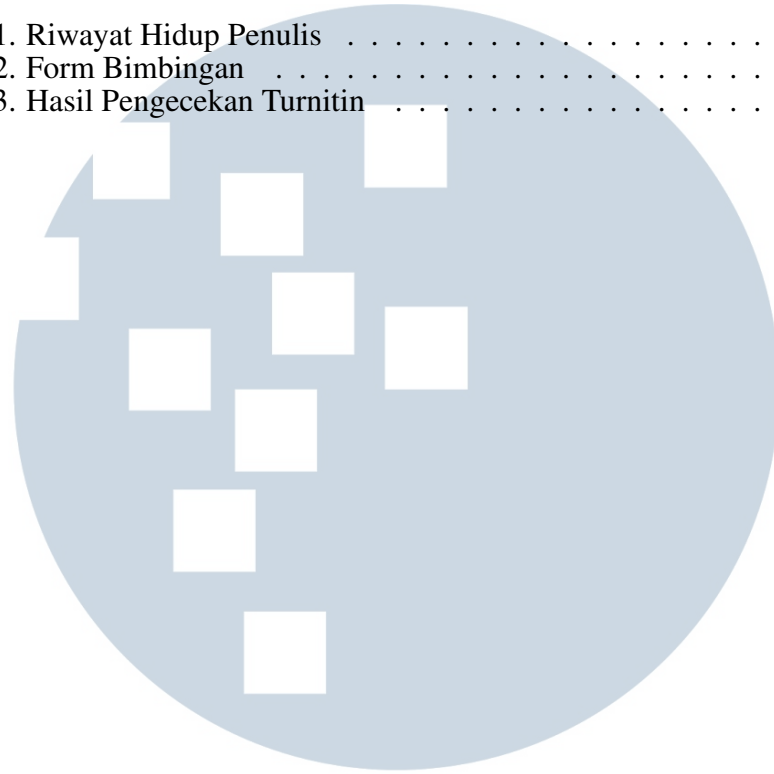
DAFTAR KODE

4.1	Potongan kode untuk membuang data yang tidak digunakan	25
4.2	Potongan kode untuk <i>function data labelling</i>	26
4.3	Potongan kode untuk penyaringan data lagu berdasarkan bahasa	26
4.4	Potongan kode untuk <i>function data cleaning</i>	27
4.5	Potongan kode untuk <i>function case folding</i>	27
4.6	Potongan kode untuk <i>function stopwords removal</i>	27
4.7	Potongan kode untuk <i>function lemmatization</i>	28
4.8	Potongan kode untuk proses tokenisasi	30
4.9	Potongan kode untuk <i>shuffle</i> dan <i>batching</i> dataset	30
4.10	Potongan kode untuk <i>split dataset</i>	31
4.11	Potongan kode untuk pembuatan model	32
4.12	Potongan kode untuk <i>training model</i>	33
4.13	Potongan kode untuk evaluasi model	34
4.14	Potongan kode untuk uji coba pemodelan tanpa <i>stopwords removal</i> dan <i>lemmatization</i>	37
4.15	Potongan kode untuk uji coba pemodelan dengan <i>stopwords</i> <i>removal</i> dan <i>lemmatization</i>	40
4.16	Potongan kode untuk uji coba pemodelan dengan <i>undersampling</i>	43
4.17	Potongan kode untuk uji coba pemodelan dengan <i>oversampling</i>	46



DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup Penulis	54
Lampiran 2. Form Bimbingan	56
Lampiran 3. Hasil Pengecekan Turnitin	57



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA