

**ANALISIS SENTIMEN DAN TREN PERCAKAPAN PUBLIK TENTANG
ISU MAFIA BBM DI MEDIA SOSIAL TIKTOK DENGAN PENDEKATAN
KLASIFIKASI BERBASIS MODEL TRANSFORMER INDOBERT**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi *SI Informatika*



disusun oleh

Winda Yulia Malaha

22.11.5071

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN DAN TREN PERCAKAPAN PUBLIK TENTANG
ISU MAFIA BBM DI MEDIA SOSIAL TIKTOK DENGAN PENDEKATAN
KLASIFIKASI BERBASIS MODEL TRANSFORMER INDOBERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI Informatika*



disusun oleh

Winda Yulia Malaha

22.11.5071

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN DAN TREN PERCAKAPAN PUBLIK TENTANG ISU MAFIA BBM DI MEDIA SOSIAL TIKTOK DENGAN PENDEKATAN KLASIFIKASI BERBASIS MODEL TRANSFORMER INDOBERT

yang disusun dan diajukan oleh

Winda Yulia Malaha

22.11.5071

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Januari 2026

Dosen Pembimbing,



I Made Artha Agastya, S.T., M.Eng., Ph.D
NIK. 190302352

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN DAN TREN PERCAKAPAN PUBLIK TENTANG ISU MAFIA BBM DI MEDIA SOSIAL TIKTOK DENGAN PENDEKTAN KLASIFIKASI BERBASIS MODEL TRANSFORMER INDOBERT

yang disusun dan diajukan oleh

Winda Yulia Malaha

22.11.5071

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rizqi Sukma Kharisma, S.Kom., M.Kom.
NIK. 190302215

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

...

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Winda Yulia Malaha
NIM : 22.11.5071

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN DAN TREN PERCAKAPAN PUBLIK TENTANG
ISU MAFIA BBM DI MEDIA SOSIAL TIKTOK DENGAN PENDEKATAN
KLASIFIKASI BERBASIS MODEL TRANSFORMER INDOBERT**

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 21 Januari 2026

Yang Menyatakan,



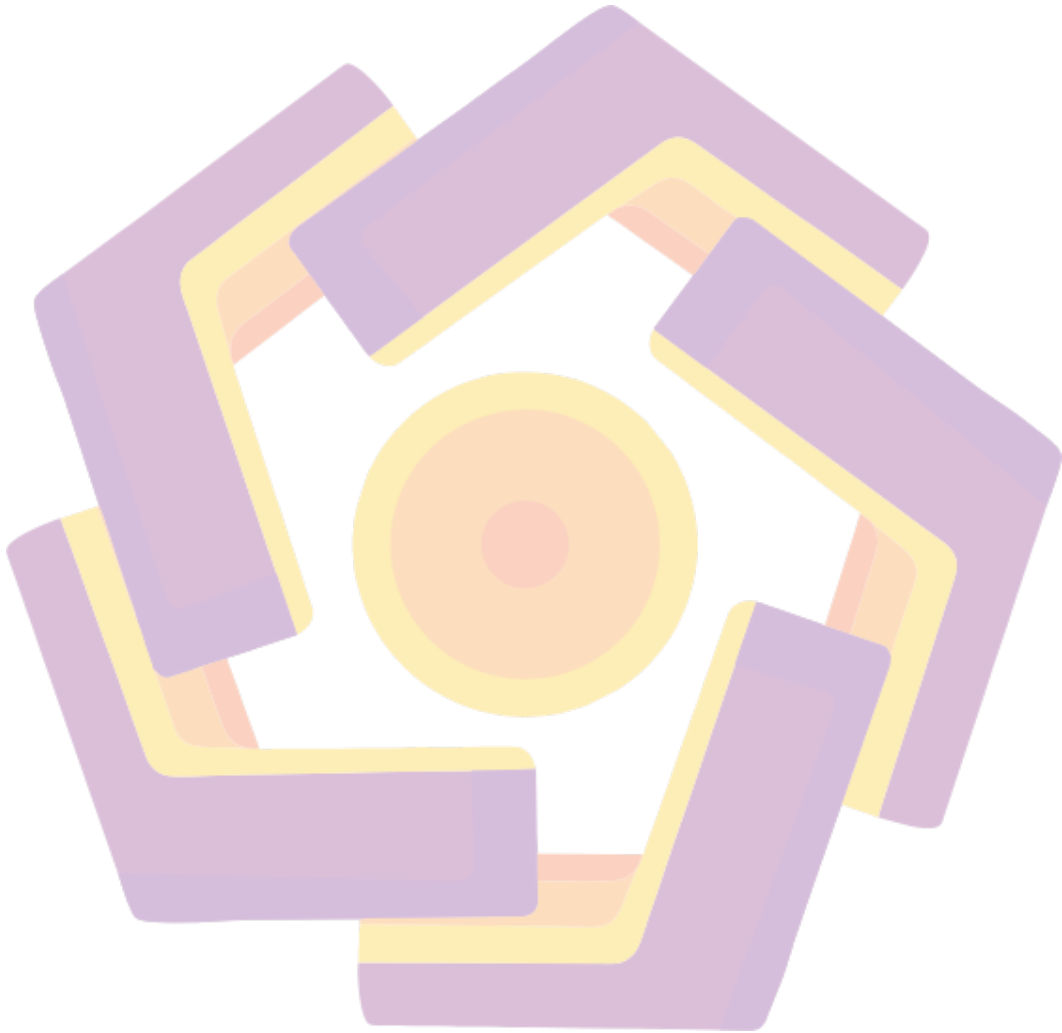
Winda Yulia Malaha

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penulis diberi kekuatan dan kemudahan dalam menyelesaikan skripsi ini. Shalawat serta salam semoga selalu tercurah kepada junjungan kita Nabi Muhammad SAW. Atas izin Allah SWT, skripsi ini dapat diselesaikan melalui berbagai usaha dan perjuangan. Skripsi ini penulis persembahkan kepada :

1. Untuk kedua orang tua tercinta Bpk Hajunil Malaha dan Ibu Yoce Altince Tampese. Terimakasih selalu meberikan kasih sayang, doa dan dukungan. Selalu hadir dalam setiap proses, menguatkan saat lelah, menenangkan saat ragu dan menyemangati ketika langka terasa berat. Terima kasih juga untuk tidak menuntut harus menjadi sesuatu, tetapi justru selalu memberi ruang untuk jalan sendiri. Terima kasih karena setiap keinginan selalu diusahakan, setiap kebutuhan selalu di penuhi, dan setiap mimpi selalu disambut dengan dukungan yang tulus. Apresiasi yang sangat besar untuk Papa dan Mama atas kerja keras, perjuangan, dan kontribusi dalam segala hal, termasuk dukungan materi maupun perhatian yang tidak pernah kurang.
2. Untuk kakek dan nenek. Terima kasih selalu mendoakan dan memberikan dukungan tanpa henti, baik secara materi maupun perhatian kepada penulis.
3. Terima kasih kepada Wahyu yang telah menjadi bagian penting dalam perjalanan hidup penulis. Terima kasih atas segala perhatian yang di berikan, mulai dari waktu, dukungan materi maupun doa.
4. Kepada Rifka dan Risma. Terima kasih selalu memberikan semangat dan menjadi tempat bercerita.
5. Teima kasih kepada teman teman IF9 dan IF5 atas kebersamaan nya selama proses perkuliahan.

6. *Last but not least*, I wanna thank me. I wanna thank me for believing in me, I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting.



KATA PENGANTAR

Segala Puji bagi Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul ***“Analisis Sentimen Dan Tren Percakapan Publik Tentang Isu Mafia BBM Dengan Pendekatan Klasifikasi Berbasis Model Transformer IndoBERT”***. Skripsi ini di susun untuk memenuhi salah satu syarat untuk memperoleh gelar sarjana pada program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penyusunan skripsi ini tidak dapat terselesaikan dengan baik tanpa adanya dukungan, bimbingan, dan bantuan dari beberapa pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar besarnya kepada :

1. Bapak Prof.Dr.Mohammad Suyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta
2. Ibu Prof.Dr.Kusrini, M.Kom. selaku dekan fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti M.Kom. selaku ketua prodi Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak I Made Artha Agastya S.T., M.Eng., Ph.D. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan selama proses penyusunan skripsi.
5. Seluruh dosen dan staf Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan ilmu, fasilitas, dan pelayanan selama penulis menempuh studi.
6. Semua pihak yang membantu dan berpartisipasi dalam penyusunan skripsi ini yang tidak bisa di sebutkan satu persatu.

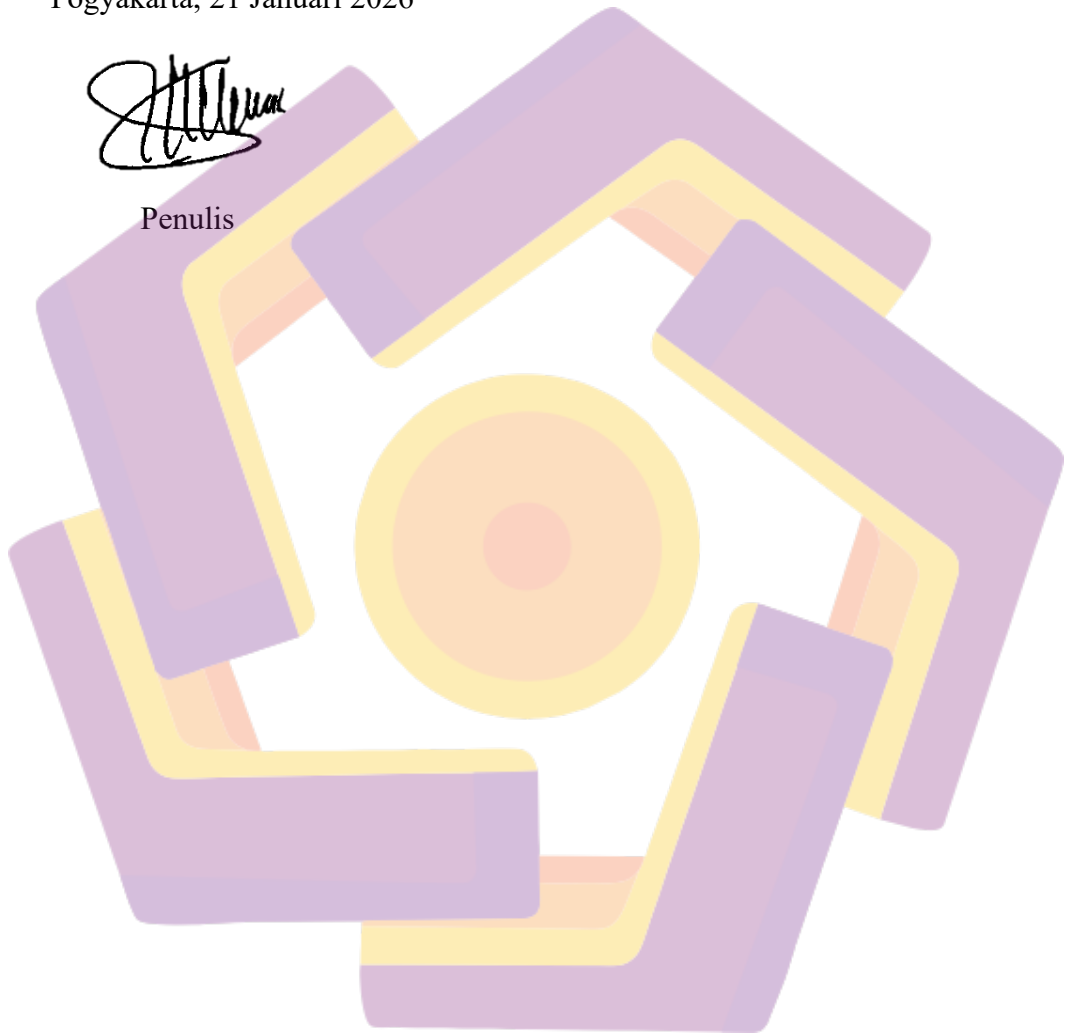
Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, baik dari segi penulisan maupun penyajian. Oleh karena itu, kritik dan saran yang membuat penulis berharap sebagai bahan pertimbangan dan perbaikan di masa mendatang.

Penulis berharap semoga skripsi ini dapat memberikan manfaat bagi penulis dan bagi pembacanya.

Yogyakarta, 21 Januari 2026



Penulis



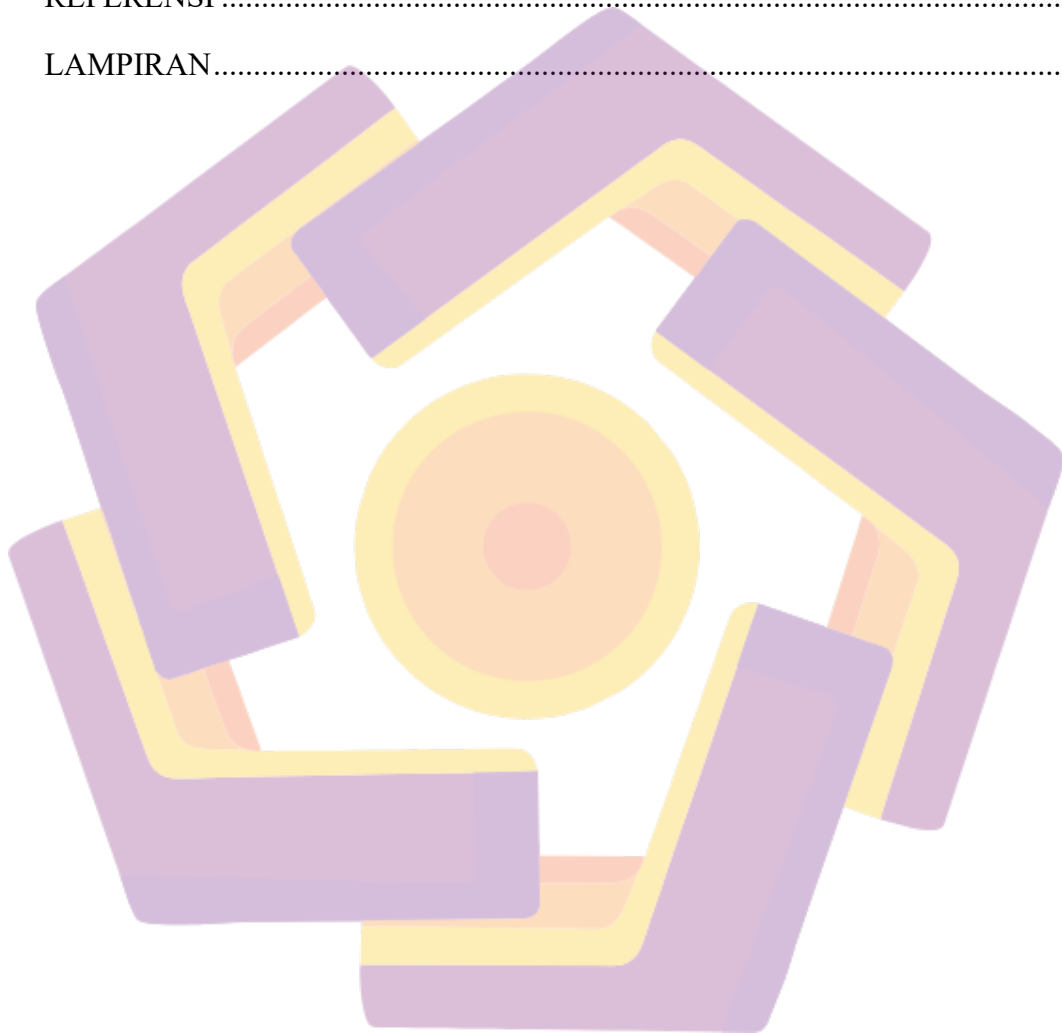
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.5.1 Manfaat Teoretis	4
1.5.2 Manfaat praktis	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7

2.1	Studi Literatur	7
2.2	Dasar Teori.....	18
2.2.1	Media Sosial TikTok.....	18
2.2.2	Text Mining	18
2.2.3	Web Scraping	20
2.2.4	Analisis Sentimen	20
2.2.5	Pelabelan Otomatis Berbasis Leksikon (InSet).....	21
2.2.6	Pengolahan Awal Teks (Text Preprocessing)	22
2.2.7	IndoBERT	23
2.2.8	Hyperparameter.....	24
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian.....	26
3.1.1	Sumber Data.....	26
3.1.2	Teknik Pengumpulan Data.....	26
3.2	Alur Penelitian	26
3.3	Tahapan Preprocessing Data.....	31
3.3.1	Cleaning Text.....	31
3.3.2	Case Folding	31
3.3.3	Normalisasi Kata.....	32
3.3.4	Stopword Removal.....	32
3.3.5	Penghapusan Data Duplikat dan Data Kosong	32
3.4	Proses Pelabelan Data Sentimen.....	33
3.5	Analisis Eksplorasi Data (Exploratory Data Analysis).....	33
3.6	Pembagian Data (Data Splitting)	34
3.7	Arsitektur dan Konfigurasi Model	34

3.8	Optimasi Hyperparameter Model IndoBERT	35
3.9	Proses Pelatihan Model.....	36
3.10	Evaluasi Akhir Model pada Data Uji.....	37
3.10.1	Accuracy	37
3.10.2	Precision, recall, dan F1-Score.....	38
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Preprocessing Data.....	40
4.2	Pelabelan Data Sentimen	43
4.3	Analisis Eksplorasi Data (EDA)	44
4.3.1	Analisis Statistik Deskriptif Dataset	44
4.3.2	Analisis Distribusi Sentimen.....	45
4.3.3	Analisis Panjang Teks Komentar	46
4.3.4	Analisis Komposisi Sentimen dalam Persentase	47
4.3.5	Analisis Frekuensi Kata	49
4.4	Pembagian Dataset.....	49
4.5	Hasil Hyperparameter Tuning.....	51
4.5.1	Hasil Hyperparameter Tuning Menggunakan Manual Tuning	51
4.5.2	Hasil Hyperparameter Tuning Menggunakan Grid Search.....	52
4.5.3	Hasil Hyperparameter Tuning Menggunakan Random Search	54
4.5.4	Hasil Hyperparameter Tuning Menggunakan Genetic Algorithm.....	55
4.5.5	Perbandingan Hasil Hyperparameter Terbaik.....	56
4.6	Proses Pelatihan Model IndoBERT	57
4.6.1	Perkembangan Nilai Loss dan Metrik Evaluasi.....	58
4.6.2	Early Stopping dan Pemilihan Model Terbaik.....	60
4.7	Evaluasi Kinerja Model Klasifikasi Sentimen	60

4.8	Prediksi Sentimen Komentar	63
BAB V PENUTUP		65
5.1	Kesimpulan	65
5.2	Saran	66
REFERENSI		69
LAMPIRAN.....		73

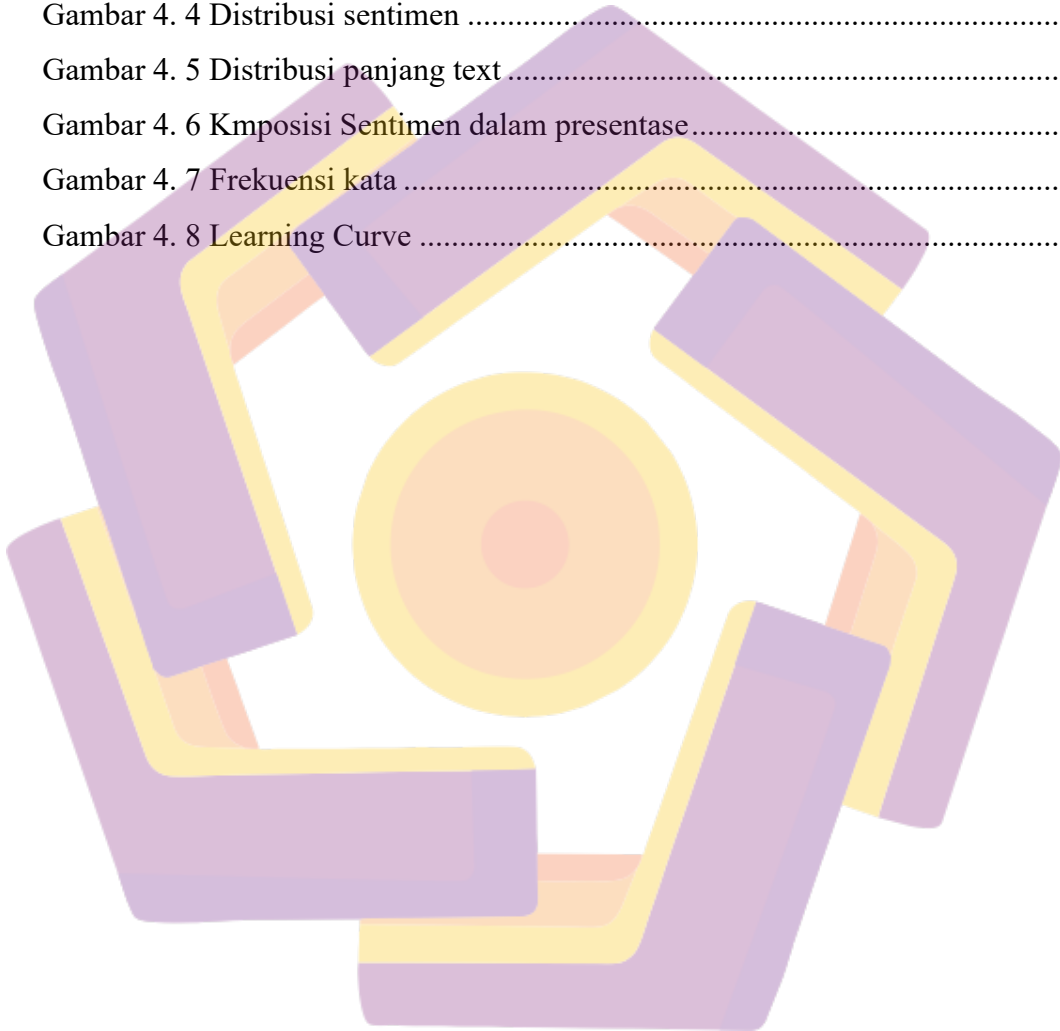


DAFTAR TABEL

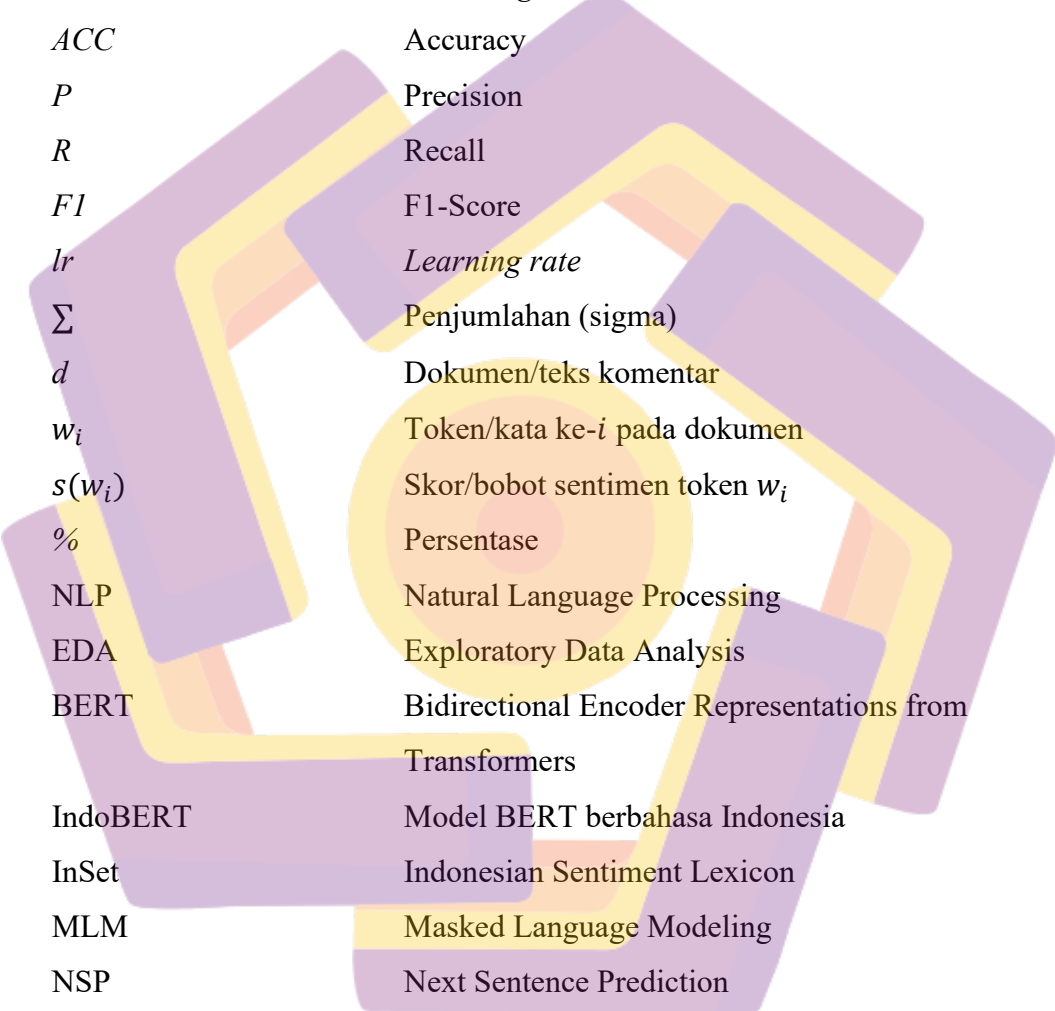
Tabel 2. 1 Keaslian Penelitian	11
Tabel 4. 1 Perbandingan Teks Sebelum dan Sesudah Preprocessing	41
Tabel 4. 2 Hasil pembagian dataset split.....	50
Tabel 4. 3 Hasil Hyperparaemetr Manual Tuning	51
Tabel 4. 4 Hasil Hyperparameter Tuning Menggunakan Grid Search	52
Tabel 4. 5 Hasil Hyperparameter Tuning Menggunakan Random Search	54
Tabel 4. 6 Hasil Hyperparameter Tuning Menggunakan Genetic Algorithm.....	55
Tabel 4. 7 Perbandingan Hasil Hyperparameter Terbaik.....	56
Tabel 4. 8 Ringkasan Hasil Pelatihan per Epoch	58
Tabel 4. 9 Ringkasan Hasil Uji (Test Set)	61
Tabel 4. 10 Metrik Evaluasi Global pada Data Uji.....	62
Tabel 4. 11 Laporan Klasifikasi pada Data Uji.....	62
Tabel 4. 12 Contoh Hasil Prediksi Sentimen Komentar	63

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	27
Gambar 4. 1 Wordcloud Perbandingan Hasil Preprocessing	43
Gambar 4. 2 Hasil pelabelan data	44
Gambar 4. 3 Hasil statistik deskriptif.....	45
Gambar 4. 4 Distribusi sentimen	46
Gambar 4. 5 Distribusi panjang text	47
Gambar 4. 6 Kmposisi Sentimen dalam presentase.....	48
Gambar 4. 7 Frekuensi kata	49
Gambar 4. 8 Learning Curve	58

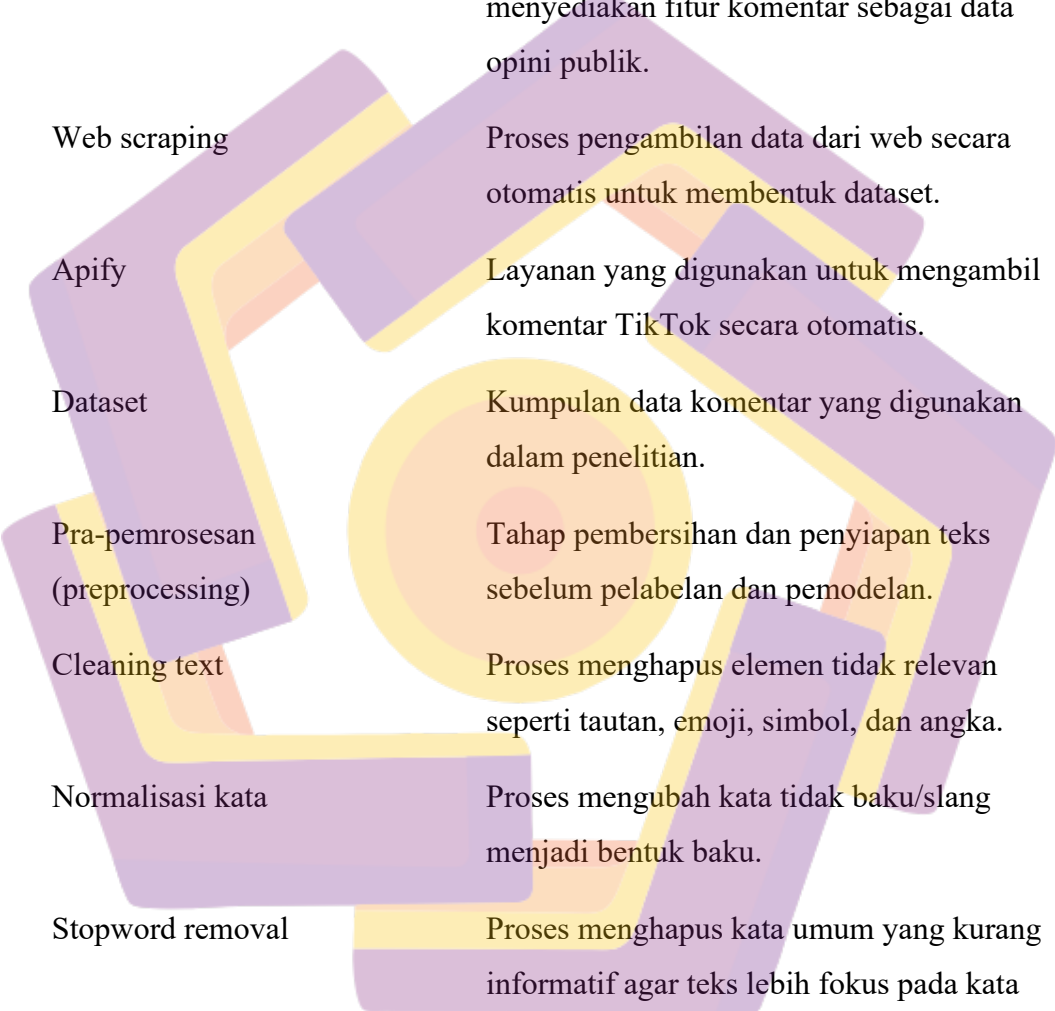


DAFTAR LAMBANG DAN SINGKATAN

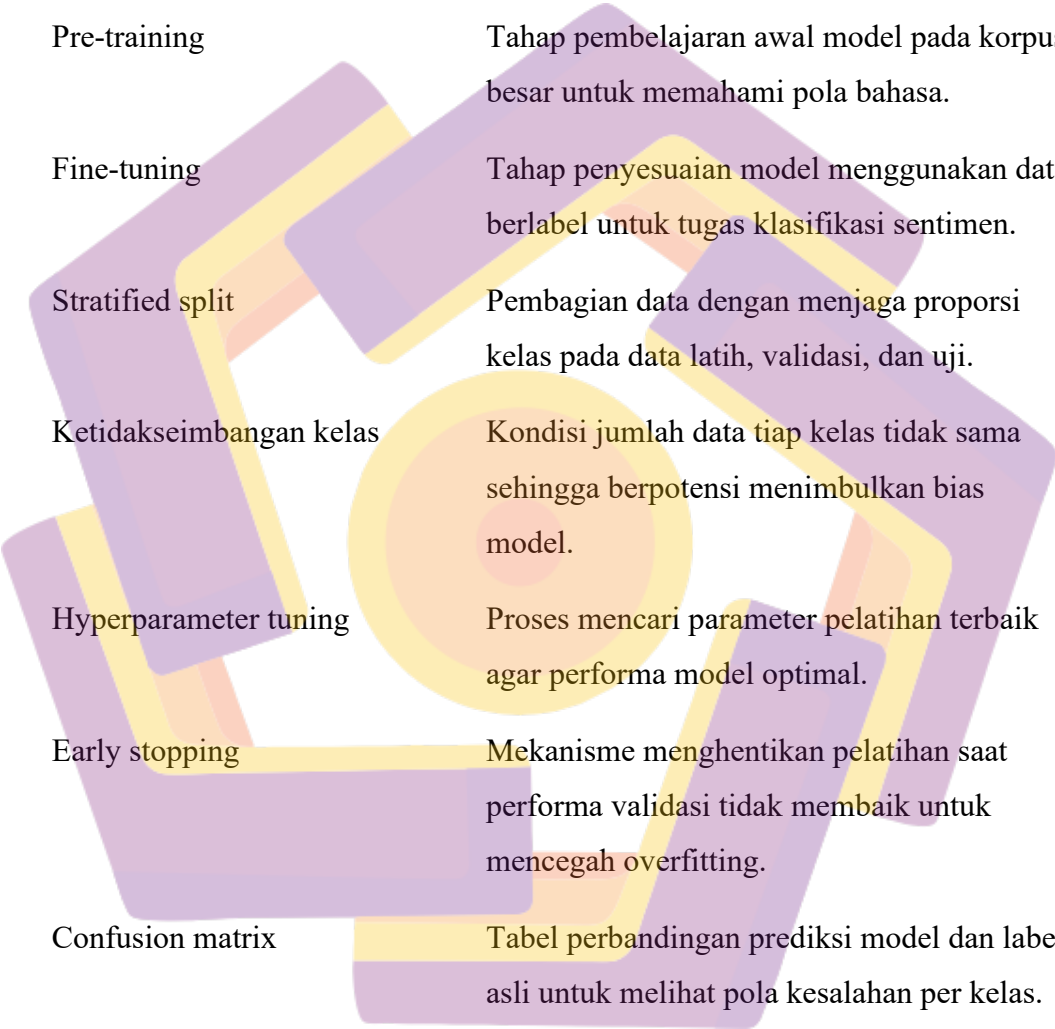


TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positiv</i>
FN	<i>False Negative</i>
ACC	Accuracy
P	Precision
R	Recall
$F1$	F1-Score
lr	<i>Learning rate</i>
Σ	Penjumlahan (sigma)
d	Dokumen/teks komentar
w_i	Token/kata ke- i pada dokumen
$s(w_i)$	Skor/bobot sentimen token w_i
$\%$	Persentase
NLP	Natural Language Processing
EDA	Exploratory Data Analysis
BERT	Bidirectional Encoder Representations from Transformers
IndoBERT	Model BERT berbahasa Indonesia
InSet	Indonesian Sentiment Lexicon
MLM	Masked Language Modeling
NSP	Next Sentence Prediction
SMOTE	Synthetic Minority Over-sampling Technique
CSV	Comma-Separated Values
URL	Uniform Resource Locator
GA	Genetic Algorithm
Bidirectional	Dua Arah

DAFTAR ISTILAH



Analisis sentimen	Proses mengelompokkan opini dalam teks ke kategori sentimen (positif, netral, negatif).
TikTok	Platform media sosial video pendek yang menyediakan fitur komentar sebagai data opini publik.
Web scraping	Proses pengambilan data dari web secara otomatis untuk membentuk dataset.
Apify	Layanan yang digunakan untuk mengambil komentar TikTok secara otomatis.
Dataset	Kumpulan data komentar yang digunakan dalam penelitian.
Pra-pemrosesan (preprocessing)	Tahap pembersihan dan penyiapan teks sebelum pelabelan dan pemodelan.
Cleaning text	Proses menghapus elemen tidak relevan seperti tautan, emoji, simbol, dan angka.
Normalisasi kata	Proses mengubah kata tidak baku/slang menjadi bentuk baku.
Stopword removal	Proses menghapus kata umum yang kurang informatif agar teks lebih fokus pada kata bermakna.
Pelabelan otomatis	Pemberian label sentimen tanpa anotasi manual, menggunakan kamus sentimen.

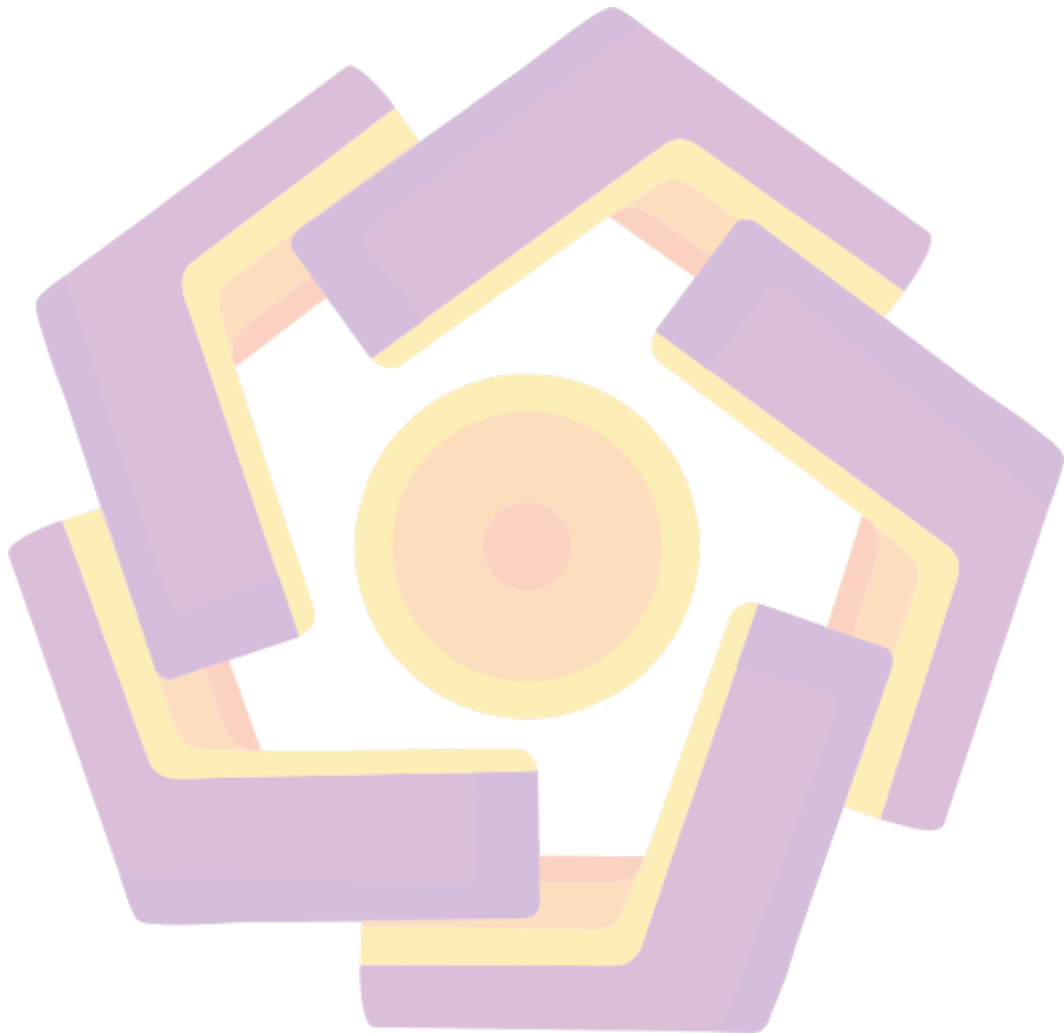


Leksikon InSet	Kamus sentimen Bahasa Indonesia berisi daftar kata positif dan negatif untuk pelabelan awal.
IndoBERT	Model bahasa berbahasa Indonesia yang digunakan untuk klasifikasi sentimen.
Pre-training	Tahap pembelajaran awal model pada korpus besar untuk memahami pola bahasa.
Fine-tuning	Tahap penyesuaian model menggunakan data berlabel untuk tugas klasifikasi sentimen.
Stratified split	Pembagian data dengan menjaga proporsi kelas pada data latih, validasi, dan uji.
Ketidakseimbangan kelas	Kondisi jumlah data tiap kelas tidak sama sehingga berpotensi menimbulkan bias model.
Hyperparameter tuning	Proses mencari parameter pelatihan terbaik agar performa model optimal.
Early stopping	Mekanisme menghentikan pelatihan saat performa validasi tidak membaik untuk mencegah overfitting.
Confusion matrix	Tabel perbandingan prediksi model dan label asli untuk melihat pola kesalahan per kelas.
Macro average	Rata-rata metrik per kelas tanpa mempertimbangkan jumlah data tiap kelas.
Komputasional	cara kerja yang mengandalkan perhitungan dengan komputer untuk mengolah dan

menganalisis data secara otomatis, terutama ketika datanya besar

Konvensional

Cara yang umum/tradisional, biasanya dilakukan secara manual atau dengan langkah sederhana.



INTISARI

Isu mafia BBM memicu beragam respon masyarakat yang terekam pada kolom komentar TikTok. Untuk melihat kecenderungan opini publik secara lebih terstruktur, dilakukan analisis sentimen menggunakan model IndoBERT. Data komentar yang bersifat teks bebas diproses melalui pembersihan teks, penyamaan format penulisan, normalisasi kata, penghapusan kata umum yang kurang membantu, serta penghapusan data kosong dan duplikat agar data lebih konsisten. Pelabelan sentimen dilakukan secara otomatis menggunakan leksikon InSet sehingga terbentuk tiga kelas sentimen negatif, netral, dan positif. Model IndoBERT dilatih menggunakan pembagian data latih, validasi, dan uji dengan proporsi kelas yang dijaga, serta dipilih konfigurasi terbaik dengan mencoba beberapa setelan pelatihan yang berbeda. Evaluasi pada data uji menghasilkan akurasi 0,8786 dengan F1 score macro 0,8785 dan F1 score weighted 0,8793. Secara per kelas, performa terbaik terdapat pada sentimen negatif dengan F1 score 0,90, disusul positif 0,88 dan netral 0,86. Kesalahan klasifikasi paling banyak terjadi pada kelas netral, yang menunjukkan bahwa komentar netral sering memiliki ciri yang mirip dengan komentar negatif atau positif. Hasil ini menunjukkan IndoBERT cukup efektif untuk klasifikasi sentimen komentar TikTok terkait isu mafia BBM dan dapat digunakan sebagai dasar pemetaan kecenderungan opini publik.

Kata kunci: Analisis sentimen, Tiktok, IndoBERT, Inset, Mafia BBM

ABSTRACT

The fuel mafia issue triggered a variety of public responses recorded in TikTok's comments section. To examine public opinion trends in a more structured manner, sentiment analysis was conducted using the IndoBERT model. The free-text comment data was processed through text cleaning, formatting, word normalization, removal of unhelpful common words, and removal of empty and duplicate data to ensure consistency. Sentiment labeling was performed automatically using the InSet lexicon, resulting in three sentiment classes: negative, neutral, and positive. The IndoBERT model was drilled using a split training, validation, and test dataset with a maintained class proportion, and the best configuration was selected by testing several different training settings. Evaluation on the test data yielded an accuracy of 0.8786 with a macro F1 score of 0.8785 and a weighted F1 score of 0.8793. Per class, the best performance was for negative sentiment with an F1 score of 0.90, followed by positive with 0.88 and neutral with 0.86. The most misclassification errors occurred in the neutral class, indicating that neutral comments often share characteristics with negative or positive comments. These results show that IndoBERT is quite effective for classifying TikTok comment sentiment related to the BBM mafia issue and can be used as a basis for mapping public opinion trends.

Keyword: Sentiment Analyst, TikTok, IndoBERT, Inset, Fuel Mafia