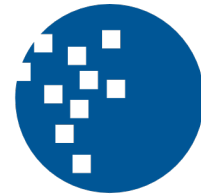


**ANALISIS SENTIMEN PENGGUNAAN AI-GENERATED ARTWORK DI
TWITTER MENGGUNAKAN INDOBERT**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Bryan Christopher
00000044058

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**ANALISIS SENTIMEN PENGGUNAAN AI-GENERATED ARTWORK DI
TWITTER MENGGUNAKAN INDOBERT**



Bryan Christopher

00000044058

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

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

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Bryan Christopher

Nomor Induk Mahasiswa : 00000044058

Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Penggunaan AI-Generated Artwork di Twitter Menggunakan IndoBERT

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis 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 maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 4 Juli 2025



(Bryan Christopher)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN PENGGUNAAN AI-GENERATED ARTWORK DI TWITTER MENGGUNAKAN INDOBERT

oleh

Nama : Bryan Christopher
NIM : 00000044058
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 28 Juli 2025

Pukul 10.00 s/s 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



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

NIDN: 0818038501

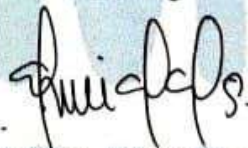
Penguji



(Fenina Adline Twince Tobing, S.Kom.,
M.Kom)

NIDN: 0406058802

Pembimbing



(Eunike Endariahna Surbakti, S.Kom., M.T.I.)

NIDN: 0322099401

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI))

NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Bryan Christopher
NIM : 00000044058
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis Sentimen Penggunaan
AI-Generated Artwork di Twitter
Menggunakan IndoBERT

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 4 Juli 2025

Yang menyatakan



Bryan Christopher

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

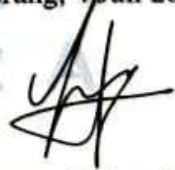
Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Terhadap Penggunaan AI-generated artwork di Twitter menggunakan IndoNLP. Laporan ini dibuat untuk memenuhi salah satu syarat untuk mendapatkan gelar Sarjana Komputer Jurusan Informatika di 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, saya akan mengalami kesulitan dalam mengerjakan laporan ini. Oleh sebab itu, saya mengucapkan terima kasih kepada:

Oleh sebab itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., 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. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI), selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Eunike Endariahna Surbakti, S.Kom., M.T.I., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

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

Tangerang, 4 Juli 2025



Bryan Christopher

ANALISIS SENTIMEN PENGGUNAAN AI-GENERATED ARTWORK DI TWITTER MENGGUNAKAN INDOBERT

Bryan Christopher

ABSTRAK

Perkembangan teknologi kecerdasan buatan (AI) telah memungkinkan terciptanya karya seni digital secara otomatis melalui algoritma, yang dikenal sebagai AI-generated artwork. Fenomena ini menimbulkan kontroversi di masyarakat, terutama di media sosial seperti Twitter, karena berbagai kekhawatiran seperti plagiarisme dan pelanggaran etika. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap penggunaan AI-generated artwork di Twitter menggunakan model IndoBERT yang termasuk dalam benchmark IndoNLP. Sebanyak 2.483 data tweet dikumpulkan melalui scraping Twitter menggunakan kata kunci “gambar AI” dan “ilustrasi AI”, kemudian dilakukan proses preprocessing mencakup penghapusan tanda baca, angka, emoji, stopword, dan lemmatization. Data yang telah diproses kemudian diklasifikasikan menggunakan model SmSA dari IndoBERT ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil labeling menunjukkan bahwa 51,8 persen dari opini pengguna tergolong negatif, 47,3 persen netral, dan hanya 0,9 persen bersentimen positif. Dari empat skenario pembagian dataset (40:30:30, 60:20:20, 70:15:15, dan 80:10:10), model terbaik diperoleh pada konfigurasi 40:30:30 pada epoch ke-7, dengan akurasi tertinggi mencapai 0,75, F1 Score 0,75, dan Precision 0,78, menggunakan parameter pelatihan optimal hasil scenario testing. Temuan ini mengindikasikan dominasi sentimen negatif dalam persepsi publik terhadap karya seni berbasis AI, yang mencerminkan kekhawatiran masyarakat terhadap legitimasi, kualitas, dan dampak etis dari penggunaan AI dalam seni.

Kata kunci: AI-generated artwork, IndoBERT, Natural Language Processing, analisis sentimen, Twitter

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 OF THE USE OF AI-GENERATED ARTWORK ON TWITTER USING INDOBERT

Bryan Christopher

ABSTRACT

The advancement of artificial intelligence (AI) technology has enabled the automatic creation of digital artworks through algorithmic processes, commonly referred to as AI-generated artwork. This phenomenon has sparked public controversy, especially on social media platforms such as Twitter, due to concerns over plagiarism, ethical violations, and artistic authenticity. This research aims to analyze the sentiments of the Indonesian public toward the use of AI-generated artwork on Twitter using the IndoBERT model, which is part of the IndoNLP benchmark. A total of 2,483 tweets were collected through Twitter scraping using the keywords “gambar AI” and “ilustrasi AI”. The collected data underwent a comprehensive preprocessing stage, including the removal of punctuation, digits, emojis, stopwords, and the application of lemmatization. The cleaned data was then classified using the SmSA model (Sentiment Analysis) from IndoBERT into three sentiment categories: positive, negative, and neutral. The labeling results indicate that 51.8 percent of public opinions were negative, 47.3 percent neutral, and only 0.9 percent positive. Among the four dataset split scenarios tested (40:30:30, 60:20:20, 70:15:15, and 80:10:10), the best model performance was achieved using the 40:30:30 configuration at epoch-7, with a highest accuracy of 0.75, an F1 Score of 0.75, and Precision of 0.78, using optimized training parameters from scenario testing. These findings highlight a dominant negative sentiment in public perception toward AI-generated artwork, reflecting concerns about legitimacy, quality, and ethical implications associated with the use of AI in the field of art.

Keywords: AI-generated artwork, IndoBERT, Natural Language Processing, sentiment analysis, Twitter

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 PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	7
2.1 Artificial Intelligence (AI)	7
2.2 Twitter	7
2.3 AI-generated artwork	8
2.4 Natural Language Processing (<i>NLP</i>)	10
2.5 Bidirectional Encoder Representations from Transformers (<i>BERT</i>)	12
2.5.1 Masked Language Model (<i>MLM</i>)	15
2.5.2 Next Sentence Prediction (<i>NSP</i>)	16
2.6 IndoNLU	17
2.6.1 IndoBERT-base	17
2.6.2 Model SmSA	17
2.7 Split Dataset	17
2.7.1 Data Training	18
2.7.2 Data Valid	18
2.7.3 Data Testing	19
2.8 Evaluasi Kinerja Model	19
2.8.1 Confusion Matrix	19
2.8.2 Accuracy	20
2.8.3 Precision	20
2.8.4 Recall	20
2.8.5 F1-Score	20
2.9 Penelitian Terkait	21
BAB 3 METODOLOGI PENELITIAN	22
3.1 Objek Penelitian	22
3.2 Alur Penelitian	22
3.3 Identifikasi Permasalahan	26
3.4 Studi Literatur	26
3.5 Pengumpulan Data	27

3.6	Data Preprocessing	29
3.6.1	Data Cleansing	30
3.6.2	Case Folding	30
3.6.3	Erase Punctuation	32
3.6.4	Erase Emoji/Emoticon	33
3.6.5	Erase Digits	34
3.6.6	Stopword	34
3.6.7	Lemmatization	35
3.6.8	Model IndoBERT	36
BAB 4	HASIL DAN DISKUSI	39
4.1	Hasil Implementasi	39
4.2	Scenario Testing	45
4.3	Evaluasi	50
4.3.1	Split Dataset	50
4.3.2	Evaluation	51
4.3.3	Analisis Sentimen	63
BAB 5	SIMPULAN DAN SARAN	65
5.1	Simpulan	65
5.2	Saran	66
DAFTAR PUSTAKA		67



DAFTAR TABEL

Tabel 2.1	Contoh Proses MLM	15
Tabel 2.2	Contoh Proses NSP	16
Tabel 3.1	Pengubahan Kapitalisasi Teks pada Data Twitter	31
Tabel 3.2	Penghapusan Tanda Baca pada Data Twitter	33
Tabel 3.3	Penghapusan Emoji pada Data Twitter	33
Tabel 3.4	Penghapusan Digit pada Data Twitter	34
Tabel 3.5	Implementasi Stopword pada Data Twitter	35
Tabel 3.6	Implementasi Lemmatization pada Data Twitter	36
Tabel 3.7	Model IndoBERT	37
Tabel 4.1	Model Training Parameters	46
Tabel 4.2	Model Training Parameters	50
Tabel 4.3	Tabel Split Dataset 40:30:30	52
Tabel 4.4	Tabel Split Dataset 60:20:20	54
Tabel 4.5	Tabel Split Dataset 70:15:15	56
Tabel 4.6	Tabel Split Dataset 80:10:10	58
Tabel 4.7	Split Dataset 60:20:20 Continuation	59
Tabel 4.8	Split Dataset 40:30:30 Continuation	60
Tabel 4.9	Split Dataset 40:30:30 Final	62



DAFTAR GAMBAR

Gambar 2.1	Contoh hasil generasi karya digital buatan AI.	8
Gambar 2.2	Website Midjourney[1]	9
Gambar 2.3	Terminologi NLP[2]	10
Gambar 2.4	Pre-Training dan Fine-Tuning BERT[3]	13
Gambar 2.5	Input-Output Embedding pada BERT[3]	14
Gambar 3.1	Metodologi Penelitian	23
Gambar 3.2	Penarikan <i>post/tweet</i> menggunakan kata kunci 'gambar ai' dalam bahasa Indonesia.	27
Gambar 3.3	Hasil data yang telah ditarik menggunakan <i>Twscraper</i> . . .	28
Gambar 3.4	Data yang sudah dirapikan dalam bentuk .csv.	29
Gambar 3.5	Data siap diproses.	30
Gambar 3.6	Semua huruf kapital diubah menjadi huruf kecil.	31
Gambar 4.1	<i>Perbandingan akurasi antar model</i>	47
Gambar 4.2	<i>Perbandingan F1 Score, Recall and Precision Chart antar model</i>	48
Gambar 4.3	<i>Loss Chart</i>	48
Gambar 4.4	<i>Perbandingan Overhead (Time) antar model</i>	49
Gambar 4.5	Hasil Split Dataset 40:30:30	51
Gambar 4.6	Hasil Split Dataset 60:20:20	53
Gambar 4.7	Hasil Split Dataset 70:15:15	55
Gambar 4.8	Hasil Split Dataset 80:10:10	57



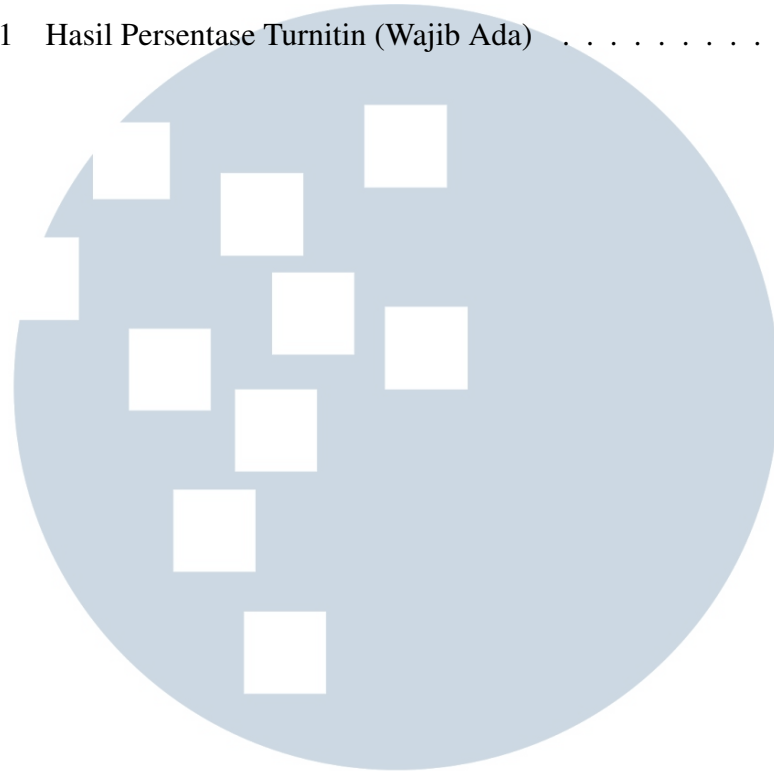
DAFTAR KODE

Kode 4.1	Kode untuk menarik data yang dibutuhkan dari hasil ekstraksi twitter.	39
Kode 4.2	Kode untuk menghilangkan stopwords.	40
Kode 4.3	Kode untuk lemmatization.	41
Kode 4.4	Kode untuk menghilangkan punctuation.	42
Kode 4.5	Kode untuk menghilangkan angka.	42
Kode 4.6	Kode untuk menghilangkan angka.	43
Kode 4.7	Kode untuk menghilangkan emoticon.	44



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin (Wajib Ada)	70
------------	---------------------------------------	----



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA