

**ANALISIS SENTIMEN PUBLIK TERHADAP STRATEGI
PEMASARAN KOLABORASI NCT DREAM DAN
TOSTOS MENGGUNAKAN INDOBERT**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

**MISHEL MILEN
00000055747**

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

**ANALISIS SENTIMEN PUBLIK TERHADAP STRATEGI
PEMASARAN KOLABORASI NCT DREAM DAN
TOSTOS MENGGUNAKAN INDOBERT**



SKRIPSI

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

MISHEL MILEN
00000055747

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Mishel Milen

Nomor Induk Mahasiswa : 00000055747

Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Publik terhadap Strategi Pemasaran Kolaborasi NCT DREAM dan TosTos 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

[meterai Rp 10.000,-]

(Mishel Milen)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN PUBLIK TERHADAP STRATEGI PEMASARAN KOLABORASI NCT DREAM DAN TOSTOS MENGGUNAKAN INDOBERT

oleh

Nama : Mishel Milen
NIM : 00000055747
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 24 Juli 2025

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

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

NIDN: 0818038501

Penguji

(Moeljono Widjaja, B.Sc., M.Sc.,
Ph.D)

NIDN: 0311106903

Pembimbing

(Dr. Maria Irmira Prasetyowati, S.Kom., M.T.)

NIDN: 725057201

Ketua Program Studi Informatika,

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

NIDN: 0315109103

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

Yang bertanda tangan di bawah ini:

Nama : Mishel Milen
NIM : 00000055747
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis Sentimen Publik terhadap
Strategi Pemasaran Kolaborasi NCT
DREAM dan TosTos 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



Mishel Milen

**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

"Have a good friends around you, have good peers, surround yourself with good people, CAUSE YOU'RE A GOOD PERSON TOO."

MARK LEE



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Publik terhadap Strategi Pemasaran Kolaborasi NCT DREAM dan TosTos Menggunakan IndoBERT dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan laporan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikannya. Oleh karena itu, penulis ingin 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, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Dr. Maria Irminda Prasetyowati, S.Kom., M.T., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan skripsi ini.
5. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D dan Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Penguji dan Ketua Sidang yang telah memberikan masukan, kritik, dan saran yang membangun selama proses sidang, sehingga kualitas laporan skripsi ini menjadi lebih baik.
6. Seluruh dosen Program Studi Informatika yang telah memberikan ilmu yang bermanfaat selama masa perkuliahan.
7. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan skripsi ini.
8. Nisa'ul Fadhilah yang telah meluangkan waktu untuk membantu proses validasi data sebagai *annatator* kedua.

9. Mark, Renjun, Jeno, Haechan, Jaemin, Chenle, dan Jisung yang selalu menjadi penyemangat dan memberikan motivasi untuk menyelesaikan laporan skripsi ini.
10. Nayasha, Abelia, Wilcoustine, Ayu, Salwa, Janice, Chelsy, Vania, Marcia, felicia, dan Angel selaku teman terbaik yang telah banyak memberikan dukungan serta bantuan berupa waktu dan tenaga dalam penyusunan laporan skripsi ini.

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

Tangerang, 4 Juli 2025



Mishel Milen



ANALISIS SENTIMEN PUBLIK TERHADAP STRATEGI PEMASARAN KOLABORASI NCT DREAM DAN TOSTOS MENGGUNAKAN INDOBERT

Mishel Milen

ABSTRAK

Analisis sentimen memegang peranan krusial dalam memahami opini publik terhadap strategi pemasaran, khususnya pada era kolaborasi *brand* lokal dengan artis internasional di media sosial. Penelitian ini bertujuan untuk mengimplementasikan model *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT) guna menganalisis sentimen publik terhadap kolaborasi NCT DREAM dan TosTos, serta mengevaluasi performa model tersebut dalam mengklasifikasikan sentimen. Data komentar publik dikumpulkan dari postingan Instagram *tostosid* periode Maret 2024 hingga Maret 2025, menghasilkan total 6.830 komentar. Proses *preprocessing data* meliputi *case folding*, penghapusan URL, penggantian emoji dan normalisasi *slang word* menggunakan kamus khusus, serta penghapusan spasi berlebih. Model IndoBERT di *fine-tune* menggunakan 1.800 data berlabel manual dengan pembagian 70:15:15 untuk *training*, *validation*, dan *test set*. Hasil evaluasi menunjukkan model IndoBERT memiliki kapabilitas yang baik dalam klasifikasi sentimen, mencapai akurasi 83% pada *validation set*, dengan *F1-score* 0.86 untuk sentimen positif dan 0.84 untuk negatif, meskipun kelas netral (*F1-score* 0.78) masih menjadi tantangan. Analisis sentimen pada seluruh data komentar mengungkapkan dominasi sentimen positif (44.7%), mencerminkan antusiasme publik yang tinggi. Namun, *word cloud* dan analisis tren menunjukkan frustrasi signifikan terkait mekanisme pengundian hadiah, yang memuncak pada kontroversi di Agustus 2024 akibat perbedaan format kode undian, memicu lonjakan sentimen negatif. Penelitian ini mengkonfirmasi efektivitas IndoBERT untuk analisis sentimen berbahasa Indonesia dan merekomendasikan *brand* untuk meningkatkan transparansi mekanisme promosi, mengelola ekspektasi konsumen, serta memanfaatkan *insight* sentimen untuk strategi pemasaran kolaborasi di masa mendatang.

Kata kunci: Analisis Sentimen, IndoBERT, Kolaborasi, NCT DREAM, TosTos

PUBLIC SENTIMENT ANALYSIS ON THE MARKETING STRATEGY OF NCT DREAM AND TOSTOS COLLABORATION USING INDOBERT

Mishel Milen

ABSTRACT

Sentiment analysis plays a crucial role in understanding public opinion towards marketing strategies, particularly in the era of collaborations between local brands and international artists on social media. This research aims to implement the Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT) model to analyze public sentiment regarding the NCT DREAM and TosTos collaboration, as well as to evaluate the model's performance in sentiment classification. Public comment data was collected from tostosid's Instagram posts between March 2024 and March 2025, yielding a total of 6,830 comments. The data preprocessing involved case folding, URL removal, emoji replacement and slang word normalization using a custom dictionary, and excess whitespace removal. The IndoBERT model was fine-tuned using 1,800 manually labeled data, split into 70:15:15 for training, validation, and test sets. Evaluation results show that the IndoBERT model demonstrates good capability in sentiment classification, achieving an accuracy of 83% on the validation set, with an F1-score of 0.86 for positive sentiment and 0.84 for negative, although the neutral class (F1-score 0.78) remains a challenge. Sentiment analysis on the entire dataset of comments revealed a dominance of positive sentiment (44.7%), reflecting high public enthusiasm. However, word cloud analysis and trend analysis indicated significant frustration related to the prize lottery mechanism, which culminated in a controversy in August 2024 due to discrepancies in lottery code formats, triggering a surge in negative sentiment. This study confirms IndoBERT's effectiveness for Indonesian sentiment analysis and recommends that brands enhance transparency in promotion mechanisms, manage consumer expectations, and leverage sentiment insights for future collaborative marketing strategies.

Keywords: *Sentiment Analysis, IndoBERT, Collaboration, NCT DREAM, TosTos*

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 KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR KODE	xiv
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.1 Penelitian Pendahulu	6
2.2 Korean Wave (Hallyu)	9
2.3 Analisis Sentimen	10
2.4 Natural Language Processing	11
2.5 Dataset Labeling and Inter-Analyst Validation	11
2.6 Text Preprocessing	13
2.7 Machine Learning	14
2.8 Bidirectional Encoder Representations from Transformers	15
2.8.1 Input dan Output BERT	17
2.8.2 Tahapan Pelatihan Model BERT	18
2.8.3 Hyperparameter Tuning	18
2.8.4 IndoBERT	20
2.9 Pembagian Dataset (Data Splitting)	21
2.10 Confusion Matrix	22
BAB 3 METODOLOGI PENELITIAN	25
3.1 Tahapan Penelitian	25
3.2 Preprocessing Data	29
3.3 Tokenisasi dan Encoding	31
3.4 Perancangan dan Pelatihan Model	32
3.5 Prediksi Data Komentar Baru	35
BAB 4 HASIL DAN DISKUSI	38
4.1 Spesifikasi Perangkat	38
4.2 Pengumpulan Dataset	38
4.3 Hasil Validasi Dataset (Cohen's Kappa)	39
4.4 Install dan Import Library	42
4.5 Data Preprocessing	42

4.6	Pembagian Data	46
4.7	Tokenisasi dan <i>Encoding</i>	47
4.8	Pelatihan Model	48
4.8.1	Hasil Eksperimen Model dan Evaluasi Performa	57
4.9	Prediksi Data Baru	59
4.10	Analisis Sentimen Publik	65
4.11	Pembahasan	71
BAB 5	SIMPULAN DAN SARAN	73
5.1	Simpulan	73
5.2	Saran	74
DAFTAR PUSTAKA		76



DAFTAR TABEL

Tabel 2.1	Penelitian pendahulu	6
Tabel 2.2	Skala interpretasi nilai koefisien Cohen's Kappa	13
Tabel 2.3	<i>Confusion matrix</i>	22
Tabel 4.1	Hasil <i>dataset</i> yang diperoleh dari <i>web scraping</i>	39
Tabel 4.2	Matriks konfusi hasil pelabelan	40
Tabel 4.3	Hasil <i>labelling</i> manual	43
Tabel 4.4	Hasil <i>preprocessing</i> komentar	45
Tabel 4.5	Hasil perubahan label sentimen menjadi numerik	46
Tabel 4.6	Hasil pelatihan dan validasi model dengan pembagian data 70:15:15	50
Tabel 4.7	Hasil evaluasi model pada data validasi dengan pembagian data 70:15:15	51
Tabel 4.8	<i>Confusion matrix</i> dari hasil evaluasi model dengan pembagian data 70:15:15	52
Tabel 4.9	Hasil pelatihan dan validasi model dengan pembagian data 80:10:10	54
Tabel 4.10	Hasil evaluasi model pada data validasi dengan pembagian data 80:10:10	54
Tabel 4.11	<i>Confusion matrix</i> dari hasil evaluasi model dengan pembagian data 80:10:10	55
Tabel 4.12	Hasil perbandingan skenario eksperimen	58
Tabel 4.13	Hasil evaluasi model pada data uji yang belum memiliki label	62
Tabel 4.14	<i>Confusion matrix</i> dari hasil evaluasi model pada data uji <i>unlabeled</i>	63
Tabel 4.15	Hasil sentimen komentar yang salah diprediksi	64

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 GAMBAR

Gambar 2.1	<i>Encoder</i>	16
Gambar 2.2	Ilustrasi proses <i>input</i> dan <i>output</i> BERT	17
Gambar 2.3	Ilustrasi proses <i>pre-training</i> dan <i>fine-tuning</i>	18
Gambar 2.4	Prosedur Optuna untuk <i>hyperparameter tuning</i>	19
Gambar 3.1	Alur penelitian	25
Gambar 3.2	<i>Flowchart preprocessing</i> data	30
Gambar 3.3	<i>Flowchart</i> tokenisasi dan <i>encoding</i>	31
Gambar 3.4	<i>Flowchart</i> perancangan dan pelatihan model	33
Gambar 3.5	<i>Flowchart</i> model <i>training</i>	34
Gambar 3.6	<i>Flowchart</i> prediksi data komentar baru	36
Gambar 4.1	Implementasi <i>emoji replacement</i>	44
Gambar 4.2	Grafik <i>training</i> dan <i>validation loss</i> pada pembagian data 70:15:15	52
Gambar 4.3	Grafik metrik performa akurasi dan <i>F1-score</i> pada pembagian data 70:15:15	53
Gambar 4.4	Grafik <i>training</i> dan <i>validation loss</i> pada pembagian data 80:10:10	56
Gambar 4.5	Grafik metrik performa akurasi dan <i>F1-score</i> pada pembagian data 80:10:10	57
Gambar 4.6	Distribusi sentimen komentar publik dalam bentuk <i>pie chart</i>	66
Gambar 4.7	Distribusi sentimen komentar publik dalam bentuk <i>bar chart</i>	66
Gambar 4.8	<i>Word cloud</i> komentar positif	67
Gambar 4.9	<i>Word cloud</i> komentar negatif	68
Gambar 4.10	<i>Word cloud</i> komentar netral	69
Gambar 4.11	Analisis tren sentimen berdasarkan tanggal komentar	70



DAFTAR KODE

Kode 4.1	Perhitungan Cohen's Kappa	41
Kode 4.2	<i>Install</i> dan <i>import library</i>	42
Kode 4.3	<i>Preprocessing</i> data	43
Kode 4.4	Implementasi kamus <i>abbreviations</i>	43
Kode 4.5	<i>Cleaning dataset</i>	44
Kode 4.6	Proses <i>label encoding</i>	45
Kode 4.7	Proses <i>split</i> data 70:15:15	46
Kode 4.8	Proses <i>split</i> data 80:10:10	47
Kode 4.9	Proses <i>tokenization</i> dan <i>encoding</i>	47
Kode 4.10	Proses perancangan model	48
Kode 4.11	Konfigurasi parameter pelatihan model dan inisialisasi <i>Trainer</i>	49
Kode 4.12	Proses memuat data komentar tanpa label	59
Kode 4.13	Proses tokenisasi dan pembuatan kelas <i>dataset</i> kustom	59
Kode 4.14	Prediksi sentimen otomatis menggunakan model terlatih	60
Kode 4.15	Proses evaluasi hasil prediksi model	61
Kode 4.16	Menampilkan komentar yang salah diprediksi	63
Kode 4.17	Perhitungan persentase kesalahan prediksi per kelas	64



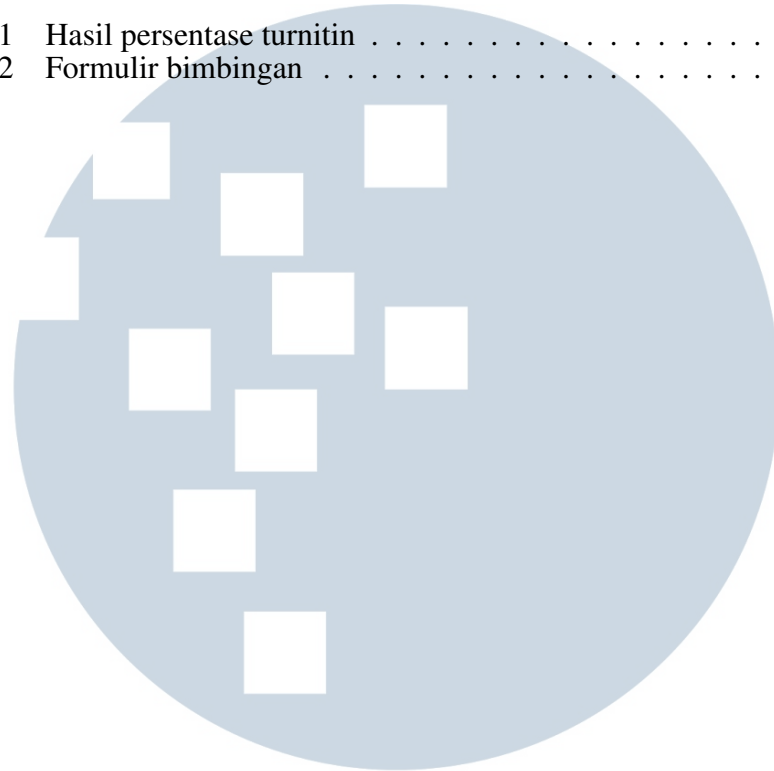
DAFTAR RUMUS

Rumus 2.1	<i>Cohen's Kappa</i>	12
Rumus 2.2	<i>Accuracy</i>	23
Rumus 2.3	<i>Precision</i>	23
Rumus 2.4	<i>Recall</i>	24
Rumus 2.5	<i>F1-score</i>	24



DAFTAR LAMPIRAN

Lampiran 1	Hasil persentase turnitin	82
Lampiran 2	Formulir bimbingan	95



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA