

**ANALISIS SENTIMEN OPINI PENGGUNA TWITTER
TERHADAP APLIKASI KAI ACCESS DENGAN METODE
BERT**

SKRIPSI

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



disusun oleh

SELI AGALISTA

20.11.3525

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN OPINI PENGGUNA TWITTER
TERHADAP APLIKASI KAI ACCESS DENGAN METODE
BERT**

SKRIPSI

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



disusun oleh
SELI AGALISTA
20.11.3525

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN OPINI PENGGUNA TWITTER
TERHADAP APLIKASI KAI ACCESS DENGAN METODE
BERT**

yang disusun dan diajukan oleh
Seli Agalista
20.11.3525

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Februari 2025

Dosen Pembimbing,



Windha Mega Pradnya Dhuhiita, M.Kom.
NIK. 190302185

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN OPINI PENGGUNA TWITTER TERHADAP
APLIKASI KAI ACCESS DENGAN METODE BERT

yang disusun dan diajukan oleh

Seli Agalista

20.11.3525

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Theopilus Bayu Sasongko , S.Kom., M.Eng
NIK. 190302375

Ahlihi Masruro, S.Kom., M.Kom
NIK. 190302148

Windha Mega Pradnya Dhuhita, S.Kom., M.Kom.
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
12 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Seli Agalista
NIM : 20.11.3525

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Opini Pengguna Twitter Terhadap Aplikasi KAI Access Dengan Metode BERT

Dosen Pembimbing : Windha Mega Pradnya Dhuhita, 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, 12 Februari 2025

Yang Menyatakan,



Seli Agalista

HALAMAN PERSEMBAHAN

Pertama saya ucapkan puji syukur kepada Allah SWT atas segala nikmat dan kasih sayangnya. Skripsi ini penulis persembahkan untuk:

1. Bapak dan Ibu yang sangat saya cintai, yang selalu memberikan dukungan tanpa henti, cinta, pengorbanan dan doa yang tak terhingga untuk penulis sehingga penulis dapat sampai dititik ini. Terima kasih telah menajdi orang tua yang sangat hebat.
2. Risa Putri selaku saudara penulis, terima kasih atas doa dan segala bentuk dukungan yang diberikan untuk penulis sehingga penulis dapat mengerjakan skripsi dengan tenang.
3. Ibu Windha Mega Pradnya Dhuhita, S.Kom., M.Kom selaku dosen pembimbing saya yang telah membimbing saya dalam menyelesaikan skripsi ini. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Ibu berikan. Semoga segala kebaikan Ibu mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.
4. Untuk Ghaluh, Raniya, Ardi, Hapis, dan Rangga sahabat di bangku kuliah saya, terima kasih atas bantuan, pengertian, dukungan serta kesediaan kalian mendengarkan setiap keluh kesah hidup penulis kapan pun dibutuhkan.

KATA PENGANTAR

Segala puji bagi Allah SWT atas limpahan rahmat, petunjuk dan karunia-Nya yang telah membantu peneliti menyelesaikan penyusunan skripsi ini. Skripsi ini disusun sebagai tahapan penting untuk memenuhi persyaratan kelulusan program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penyusunan skripsi ini tidak lepas dari doa, bantuan, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Keluarga tersayang, yang senantiasa memberikan doa terbaik, dukungan moral, materi, serta semangat dalam perjalanan penyusunan skripsi ini. Tanpa kasih sayang dan motivasi yang tiada henti, pencapaian ini tidak akan terwujud.
2. Windha Mega Pradnya Duhita, S.Kom., M.Kom selaku dosen pembimbing, atas dukungan, arahan, serta masukan berharga yang telah membimbing langkah-langkah penyusunan skripsi ini.
3. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.
4. Semua individu yang turut menyumbangkan gagasan, pandangan, dan sokongan dalam berbagai bentuk, yang telah membantu kelancaran penyelesaian skripsi ini. Setiap bantuan yang diberikan, sekecil apa pun, memiliki makna besar bagi penulis.

Yogyakarta, 12 Februari 2025

Penulis

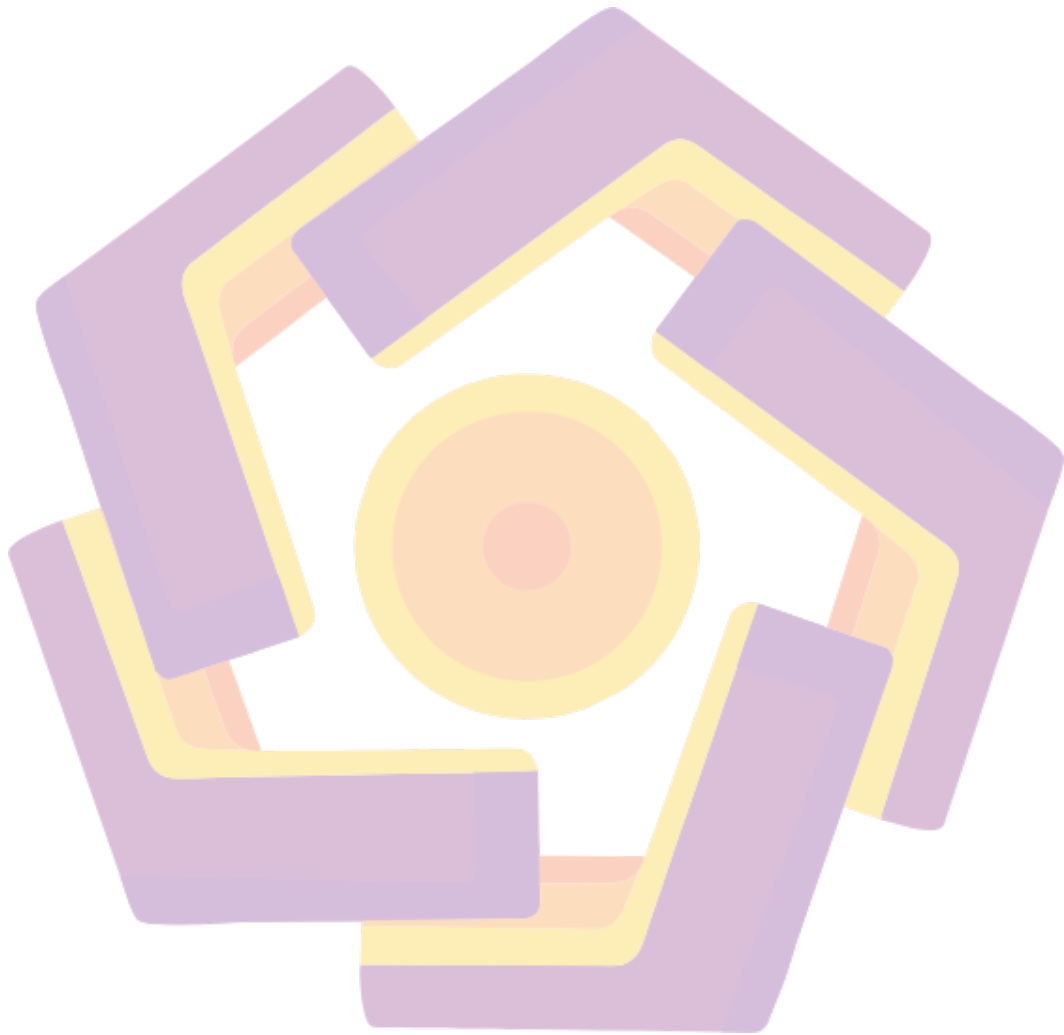
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT.....	xvii
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 II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	13
2.2.1 KAI ACCESS.....	13
2.2.2 Analisis sentiment	13

2.2.3	X/Twitter	14
2.2.4	Crawling Data	14
2.2.5	Pre-processing	15
2.2.5.1	Case folding	15
2.2.5.2	Data cleaning	15
2.2.5.3	Normalisasi	16
2.2.6	Labeling.....	16
2.2.7	BERT Multilingual.....	16
2.2.7.1	Tokenizing	18
2.2.7.2	Encode.....	18
2.2.7.3	Confusion Matrix	18
BAB III METODE PENELITIAN		20
3.1	Objek Penelitian	20
3.2	Alur Penelitian.....	20
3.3.1	Pengumpulan Data	21
3.3.2	Data Tweet	21
3.3.3	Data Normalisasi	21
3.3.4	Pre-processing	22
3.3.4.1	Case Folding	22
3.3.4.2	Cleaning	23
3.3.4.3	Normalisasi	23
3.3.5	Labeling.....	23
3.3.6	BERT Multilingual.....	24
3.3.6.1	Tokenizing	24
3.3.6.2	Encoding	25

3.3.6.3	Evaluasi BERT Multilingual.....	25
3.3.6.4	Evaluasi Set Validasi	25
3.3.6.5	Evaluasi Set Pengujian.....	26
BAB IV HASIL DAN PEMBAHASAN		27
4.1	Pengumpulan Data	27
4.4.1	Data Tweet	27
4.4.2	Data Normalisasi.....	29
4.2	Pre-Processing	30
4.4.1	Case Folding	30
4.4.2	Cleaning	32
4.4.3	Normalisasi	33
4.3	Labelling.....	38
4.4	BERT Multilingual.....	41
4.4.1	Tokenizing	41
4.4.2	Encode.....	43
4.5	Splitting Data.....	45
4.5.1	Training dan Temporary Set	46
4.5.2	Validation dan Testing Set	46
4.6	Implementasi BERT Multilingual.....	47
4.7	Evaluasi BERT Multilingual.....	49
4.7.1	Evaluasi Set Validasi	52
4.7.2	Evaluasi Set Pengujian.....	53
BAB V PENUTUP		56
5.1	Kesimpulan.....	56
5.2	Saran.....	56

REFERENSI	58
-----------------	----



DAFTAR TABEL

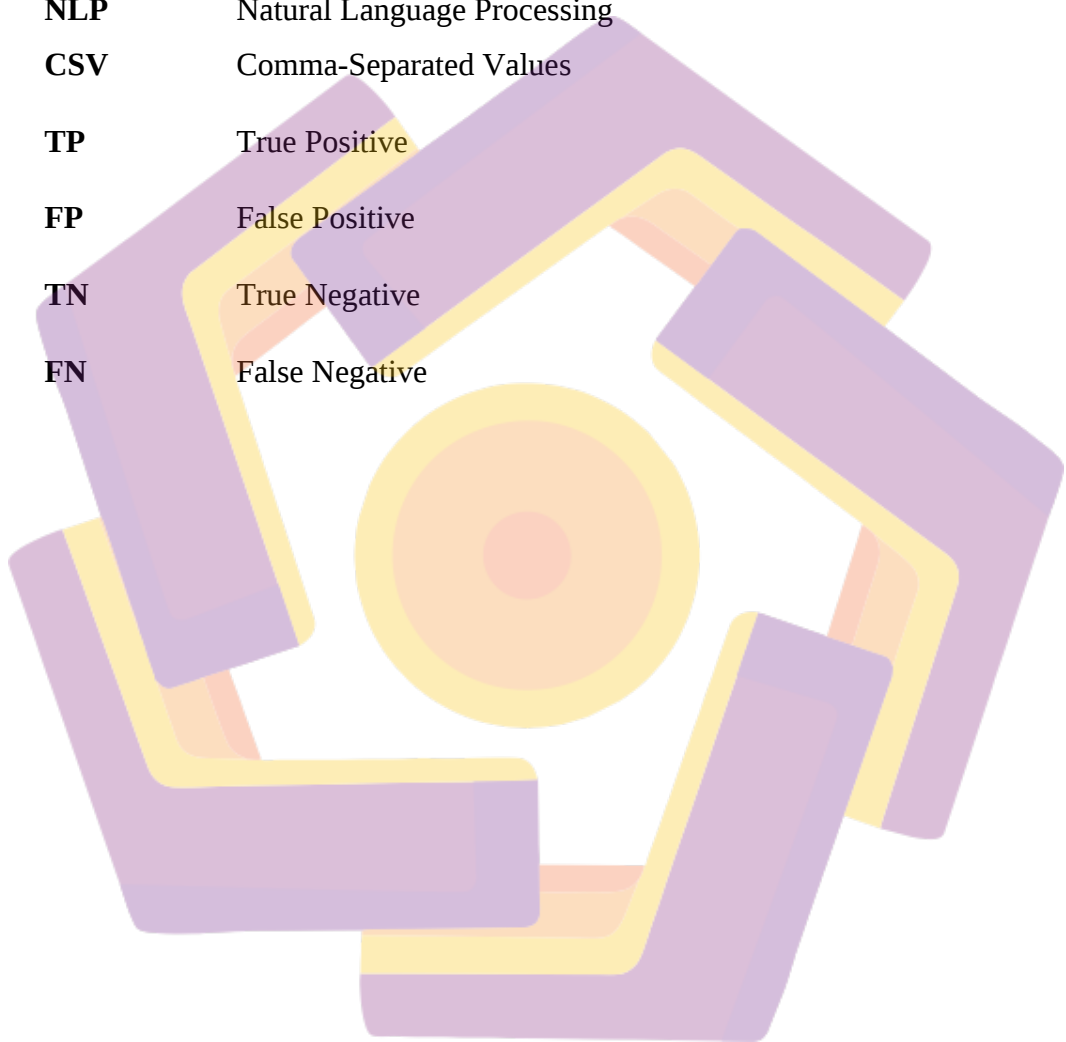
Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 4. 1 Hasil Penggabungan Dataset.....	28
Tabel 4. 2 Dataset Normalisasi	30
Tabel 4. 3 Sampel Kalimat Slang Tambahan.....	30
Tabel 4. 4 Hasil Penerapan Case Folding	31
Tabel 4. 5 Hasil Penerapan Cleaning	32
Tabel 4. 6 Hasil Normalisasi	35
Tabel 4. 7 Hasil Identifikasi Bahasa Komentar	37
Tabel 4. 8 Proses Tokenizing.....	42
Tabel 4. 9 Hasil Encoding, Padding, dan Attention Mark	45
Tabel 4. 10 Hasil Pembagian Data.....	46
Tabel 4. 11 Hyperparameter BERT	49
Tabel 4. 12 Tabel Rata-rata Training Loss	50
Tabel 4. 13 Performansi Model Set Validation	52
Tabel 4. 14 Performansi Model Set Test.....	53

DAFTAR GAMBAR

Gambar 2. 1 Visualisasi Kerja BERT	17
Gambar 4. 1 Informasi Dataset Normalisasi	33
Gambar 4. 2 Informasi Dataset Normalisasi Setelah Hapus Duplikat	34
Gambar 4. 3 Code Normalisasi Konversi Teks Slang	34
Gambar 4. 4 Pustaka Deteksi Bahasa	36
Gambar 4. 5 Visualisasi Label Positive dan Negative	40
Gambar 4. 6 Grafik Training Loss.....	50
Gambar 4. 7 Diagram Confusion Matrix Set Validation	53
Gambar 4. 8 Diagram Confusion Matrix Set Test.....	54

DAFTAR LAMBANG DAN SINGKATAN

BERT	Bidirectional Encoder Representations from Transformers
KAI	Kereta Api Indonesia
X	Twitter
NLP	Natural Language Processing
CSV	Comma-Separated Values
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative



DAFTAR ISTILAH

Slang	Kata gaul
Case Folding (lowercase)	Proses mengubah semua data teks menjadi huruf kecil
Normalisasi	Tahap mengubah kata-kata tidak baku menjadi baku
Tokenizing	Memisahkan kata menjadi kalimat terpisah
Encode angka	Mengubah kata dalam sebuah kalimat menjadi representasi



INTISARI

Di era digital, media sosial menjadi platform utama masyarakat untuk menyampaikan pendapat, termasuk mengenai aplikasi seperti KAI Access. Aplikasi ini merupakan layanan resmi PT Kereta Api Indonesia untuk mempermudah pengguna dalam membeli tiket dan mendapatkan informasi terkait perjalanan kereta. Penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat terhadap aplikasi KAI Access yang disampaikan melalui platform X (Twitter). Model BERT (*Bidirectional Encoder Representations from Transformers*) digunakan untuk klasifikasi sentimen karena kemampuannya memahami konteks bahasa secara dua arah.

Penelitian dilakukan dengan mengumpulkan 12.477 tweet selama November 2023 hingga April 2024 menggunakan teknik *crawling data*. Setelah melalui proses *pre-processing*, termasuk *case folding*, *cleaning*, normalisasi, dan tokenisasi, data dibagi menjadi 8.511 data berbahasa Indonesia dan dianalisis menggunakan model Multilingual BERT yang dioptimasi. Model ini dilatih dengan *hyperparameter learning rate* $2e-5$, *batch size* 16, dan 5 *epoch*, serta dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Hasil penelitian menunjukkan akurasi model mencapai 90% pada set validasi dan pengujian. Analisis sentimen mengungkapkan bahwa sebagian besar opini pengguna terhadap aplikasi KAI Access bersifat positif (51,6%) dibandingkan negatif (48,4%). Studi ini membuktikan efektivitas Multilingual BERT dalam memahami dan mengklasifikasikan sentimen berbahasa Indonesia, termasuk variasi informal dan slang. Temuan ini memberikan wawasan berharga bagi pengembang untuk meningkatkan kualitas layanan aplikasi KAI Access.

Kata kunci: Analisis sentimen, KAI Access, Twitter, BERT, Multilingual BERT

ABSTRACT

In the digital era, social media serves as a primary platform for expressing opinions, including those about applications such as KAI Access. This application, developed by PT Kereta Api Indonesia, facilitates users in purchasing tickets and accessing travel-related information. This study aims to analyze public sentiment toward the KAI Access application as expressed through the X (Twitter) platform. The BERT (Bidirectional Encoder Representations from Transformers) model was employed for sentiment classification due to its ability to comprehend language context bidirectionally.

The research involved collecting 12,477 tweets from November 2023 to April 2024 using data crawling techniques. After pre-processing, which included case folding, cleaning, normalization, and tokenization, 8,511 Indonesian-language tweets were analyzed using the optimized Multilingual BERT model. The model was trained using hyperparameters of a $2e-5$ learning rate, a batch size of 16, and 5 epochs, and evaluated with metrics such as accuracy, precision, recall, and F1-score.

The results showed that the model achieved 90% accuracy on validation and test sets. Sentiment analysis revealed that the majority of user opinions about the KAI Access application were positive (51.6%) compared to negative (48.4%). This study demonstrates the effectiveness of Multilingual BERT in understanding and classifying Indonesian-language sentiments, including informal language and slang. These findings provide valuable insights for developers to enhance the quality of the KAI Access application.

Keywords: sentiment analysis, KAI Access, Twitter, BERT, Multilingual BERT