

**ANALISIS SENTIMEN KOMENTAR YOUTUBE PADA
INSTITUSI KEPOLISIAN INDONESIA MENGGUNAKAN
ALGORITMA BERT
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

BAYU APRIANSYAH

20.11.3506

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**SENTIMEN ANALISIS KOMENTAR YOUTUBE PADA
INSTITUSI KEPOLISIAN INDONESIA MENGGUNAKAN
ALGORITMA BERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

BAYU APRIANSYAH

20.11.3506

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR YOUTUBE PADA INSTITUSI KEPOLISIAN
INDONESIA MENGGUNAKAN ALGORITMA BERT**

yang disusun dan diajukan oleh

Bayu Apriansyah

20.11.3506

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Juni 2025

Dosen Pembimbing,



Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN KOMENTAR YOUTUBE PADA INSTITUSI KEPOLISIAN
INDONESIA MENGGUNAKAN ALGORITMA BERT

yang disusun dan diajukan oleh

Bayu Apriansyah

20.11.3506

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392

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

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Bayu Apriansyah
NIM : 20.11.3506

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Komentar Youtube Pada Institusi Kepolisian Indonesia
Menggunakan Algoritma BERT**

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

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, 30 Juli 2025

Yang Menyatakan,



Bayu Apriansyah

HALAMAN PERSEMBAHAN

Puji dan Syukur Saya panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya sehingga skripsi ini dapat diselesaikan dengan baik. Oleh karena itu Saya ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, Bapak Subandi, S.Pdi dan Ibu Ismarmi yang selalu memberikan cinta dan doa yang tak henti serta menjadi faktor utama dalam setiap pengambilan keputusan.
2. Saudara dan keluarga besar, yang selalu memberikan dorongan dan kehangatan yang tak tergantikan
3. Para Dosen pembimbing, yang telah memberikan ilmu, bimbingan serta motivasi dalam menyelesaikan skripsi ini. Terimakasih atas segala arahan dan dukungannya.
4. Terima kasih kepada Chandra Niko Pasaribu dengan segala bantuan teknis, kesabaran dalam menemani proses *coding*, serta diskusi produktif yang membuka wawasan penulis mengenai alur penelitian. Tanpa bantuannya, penulis akan menghadapi banyak kesulitan dalam menyelesaikan aspek teknis pada penelitian ini.
5. Penulis juga berterima kasih kepada Cindy Ellysa atas segala dukungan, semangat, dan kepercayaan yang tak henti-hentinya diberikan. Kehadirannya menjadi pengingat bagi penulis untuk terus berjuang dan menyelesaikan tanggung jawab ini dengan sebaik-baiknya.
6. Teman-teman yang ikut terlibat dalam proses penyelesaian skripsi.
7. Almamater tercinta, yang telah menjadi tempat saya menimba ilmu dan mengembangkan diri

Semoga skripsi ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi pembaca dan pengembangan ilmu pengetahuan.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul **"ANALISIS SENTIMEN KOMENTAR YOUTUBE PADA INSTITUSI KEPOLISIAN INDONESIA MENGGUNAKAN ALGORITMA BERT"** dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa tidak sedikit bantuan, bimbingan, dan dukungan dari berbagai pihak yang telah banyak berkontribusi. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Majid Rahardi, S.Kom., M.Eng., selaku dosen pembimbing yang dengan penuh kesabaran, memberikan bimbingan, arahan, dan motivasi dalam proses penyusunan skripsi ini.
2. Dosen Penguji, yang telah memberikan kritik, saran, serta masukan yang membangun demi kesempurnaan skripsi ini.
3. Orang tua tercinta dan keluarga besar yang selalu memberikan doa dan dukungan tiada henti.
4. Seluruh pihak yang tidak bisa disebutkan satu per satu, yang telah membantu penulis, baik secara langsung maupun tidak langsung, dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, baik dari segi isi maupun penyajiannya. Oleh karena itu, penulis terbuka terhadap segala bentuk kritik dan saran yang membangun untuk perbaikan di masa mendatang.

Yogyakarta, 29 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3

1.6	Sistematika Penulisan	3
	BAB I Pendahuluan.....	4
	Bab II Landasan Teori	4
	Bab V Penutup	4
	BAB II TINJAUAN PUSTAKA	5
2.1	Studi Literatur	5
2.2	Dasar Teori.....	14
2.2.1	Institusi Kepolisian Negara Republik Indonesia (Polri)	14
2.2.2	Analisis Sentimen	14
2.2.3	YouTube	15
2.2.4	Natural Language Processing (NLP)	15
2.2.5	Pengumpulan Data	16
2.2.6	Text Preprocessing.....	16
2.2.7	Arsitektur Transformer	16
2.2.8	BERT	18
2.2.9	IndoBERT	20
2.2.10	Prediksi Dan Evaluasi	21
	BAB III METODE PENELITIAN	24
3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	25
3.2.1	Pengumpulan Data	26
3.3	Alat dan Bahan.....	33

BAB IV HASIL DAN PEMBAHASAN	34
4.1 Pengumpulan Data	34
4.2 Preprocessing Data.....	34
4.2.1 Casefolding	35
4.2.2 Text Cleaning.....	35
4.2.3 Text Normalization	36
4.3 Labeling	37
4.4 Analisis Data Eksplorasi	38
4.5 Data Preparation.....	43
4.5.1 Pembagian Dataset.....	43
4.5.2 Tokenizing	44
4.5.3 Data Loader.....	46
4.6 Implementasi Dan Evaluasi Model.....	47
4.6.1 Fine Tuning BERT.....	47
4.6.2 Hasil Evaluasi Model.....	50
BAB V PENUTUP	54
5.1 Kesimpulan	54
5.2 Saran	54
REFERENSI	56

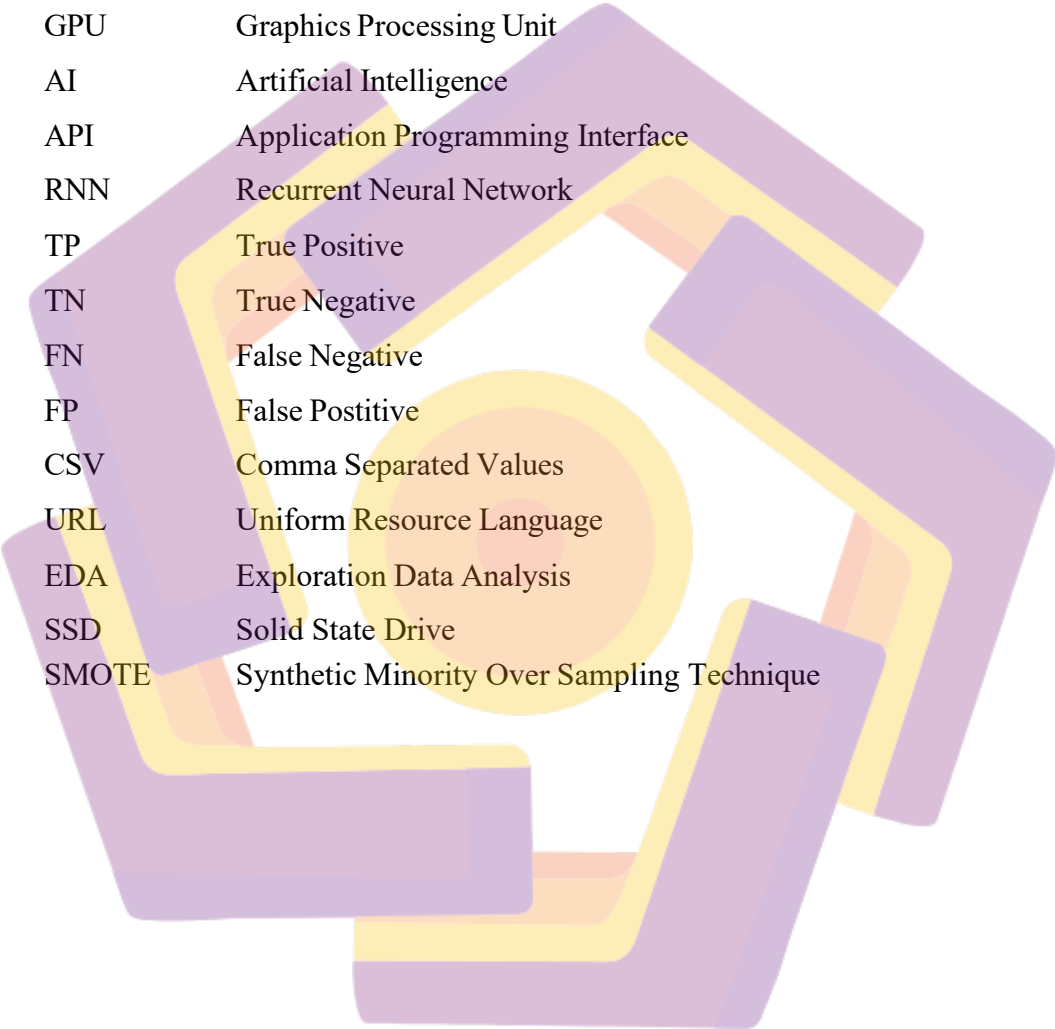
DAFTAR TABEL

Tabel 2. 1 Tabel Keaslian	9
Tabel 4. 1 Hasil Text Cleaning	35
Tabel 4. 2 Text Normalization	37
Tabel 4. 3 Labeling	38
Tabel 4. 4 Jumlah Data Sentimen	38
Tabel 4. 5 Hasil Pembagian Dataset	44
Tabel 4. 6 Hasil Evaluasi Model Pada Data Uji	50
Tabel 4. 7 Laporan Klasifikasi Pada Data Uji	51

DAFTAR GAMBAR

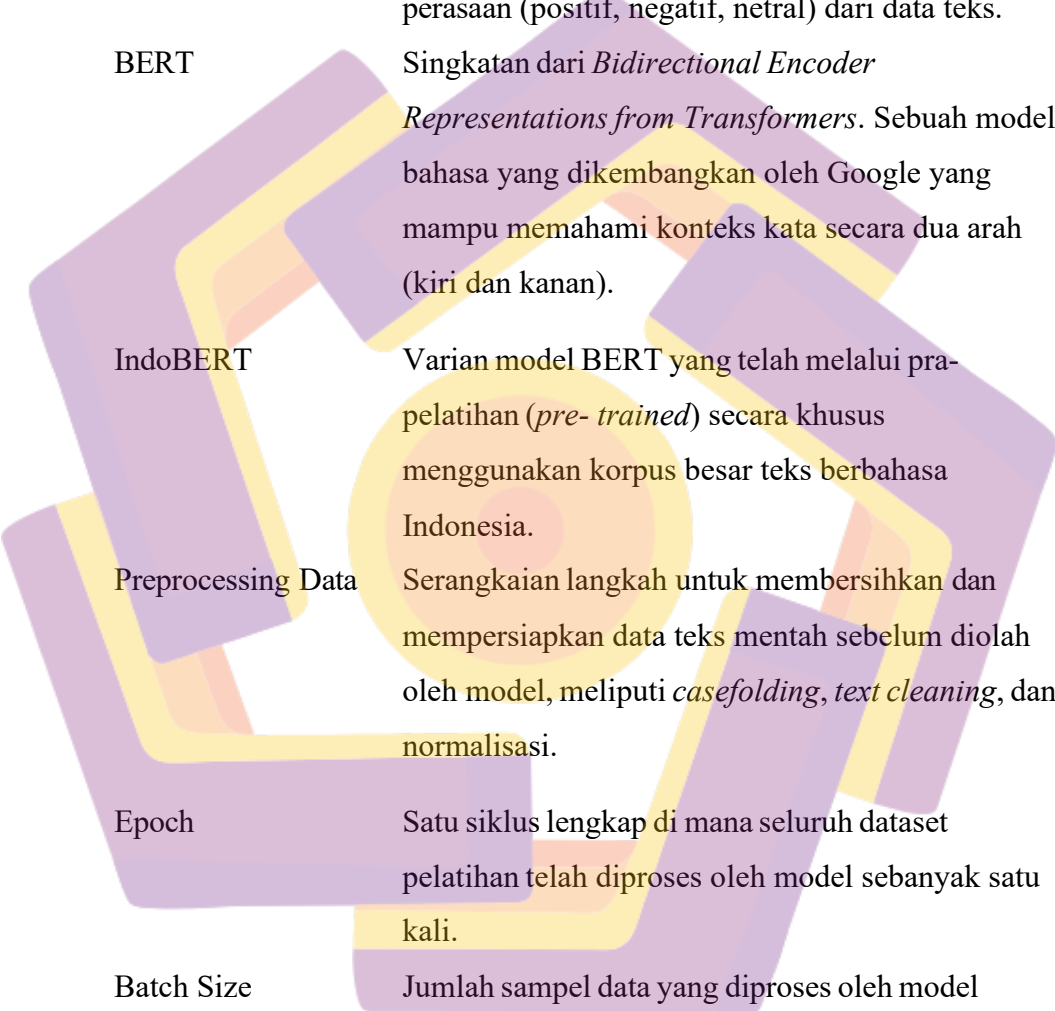
Gambar 2. 1 Arsitektur Transformer	17
Gambar 2. 2 Fine Tuning BERT	20
Gambar 3. 1 Alur Penelitian	25
Gambar 3. 2 Pengumpulan Data	26
Gambar 3. 3 Case Folding	27
Gambar 3. 4 Text Normalization	28
Gambar 4. 1 Hasil Scraping Data	34
Gambar 4. 2 Distribusi Kelas Sentimen	39
Gambar 4. 3 Visualisasi Kategori Positif	40
Gambar 4. 4 Visualisasi Kategori Negatif	41
Gambar 4. 5 Visualisasi Kategori Netral	42
Gambar 4. 6 Visualisasi Semua Kategori	42
Gambar 4. 7 Hasil Encode Dengan Menambah Token Khusus	45
Gambar 4. 8 Hasil Tokenizing Padding Dan Attention Mask	46
Gambar 4. 9 Hasil Training & Validation Loss	48
Gambar 4. 10 Grafik Training & Validation Loss	49
Gambar 4. 11 Grafik Akurasi Pada Data Validasi	50
Gambar 4. 12 Confusion Matrix	52

DAFTAR LAMBANG DAN SINGKATAN



BERT	Bidirectional Encoder Representations from Transformers
NLP	Natural Language Processing
GPU	Graphics Processing Unit
AI	Artificial Intelligence
API	Application Programming Interface
RNN	Recurrent Neural Network
TP	True Positive
TN	True Negative
FN	False Negative
FP	False Postitive
CSV	Comma Separated Values
URL	Uniform Resource Language
EDA	Exploration Data Analysis
SSD	Solid State Drive
SMOTE	Synthetic Minority Over Sampling Technique

DAFTAR ISTILAH



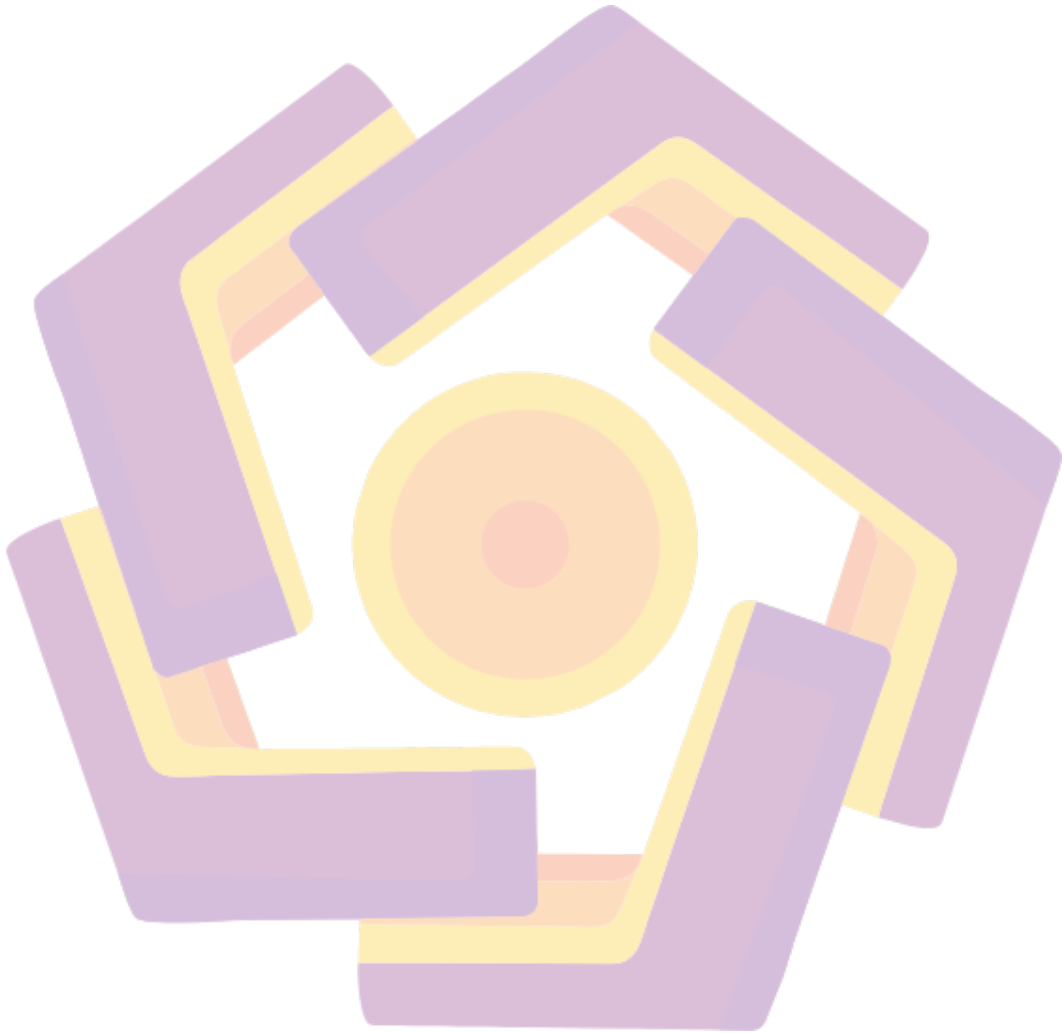
Analisis Sentimen	Proses komputasional untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini atau perasaan (positif, negatif, netral) dari data teks.
BERT	Singkatan dari <i>Bidirectional Encoder Representations from Transformers</i> . Sebuah model bahasa yang dikembangkan oleh Google yang mampu memahami konteks kata secara dua arah (kiri dan kanan).
IndoBERT	Varian model BERT yang telah melalui pra-pelatihan (<i>pre-trained</i>) secara khusus menggunakan korpus besar teks berbahasa Indonesia.
Preprocessing Data	Serangkaian langkah untuk membersihkan dan mempersiapkan data teks mentah sebelum diolah oleh model, meliputi <i>casefolding</i> , <i>text cleaning</i> , dan normalisasi.
Epoch	Satu siklus lengkap di mana seluruh dataset pelatihan telah diproses oleh model sebanyak satu kali.
Batch Size	Jumlah sampel data yang diproses oleh model dalam satu iterasi sebelum parameter model diperbarui.
Recall	Metrik yang mengukur kemampuan model untuk menemukan kembali semua sampel positif yang sebenarnya dari suatu kelas.
F1-Score	Rata-rata harmonik dari Presisi dan <i>Recall</i> , memberikan ukuran kinerja yang seimbang

Overfitting

terutama pada dataset yang tidak seimbang.

Sebuah kondisi di mana model terlalu

"menghafal" data latih sehingga kinerjanya sangat baik pada data latih tetapi buruk pada data baru yang belum pernah dilihat



INTISARI

Media sosial seperti *Youtube* telah menjadi platform penting bagi masyarakat untuk menyuarakan opini mereka, termasuk terhadap institusi pemerintah seperti Kepolisian Negara Republik Indonesia (Polri). Komentar-komentar publik ini merupakan sumber data yang berharga untuk memahami persepsi masyarakat. Penelitian ini bertujuan untuk analisis sentiment terhadap komentar Youtube mengenai institusi Polri dengan menggunakan algoritma *Bidirectional Encoder Representations from Transformers (BERT)*. Metode penelitian ini dimulai dengan pengumpulan data komentar dari sebuah video relevan menggunakan Youtube Data API, diikuti dengan tahap pra-pemrosesan data dan pelabelan sentiment otomatis ke dalam tiga kelas (positif, negative, netral). Hasil evaluasi pada data uji menunjukkan bahwa model *IndoBERT* yang telah diadaptasi memiliki kinerja yang sangat baik, dengan mencapai akurasi keseluruhan sebesar 91.02% dan *weighted F1-Score* 90.71% (presisi 90.63%, Recall 91.02%). Model menunjukkan performa tertinggi dalam mengklasifikasikan sentiment negative (*F1-Score* 95.27%), yang merupakan kelas mayoritas, dan menghadapi tantangan pada kelas netral yang lebih ambigu. Penelitian ini membuktikan efektivitas arsitektur *BERT* untuk menganalisis teks informal berbahasa Indonesia dan menyajikan Gambaran kuantitatif mengenai persepsi public yang didominasi sentiment negative terhadap Polri berdasarkan sampel data yang dianalisis.

Kata kunci: Analisis Sentimen, NLP, IndoBERT, Komentar Youtube, Kepolisian Indonesia.

ABSTRACT

Social media such as Youtube has become an important platform for the public to voice their opinions, including towards government institutions like the Indonesian National Police (Polri). These public comments are a valuable source of data for understanding public perception. This study aims to conduct a sentiment analysis of Youtube comments regarding the Polri institution using the Bidirectional Encoder Representations from Transformers (BERT) algorithm. The research method began with the collection of 7,711 comment data from a relevant video using the Youtube Data API, followed by a data preprocessing stage, and automatic sentiment labeling into three classes (positive, negative, neutral). The evaluation results on the test data show that the adapted IndoBERT model performs very well, achieving an overall accuracy of 91.02% and a weighted F1-Score of 90.71% (precision 90.63%, recall 91.02%). The model shows the highest performance in classifying negative sentiment (F1-score 95.27%), which is the majority class, and faces challenges with the more ambiguous neutral class. This research proves the effectiveness of the BERT architecture for analyzing informal Indonesian language texts and presents a quantitative overview of public perception, which is dominated by negative sentiment towards Polri based on the analyzed data sample.

Keyword: Sentiment Analysis, NLP, IndoBERTm Youtube Comments, Indonesian Police