

**ANALISIS SENTIMEN TERHADAP KERETA CEPAT
WHOOSH MELALUI KOMENTAR TIKTOK
MENGUNAKAN INDOBERT**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

M.TAUFIK PRATAMA

22.11.5169

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN TERHADAP KERETA CEPAT
WHOOSH MELALUI KOMENTAR TIKTOK
MENGUNAKAN INDOBERT
SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh
M.TAUFIK PRATAMA
22.11.5169

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP KERETA CEPAT WHOOSH
MELALUI KOMENTAR TIKTOK MENGGUNAKAN INDOBERT

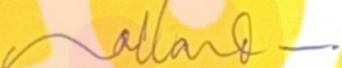
yang disusun dan diajukan oleh

M.Taufik Pratama

22.11.5169

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Mei 2026

Dosen Pembimbing,


Akhmad Dahlan, S.Kom., M.Kom.
NIK. 190302174

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP KERETA CEPAT WHOOSH
MELALUI KOMENTAR TIKTOK MENGGUNAKAN INDOBERT

yang disusun dan diajukan oleh

M.Taufik Pratama

22.11.5169

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232



Ike Verawati, S.Kom., M.Kom.
NIK. 190302237



Akhsad Dahlan, S.Kom., M.Kom.
NIK. 190302174



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : M. Taufik Pratama
NIM : 22.11.5169

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS SENTIMEN TERHADAP KERETA CEPAT WHOOSH MELALUI KOMENTAR TIKTOK MENGGUNAKAN INDOBERT

Dosen Pembimbing : Akhmad Dahlan, S.Kom., M.Kom.

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, 19 Mei 2026

Yang Menvatakan



(M. Taufik Pratama)

HALAMAN PERSEMBAHAN

Segala puji bagi Allah, Tuhan semesta alam, yang telah melimpahkan rahmat, karunia, dan petunjuk-Nya sehingga penulis dapat menyelesaikan tugas skripsi ini dengan sebaik-baiknya.

Skripsi ini penulis persembahkan kepada:

1. Kedua orang tua tercinta yang senantiasa memberikan doa, dukungan, kasih sayang, pengorbanan, serta motivasi yang tiada henti kepada penulis dalam menempuh pendidikan.
2. Keluarga dan saudara yang selalu memberikan semangat, perhatian, dan dukungan selama proses penyusunan skripsi ini.
3. Bapak Akhmad Dahlan, M.Kom., dan seluruh dosen di Program Studi Informatika Universitas AMIKOM Yogyakarta yang telah memberikan ilmu, arahan, serta bimbingan kepada penulis selama masa perkuliahan hingga penyelesaian skripsi.
4. Sahabat dan rekan seperjuangan yang telah memberikan bantuan, dukungan, dan motivasi selama proses penyusunan skripsi ini.
5. Semua pihak yang telah membantu dan memberikan kontribusi, baik secara langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Semoga skripsi ini dapat memberikan manfaat dan menjadi kontribusi yang positif bagi perkembangan ilmu pengetahuan.

KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala berkat, rahmat, hidayah dan limpahan karunia-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “ANALISIS SENTIMEN TERHADAP KERETA CEPAT WHOOSH MELALUI KOMENTAR TIKTOK MENGGUNAKAN INDOBERT”. Skripsi ini ditulis untuk memenuhi persyaratan untuk mencapai gelar Sarjana Komputer dalam Program Studi Informatika Fakultas Ilmu Komputer.

Skripsi ini dapat diselesaikan atas doa, dukungan bantuan serta bimbingan dari berbagai pihak, baik bantuan berupa moril maupun materil. Dalam kesempatan ini dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Kedua Orang Tua serta seluruh keluarga yang tidak pernah berhenti mendoakan dan memberikan dukungannya selama ini.
2. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
3. Ibu Prof. Dr. Kusri., M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
4. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
5. Bapak Akhmad Dahlan, M.Kom., selaku dosen pembimbing yang telah memberikan waktu, saran, kritik, motivasi dan bimbingan dalam skripsi ini sehingga skripsi ini dapat selesai dengan baik.
6. Seluruh Staff Pengajar Khususnya Dosen yang mengajar di Program Studi Informatika, Universitas Amikom Yogyakarta yang telah memberikan bantuan, masukan, dan informasi serta ilmunya kepada penulis selama mengikuti perkuliahan.
7. Teman-teman penulis di FIK Universitas Amikom Yogyakarta angkatan 2022 yang tidak dapat disebutkan namanya satu persatu. Terima kasih atas pertemanan selama ini.

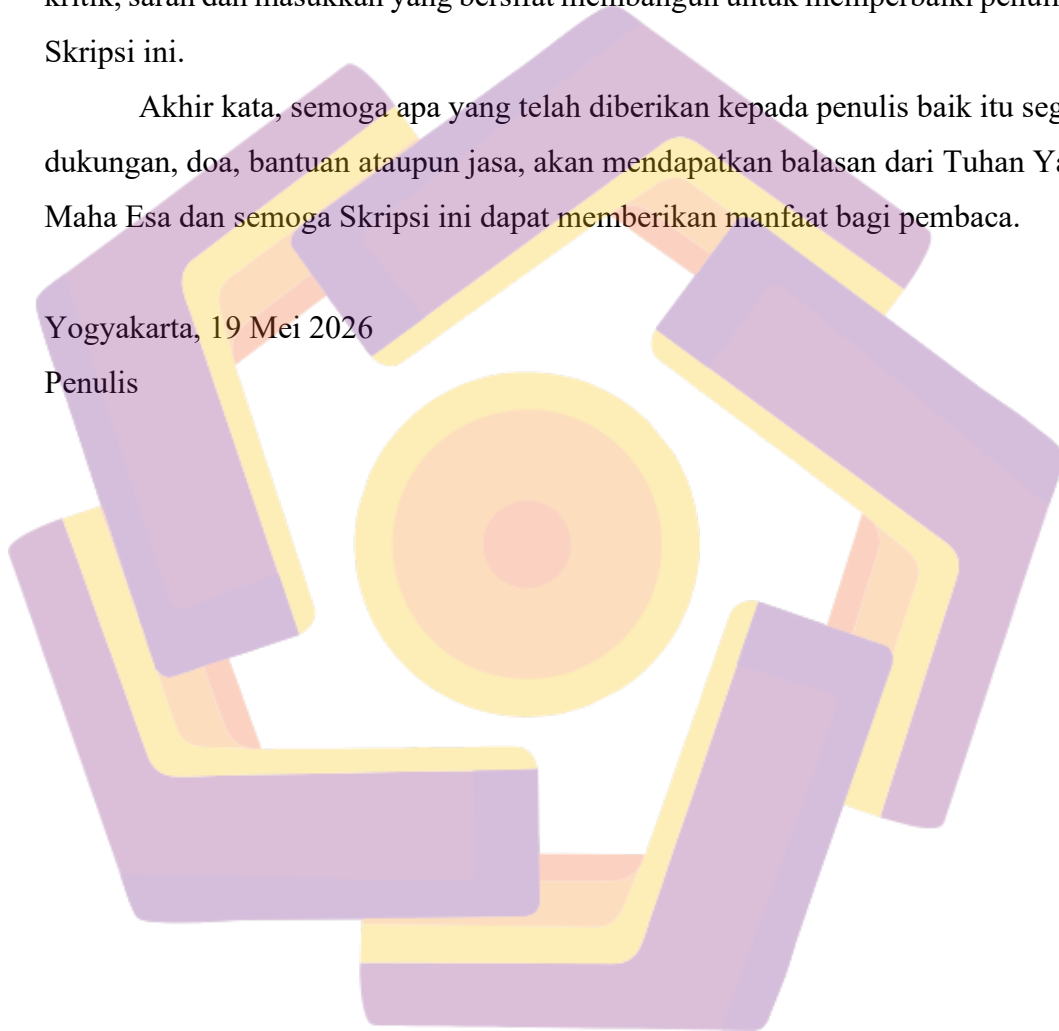
8. Gank kontrakan MGM yang telah memberikan dukungan, masukan, dan keceriaan.
9. Terakhir, diri sendiri yang sudah bertahan selama ini.

Penulis menyadari bahwa keterbatasan pengetahuan dan pengalaman yang masih jauh dari harapan dan kesempurnaan, untuk itu penulis sangat mengharapkan kritik, saran dan masukan yang bersifat membangun untuk memperbaiki penulisan Skripsi ini.

Akhir kata, semoga apa yang telah diberikan kepada penulis baik itu segala dukungan, doa, bantuan ataupun jasa, akan mendapatkan balasan dari Tuhan Yang Maha Esa dan semoga Skripsi ini dapat memberikan manfaat bagi pembaca.

Yogyakarta, 19 Mei 2026

Penulis

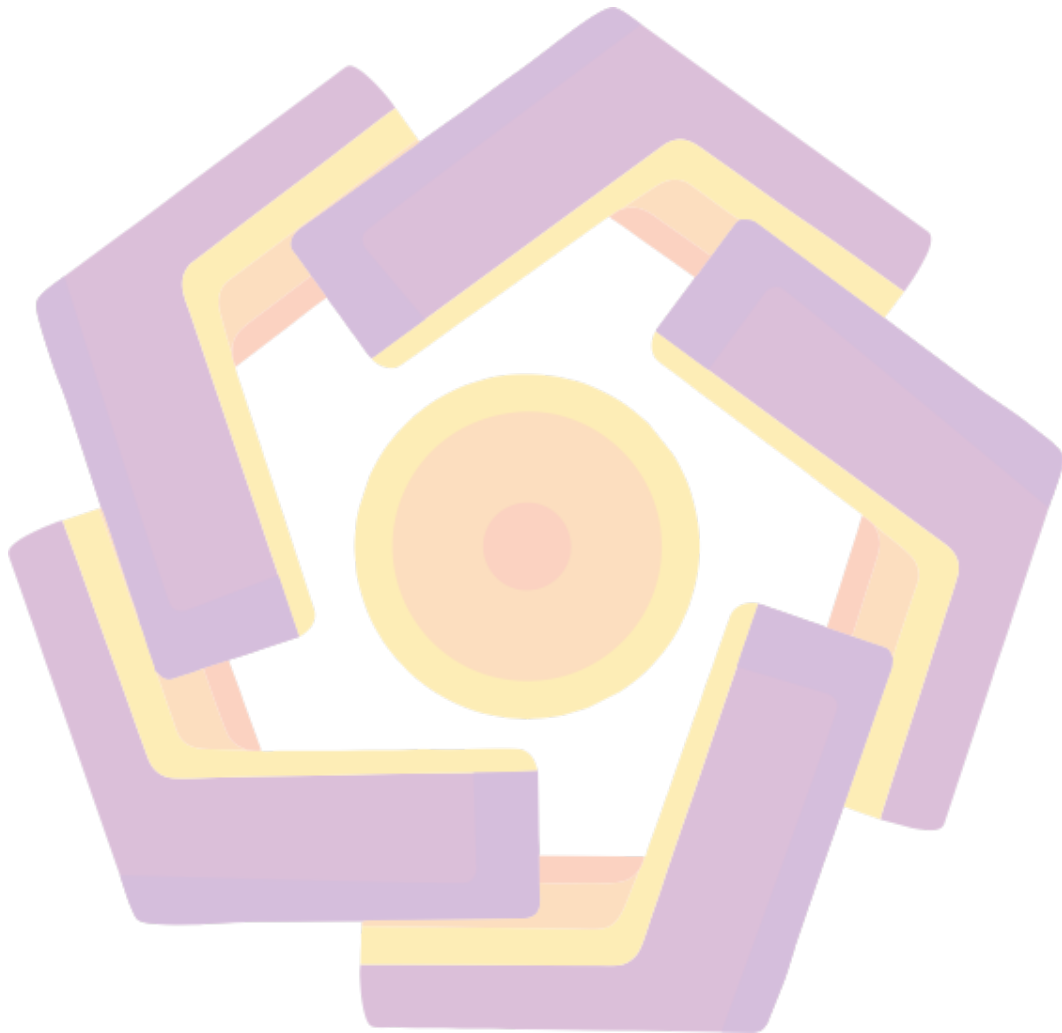


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvii
ABSTRACT	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	8
2.1 Studi Literatur	8
2.2 Dasar Teori.....	16
2.2.1 Whoosh	16
2.2.2 TikTok.....	16
2.2.3 Analisis Sentimen	17
2.2.4 Bahasa Pemrosesan Alami (Natural Language Processing)	17
2.2.5 Preprocessing	18

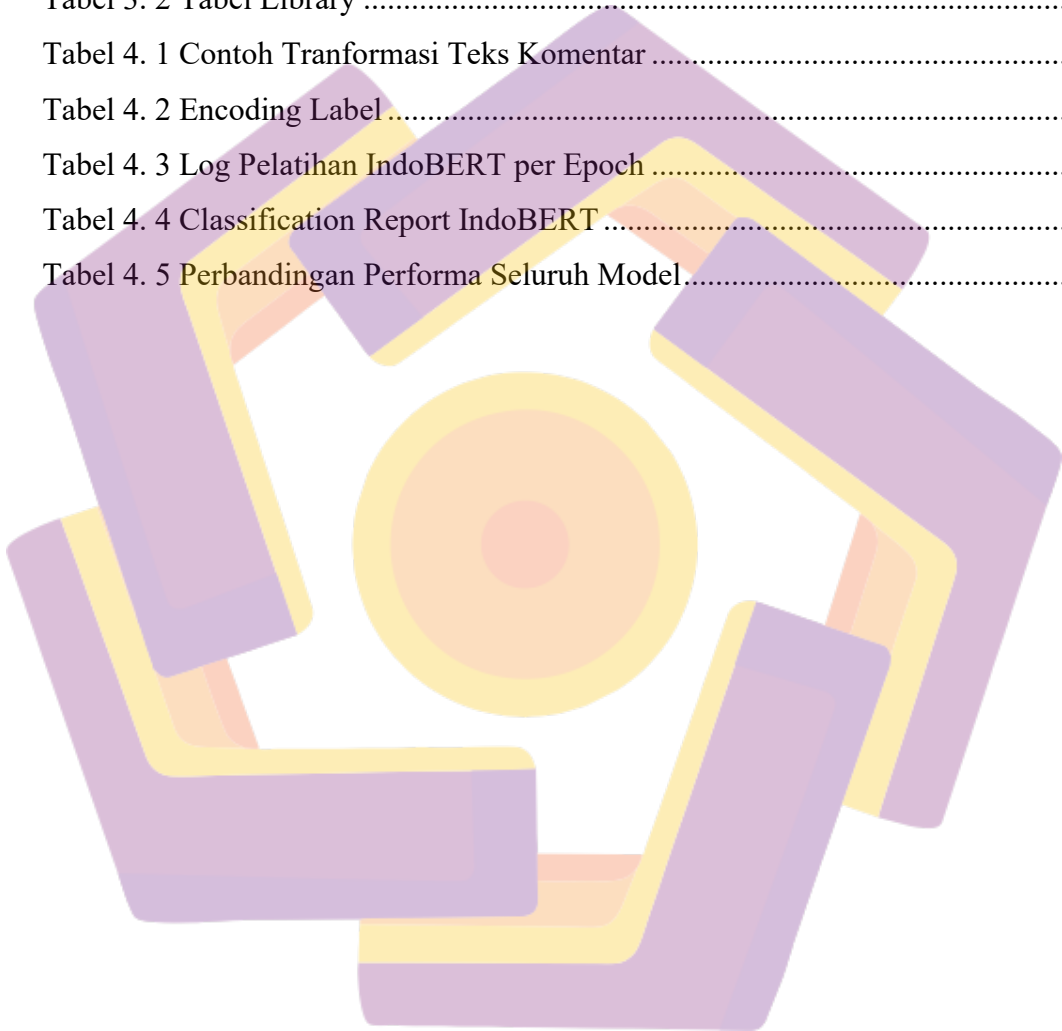
2.2.6	IndoBERT	20
2.2.7	Fine-tuning	21
2.2.8	Confusion matrix	21
2.2.9	Lexicon-Based Sentiment Analysis	23
2.2.10	Support Vector Machine (SVM).....	24
2.2.11	Long Short-Term Memory (LSTM)	25
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian.....	26
3.2	Alur Penelitian	26
3.2.1	Tahap Pendahuluan.....	27
3.2.2	Tahap Pengolahan Data	28
3.3	Alat dan Bahan.....	29
3.3.1	Alat	29
3.3.2	Bahan	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1.	Pengolahan Data	33
4.1.1.	Pengumpulan Data	33
4.1.2.	Pelabelan Data	34
4.1.3.	Pemrosesan Data Awal	35
4.2.	Tahap Persiapan Data.....	39
4.2.1.	Encoding Label	39
4.2.2.	Pembagian Dataset (Train-Test Split).....	40
4.3.	Pembangunan dan Pelatihan Model.....	41
4.5.1.	Model Utama IndoBERT	41
4.4.	Metode Pembandingan	44
4.4.1.	Metode Lexicon-Based	44
4.4.2.	Metode Support Vector Machine (SVM).....	45
4.4.3.	Metode Long Short-Term Memory (LSTM)	45
4.5.	Evaluasi Model	46
4.5.1.	Hasil Evaluasi IndoBERT (Best Checkpoint).....	46
4.5.2.	Perbandingan Seluruh Model.....	47
BAB V PENUTUP		49

5.1	Kesimpulan	49
5.2	Saran	49
REFERENSI.....		51
LAMPIRAN.....		53



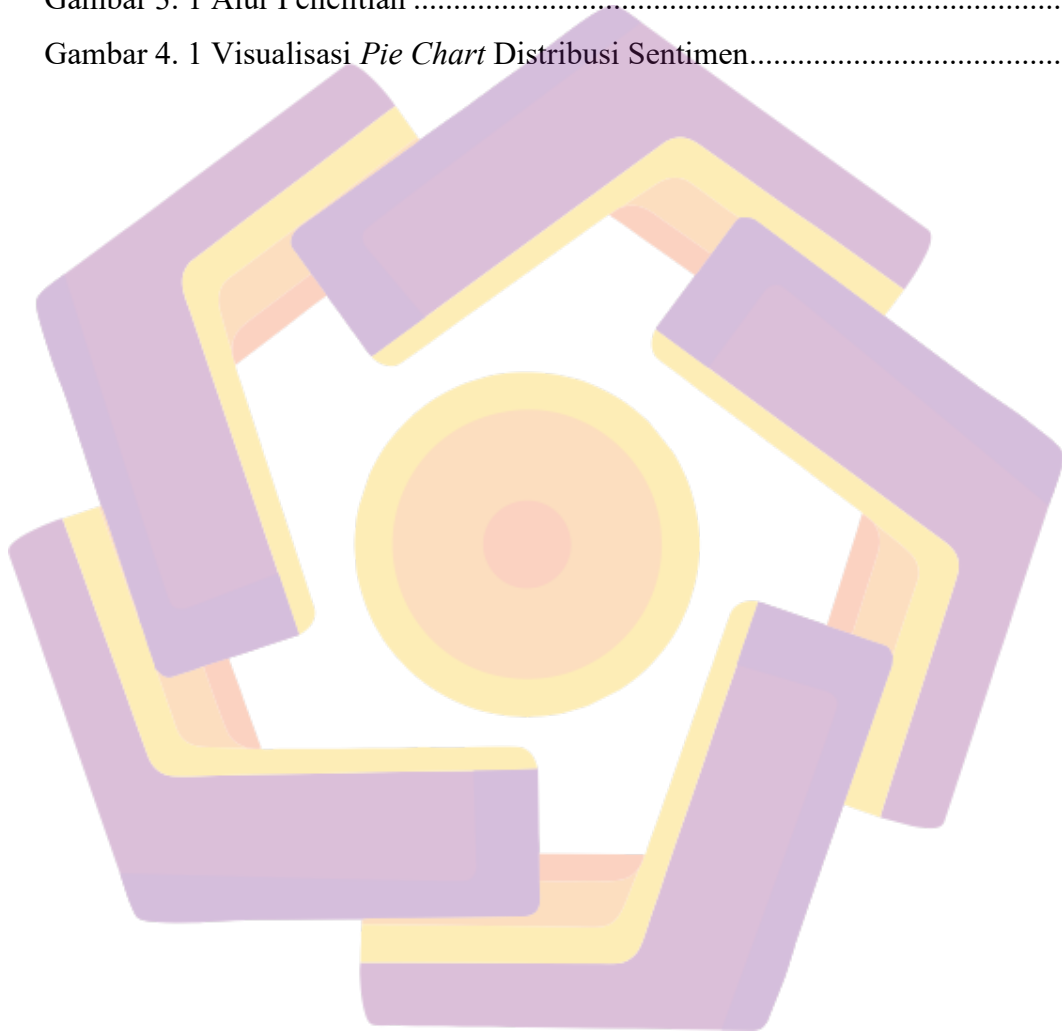
DAFTAR TABEL

Tabel 2. 1 Perbandingan Studi Literatur	12
Tabel 2. 2 Tabel perbandingan Bert.....	21
Tabel 3. 1 Komponen dan Spesifikasi Komputer	29
Tabel 3. 2 Tabel Library	30
Tabel 4. 1 Contoh Tranformasi Teks Komentar	39
Tabel 4. 2 Encoding Label	40
Tabel 4. 3 Log Pelatihan IndoBERT per Epoch	43
Tabel 4. 4 Classification Report IndoBERT	46
Tabel 4. 5 Perbandingan Performa Seluruh Model.....	47



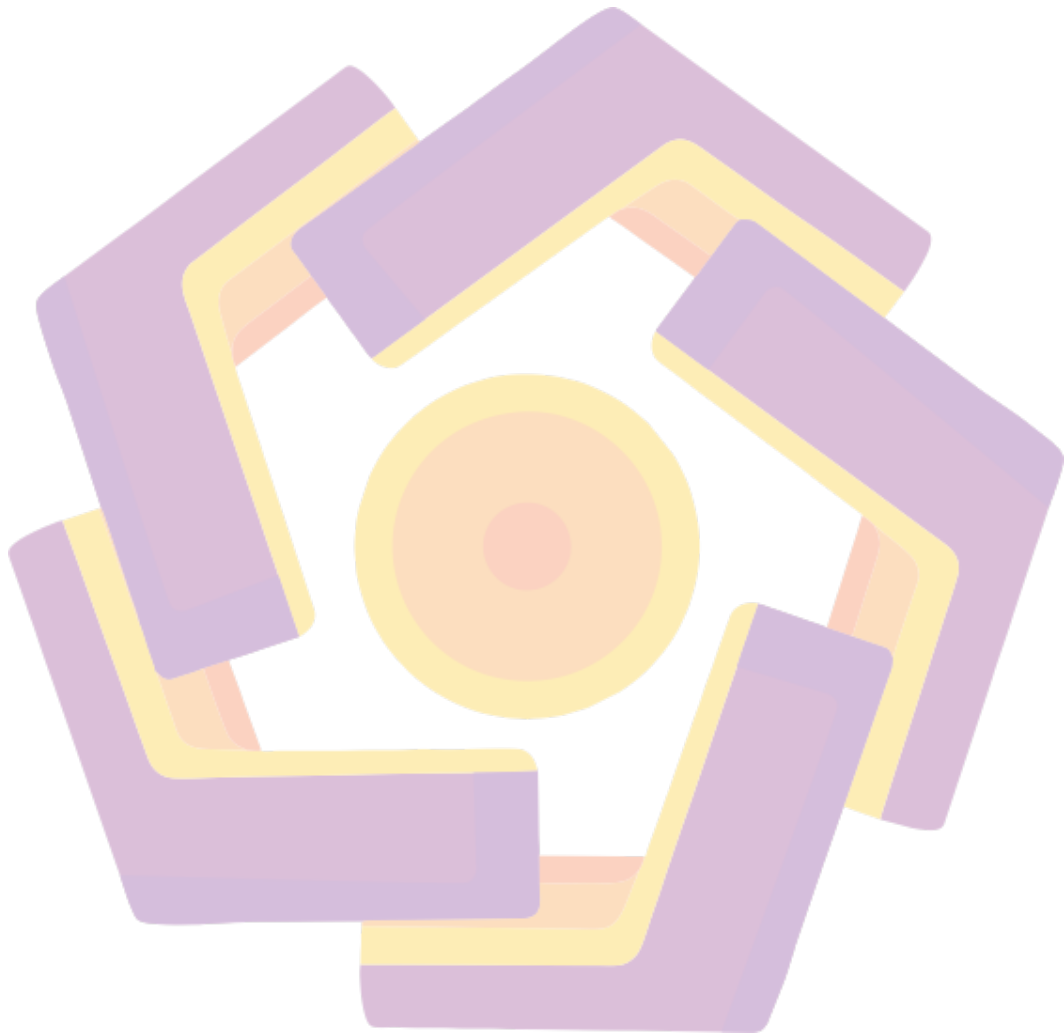
DAFTAR GAMBAR

Gambar 1. 1 Grafik pertumbuhan pengguna TikTok di Indonesia	1
Gambar 2. 1 Grafik pengguna TikTok Indonesia berdasarkan usia	17
Gambar 2. 2 Confussion Matrix.....	22
Gambar 3. 1 Alur Penelitian	27
Gambar 4. 1 Visualisasi <i>Pie Chart</i> Distribusi Sentimen.....	36

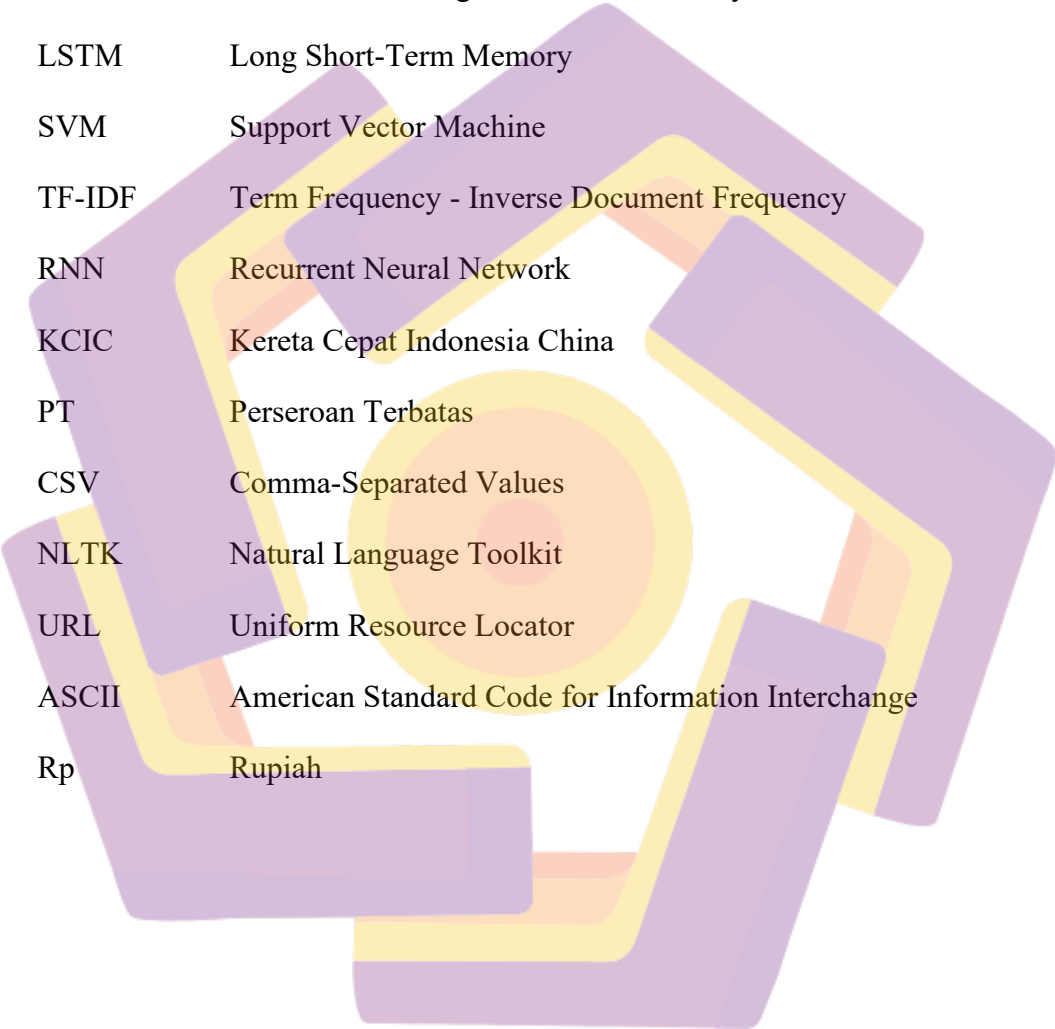


DAFTAR LAMPIRAN

Lampiran 1. Source code:	53
Lampiran 2. Prototype	53
Lampiran 3. Dataset:	53



DAFTAR LAMBANG DAN SINGKATAN



BERT	Bidirectional Encoder Representations from Transformers
NLP	Natural Language Processing
BILSTM	Bidirectional Long Short-Term Memory
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverse Document Frequency
RNN	Recurrent Neural Network
KCIC	Kereta Cepat Indonesia China
PT	Perseroan Terbatas
CSV	Comma-Separated Values
NLTK	Natural Language Toolkit
URL	Uniform Resource Locator
ASCII	American Standard Code for Information Interchange
Rp	Rupiah

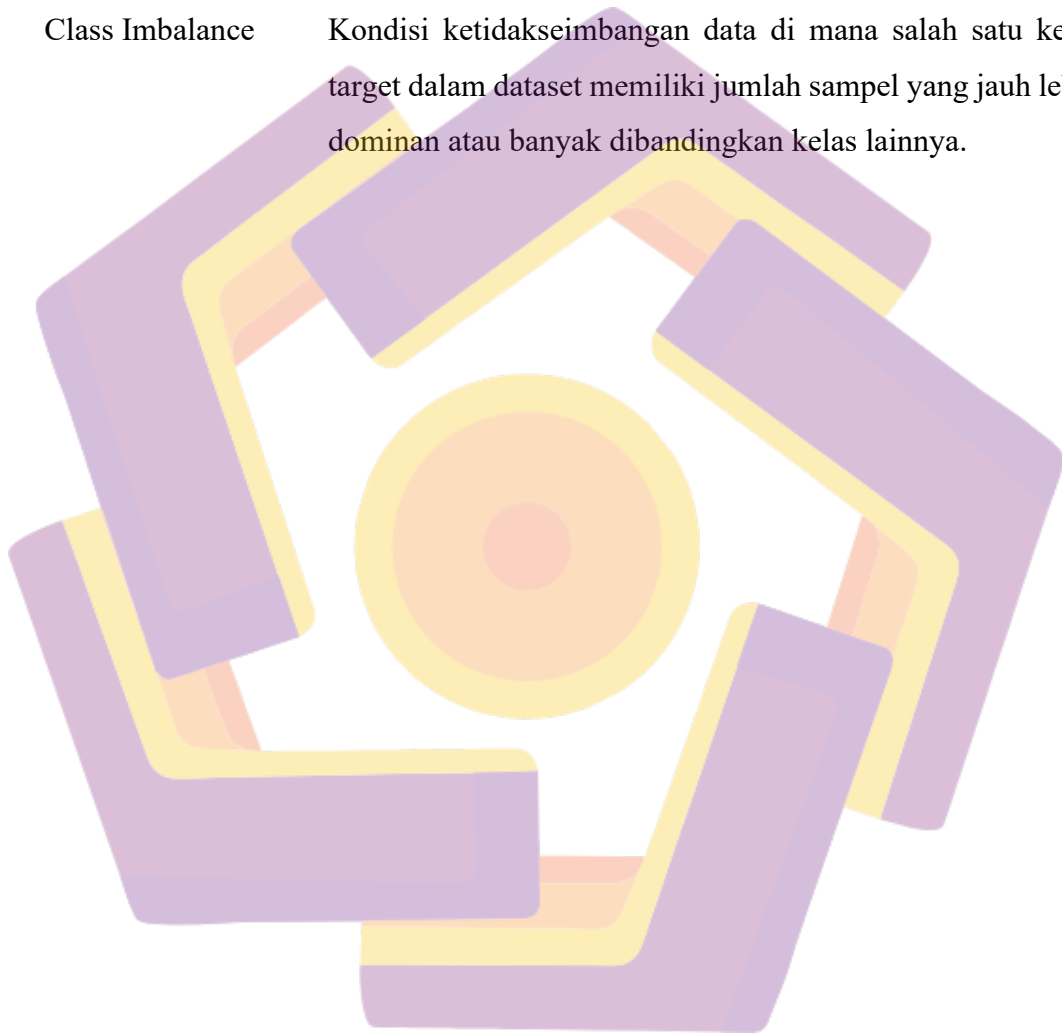
DAFTAR ISTILAH

Heterogenitas	Perbedaan atau keragaman jenis atau susunan elemen atau konstituen penyusunnya.
Fine-tuning	Pendekatan atau proses melatih kembali model yang telah dilatih sebelumnya (<i>pre-trained model</i>) dengan menambahkan atau menggunakan dataset spesifik yang baru untuk tugas tertentu.
Confusion Matrix	Matriks atau tabel evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan nilai aktual (<i>ground truth</i>) dengan nilai hasil prediksi.
Lexicon-Based	Pendekatan analisis sentimen yang proses klasifikasi atau pelabelannya berbasis pada kamus kata (<i>leksikon</i>) yang telah memiliki sentimen tertentu.
Scraping/crawling	Proses pengambilan atau pengumpulan data teks atau komentar secara otomatis dari platform digital atau situs web tertentu.
Deep Learning	Sub bidang dari pembelajaran mesin (<i>machine learning</i>) berbasis jaringan saraf tiruan dengan banyak lapisan yang mampu pola dari kata kompleks.
Tokenizing	Proses memecah teks atau dokumen menjadi potongan-potongan kecil yang disebut token.
Stemming	Proses mengubah kata berimbuhan menjadi bentuk dasarnya untuk mengurangi variasi morfologis kata.
Stopword Removal	Tahapan untuk menghapus kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna yang signifikan dalam analisis sentimen.
Case Folding	Proses mengubah seluruh huruf di dalam teks menjadi huruf

kecil (lowercase) untuk meminimalkan redundansi kosakata akibat perbedaan kapitalisasi.

Dataset Kumpulan data terstruktur yang dikumpulkan dan digunakan sebagai objek utama dalam proses pelatihan serta pengujian model penelitian.

Class Imbalance Kondisi ketidakseimbangan data di mana salah satu kelas target dalam dataset memiliki jumlah sampel yang jauh lebih dominan atau banyak dibandingkan kelas lainnya.



INTISARI

Pesatnya perkembangan media sosial TikTok telah menjadikannya platform krusial bagi masyarakat untuk menyampaikan opini secara real-time terhadap berbagai isu, termasuk operasional kereta cepat Whoosh. Namun, tingginya intensitas komentar yang bersifat informal dan ambigu menimbulkan tantangan dalam memahami kecenderungan opini publik secara sistematis. Dampak dari masalah ini adalah sulitnya pihak pengelola maupun pemerintah dalam memetakan persepsi publik yang sebenarnya guna melakukan evaluasi layanan. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna TikTok mengenai Whoosh menggunakan model IndoBERT dengan pendekatan fine-tuning. Data sebanyak 1.337 komentar dikumpulkan menggunakan TikTok Comments Scraper dan melalui tahap preprocessing yang meliputi pembersihan teks, penghapusan emoji, serta pelabelan manual ke dalam kategori positif, negatif, dan netral. Model IndoBERT kemudian dibandingkan kinerjanya dengan tiga metode lain, yaitu Lexicon-Based Sentiment Analysis, Support Vector Machine (SVM), dan Bidirectional Long Short-Term Memory (BiLSTM). Hasil penelitian menunjukkan dominasi sentimen negatif yang ekstrem sebesar 87,3%, yang terutama dipicu oleh keluhan mengenai harga tiket yang dianggap tidak terjangkau. Secara teknis, evaluasi menggunakan metrik F1 Macro mengungkapkan bahwa BiLSTM mencapai performa tertinggi (0,4277), diikuti oleh SVM (0,4096), dan IndoBERT (0,3395). Hal ini membuktikan bahwa pada dataset kecil dan tidak seimbang, model dengan kompleksitas lebih rendah mampu mengungguli model Transformer. Hasil penelitian ini dapat dimanfaatkan oleh manajemen PT KCIC sebagai bahan evaluasi kebijakan harga dan layanan.

Kata kunci: Analisis Sentimen, Komentar TikTok, BERT, Deep Learning, Opini Publik.

ABSTRACT

Rapid development of TikTok as a social media platform has made it a crucial venue for public discourse, where users share real-time opinions on various issues, including the "Whoosh" high-speed railway operations. However, the high intensity of informal and ambiguous comments creates challenges in systematically understanding public opinion trends. The impact of this issue is the difficulty for operators and policymakers to accurately map public perception for service evaluation. This study aims to analyze the sentiment of TikTok user comments regarding Whoosh using the IndoBERT model with a fine-tuning approach. A total of 1.337 comments were collected via TikTok Comments Scraper and underwent preprocessing, including text cleaning, emoji removal, and manual labeling into positive, negative, and neutral categories. The IndoBERT model's performance was compared against three other methods: Lexicon-Based Sentiment Analysis, Support Vector Machine (SVM), and Bidirectional Long Short-Term Memory (BiLSTM). The findings reveal an extreme dominance of negative sentiment at 87.3%, primarily driven by concerns over ticket pricing perceived as unaffordable. Technical evaluation using the F1 Macro metric showed that BiLSTM achieved the highest performance (0.4277), followed by SVM (0.4096), and IndoBERT (0.3395). This proves that on small and imbalanced datasets, lower-complexity models can outperform Transformer models. These results can be utilized by PT KCIC management as a basis for evaluating pricing policies and services.

Keyword: Sentiment Analysis, TikTok Comments, BERT, Deep Learning, Public Opinion.