

**ANALISIS SENTIMEN APLIKASI GOJEK MENGGUNAKAN
ALGORITMA BERT**

SKRIPSI

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



disusun oleh

NUGROHO WISNU SAPUTRO

20.11.3415

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN APLIKASI GOJEK MENGGUNAKAN
ALGORITMA BERT**

SKRIPSI

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



disusun oleh

NUGROHO WISNU SAPUTRO

20.11.3415

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN APLIKASI GOJEK MENGGUNAKAN
ALGORITMA BERT**

yang disusun dan diajukan oleh

Nugroho Wisnu Saputro

20.11.3415

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Februari 2025

Dosen Pembimbing,



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

NIK. 190302375

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN APLIKASI GOJEK MENGGUNAKAN
ALGORITMA BERT

yang disusun dan diajukan oleh

Nugroho Wisnu Saputro

20.11.3415

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Acihmah Sidauruk, S.Kom., M.Kom.
NIK. 190302238



Nafiatun Sholihah, S.Kom., M.Cs.
NIK. 190302524



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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 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 : Nugroho Wisnu Saputro
NIM : 20.11.3415

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Aplikasi Gojek Menggunakan Algoritma BERT

Dosen Pembimbing : Theopilus Bayu Sasongko, 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, 25 Februari 2025

Yang Menyatakan,



Nugroho Wisnu Saputro

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamin, skripsi ini saya persembahkan untuk:

1. Allah SWT karena sudah memberikan kemudahan kepada penulis untuk menyelesaikan skripsi ini.
2. Bapak, Ibu, kakak, dan keluarga besar karena menjadi motivasi dan selalu memberikan dukungan, semangat dan doa untuk penulis agar dapat menyelesaikan skripsi.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng., selaku dosen pembimbing yang selalu membantu dan memberi masukan untuk penulis agar laporan skripsi ini selesai dengan baik.
4. Teman seperjuangan S1 Informatika terutama sahabat saya Muhamad Hanif Fadhilah dan Athariq Raffi Fahrezi Ediningtyas dari IF-02 yang selalu memberikan dukungan, motivasi dan petunjuk untuk menyelesaikan tugas akhir skripsi ini.
5. Terakhir, saya ucapkan terimakasih untuk Nugroho Wisnu Saputro, diri saya sendiri.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul Analisis Sentimen Aplikasi Gojek Menggunakan Algoritma BERT ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer di Universitas Amikom Jogjakarta.

Penyusunan skripsi ini tidak terlepas dari dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng., selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan motivasi dalam penyusunan skripsi ini.
2. Orang tua dan keluarga tercinta, yang selalu memberikan doa, dukungan moral, serta semangat dalam menyelesaikan skripsi ini.
3. Teman-teman seperjuangan, yang selalu memberikan bantuan, dukungan, serta kebersamaan selama masa studi.
4. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dalam proses penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi kontribusi yang berharga

Yogyakarta, 4 November 2024

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.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.1 Rumusan Masalah.....	2
1.2 Batasan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	3
1.5 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.2 Dasar Teori	10
2.2.1 Gojek	10
2.2.2 Analisis Sentimen.....	10
2.2.3 BERT (<i>Bidirectional Encoder Representations from Transformers</i>)....	11
2.2.3.1 Tahap-tahap BERT.....	15
2.2.4 <i>Deep Learning</i>	17
2.2.5 Data Splitting.....	18
2.2.6 Confusion Matrix.....	19
2.2.7 Flask.....	21
BAB III METODE PENELITIAN.....	23
3.1 Objek Penelitian	23

3.2	Alur Penelitian	23
3.2.1	Mencari Dataset.....	24
3.2.2	Data Processing.....	24
3.2.2.1	<i>Case Folding</i>	25
3.2.2.2	<i>Normalisasi</i>	25
3.2.2.4	Tokenizing	28
3.2.2.5	Stemming.....	29
3.2.3	Labeling	31
3.3	Balancing Dataset Menggunakan SMOTE.....	32
3.4	EDA (Exploratory Data Analyst)	33
3.5	Data preparation	34
3.2.6.1	Encoding BERT.....	34
3.2.6.2	Data Splitting.....	34
3.6	Modeling BERT.....	34
3.6.1	<i>Pretrained Model</i>	35
3.6.2	Optimizer	35
3.7	Prediction dan Evaluation.....	35
3.3	Deployment Web App menggunakan Flask.....	35
3.4	Alat dan Bahan.....	36
3.4.1	Sentimen Alisis :	36
3.4.2	Deployment Web App	37
BAB IV HASIL DAN PEMBAHASAN.....		38
4.1	Pengumpulan Data	38
4.2	Processing dataset.....	38
4.2.1	Case Folding.....	39
4.2.2	Normalisasi.....	39
4.2.3	Stopwords.....	40
4.2.4	Tokenisasi.....	41
4.2.5	Stemming.....	41
4.3	Balancing Dataset Menggunakan SMOTE.....	42
4.4	EDA	43
4.4.1	Visualisasi Persentase.....	43
4.4.2	WordCloud	44
4.4.2.1	Positif.....	44

4.4.2.2	Negatif.....	45
4.4.2.3	Netral.....	45
4.5	Data Preparation	46
4.6	Tokenizer dan Encoding BERT.....	46
4.7	Data Splitting	48
4.8	Data Loader	49
4.8	Modeling BERT	49
4.8.1	BERT Pre-trained Model.....	50
4.8.2	Optimizer dan Scheduler	51
4.8.3	Training BERT	51
4.9	Prediction dan Evaluation	52
4.10	Deployment Web App Analisis Sentimen	55
4.1.1	Desain dan Arsitektur Aplikasi.....	57
4.1.2	Dataset dan Pelatihan Model.....	58
BAB V	PENUTUP	59
5.1	Kesimpulan.....	59
5.2	Saran	60
REFERENSI		61
LAMPIRAN		Error! Bookmark not defined.

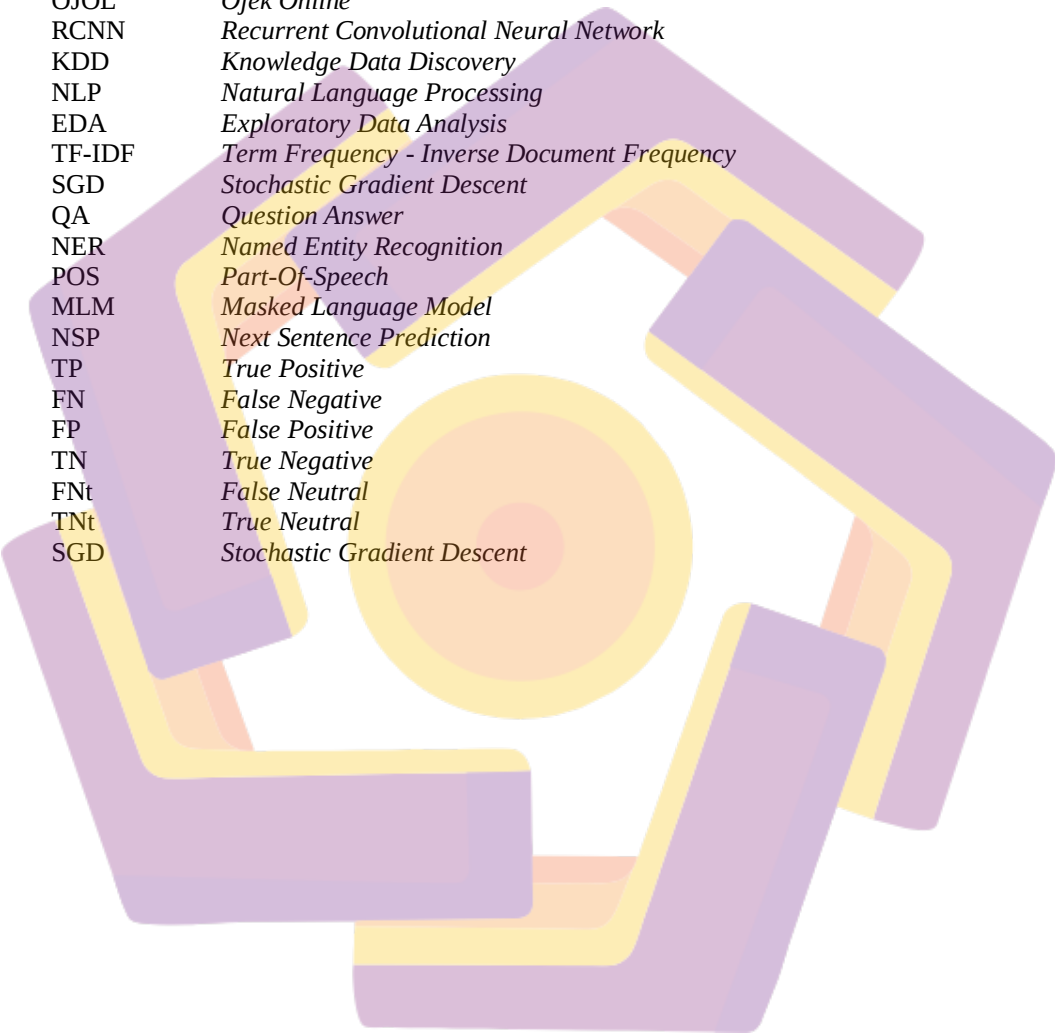
DAFTAR TABEL

Tabel 2 1 Daftar Keaslian Penelitian	7
Tabel 2 2 Confision Matrix	19
Tabel 4.1 Dataset	38
Tabel 4 2 Case folding	39
Tabel 4.3 Hasil Normalisasi	40
Tabel 4 4 Hasil Stopwords	40
Tabel 4 5 Hasil Tokenisasi	41
Tabel 4 6 Hasil Stemming	41
Tabel 4 7 Hasil Balancong SMOTE	42
Tabel 4 8 Data Splitting	48
Tabel 4 9 Classification Report	54

DAFTAR GAMBAR

Gambar 2.1 Proses Analisis Sentimen.....	10
Gambar 2.2 Cara Kerja BERT	13
Gambar 2.3 Ilustrasi Deep Learning.....	17
Gambar 2.4 Flowchart Flask	21
Gambar 3.1 Alur Penelitian.....	23
Gambar 3.2 Flowchart Casefolding.....	25
Gambar 3.3 Flowchart Normalisasi.....	26
Gambar 3.4 Flowchart Stopwords.....	27
Gambar 3.5 Flowchart Tokenizing.....	29
Gambar 3.6 Stemming	30
Gambar 3.7 flowchart labeling	31
Gambar 3.8 Flowchart Balancing.....	33
Gambar 4.1 visualisasi sebelum dan setelah SMOTE	42
Gambar 4.2 Pie Chart.....	43
Gambar 4.3 Wordcloud Positif.....	44
Gambar 4.4 Wordcloud Negatif	45
Gambar 4.5 Wordcloud Netral	45
Gambar 4.6 Hasil Tokenizier BERT & Attention Mask	47
Gambar 4.7 Pretrained Model	50
Gambar 4.8 Visualisasi Training & Validation Loss.....	52
Gambar 4.9 Confisuion Matrix	53
Gambar 4.10 Web App Analisis Sentimen	56

DAFTAR LAMBANG DAN SINGKATAN



BERT	<i>Bidirectional Encoder Representation from Transformers</i>
ANN	<i>Artificial Neural Network</i>
SVM	<i>Support Vector Machine</i>
OJOL	<i>Ojek Online</i>
RCNN	<i>Recurrent Convolutional Neural Network</i>
KDD	<i>Knowledge Data Discovery</i>
NLP	<i>Natural Language Processing</i>
EDA	<i>Exploratory Data Analysis</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
SGD	<i>Stochastic Gradient Descent</i>
QA	<i>Question Answer</i>
NER	<i>Named Entity Recognition</i>
POS	<i>Part-Of-Speech</i>
MLM	<i>Masked Language Model</i>
NSP	<i>Next Sentence Prediction</i>
TP	<i>True Positive</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
TN	<i>True Negative</i>
FNt	<i>False Neutral</i>
TNt	<i>True Neutral</i>
SGD	<i>Stochastic Gradient Descent</i>

DAFTAR ISTILAH



Neuron	Unit dasar dalam jaringan saraf yang melakukan
Epoch	komputasi, mirip dengan sel saraf di otak. Satu putaran penuh melalui seluruh dataset selama
Adam	pelatihan model. Optimizer yang digunakan dalam pelatihan jaringan
PyTorch	saraf. Pustaka machine learning berbasis Python yang mendukung komputasi tensor dengan akselerasi
Encoder	GPU dan pelatihan model neural network secara fleksibel dan efisien.
Fine Tuning	Komponen dalam model yang mengubah input mentah menjadi representasi yang lebih berguna,
Softmax	sering digunakan dalam model sekuensial. Proses penyesuaian model yang telah dilatih sebelumnya pada dataset baru untuk meningkatkan
Dropout Layer	kinerja. Fungsi aktivasi yang mengubah output model menjadi probabilitas, sering digunakan dalam
Overfitting	klasifikasi multi-kelas. Teknik regularisasi yang acak mematikan neuron
Confusion Matrix	during pelatihan untuk mencegah overfitting. Ketika model belajar terlalu banyak detail dari data
Wordcloud	pelatihan, sehingga tidak bekerja baik pada data baru. Tabel yang menunjukkan performa model
	klasifikasi dengan membandingkan prediksi dan label sebenarnya.
	Visualisasi frekuensi kata dalam bentuk awan, di
	mana ukuran kata menunjukkan seberapa sering kata tersebut muncul.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Gojek menggunakan algoritma BERT (Bidirectional Encoder Representations from Transformers). Aplikasi Gojek, sebagai salah satu platform layanan terkemuka di Indonesia, menerima banyak ulasan dari pengguna yang mencerminkan pengalaman mereka. Analisis sentimen ini penting untuk memahami persepsi pengguna dan membantu perusahaan dalam meningkatkan layanan. Metode yang digunakan adalah BERT, sebuah model berbasis deep learning yang mampu memahami konteks teks secara bidirectional. Dataset yang digunakan terdiri dari 210.621 ulasan pengguna yang diambil dari Google Play Store. Proses penelitian meliputi pengumpulan data, preprocessing (seperti case folding, tokenisasi, dan stemming), balancing dataset menggunakan SMOTE, data splitting dibagi menjadi 70:20:10 untuk latih, validasi, dan uji data dan menggunakan optimizer AdamW dengan learning-rate $2e-6$ dan 3 epoch. Hasil penelitian menunjukkan bahwa model BERT mampu mengklasifikasikan sentimen dengan akurasi setinggi 0.82(82%) dengan model BERT yang recall sebesar 98%, precision sebesar 93% dan *f1-score* sebesar 95%, terutama untuk sentimen positif. Model ini juga diimplementasikan dalam aplikasi web menggunakan framework Flask untuk memudahkan analisis sentimen secara otomatis. Penelitian ini memberikan kontribusi dalam pemanfaatan teknologi machine learning untuk analisis sentimen, khususnya dalam konteks bahasa Indonesia.

Kata kunci: Analisis Sentimen, BERT, Gojek, SMOTE, Flask, Bahasa Indonesia.

ABSTRACT

This research aims to analyze user sentiment towards the Gojek application using the BERT (Bidirectional Encoder Representations from Transformers) algorithm. The Gojek application, as one of the leading service platforms in Indonesia, receives many reviews from users reflecting their experiences. This sentiment analysis is important to understand user perceptions and help companies improve services. The method used is BERT, a deep learning-based model that is able to understand text context bidirectionally. The dataset used consists of 210,621 user reviews taken from the Google Play Store. The research process includes data collection, preprocessing (such as case folding, tokenization, and stemming), balancing the dataset using SMOTE, data splitting divided into 70:20:10 for training, validation, and test data and using the AdamW optimizer with a learning-rate of $2e-6$ and 3 epochs. The research results show that the BERT model is able to classify sentiment with an accuracy as high as 0.82 (82%) with the BERT model having a recall of 98%, precision of 93% and f1-score of 95%, especially for positive sentiment. This model is also implemented in a web application using the Flask framework to facilitate automatic sentiment analysis. This research contributes to the use of machine learning technology for sentiment analysis, especially in the Indonesian language context.

Keyword: Sentiment Analysis, BERT, Gojek, SMOTE, Flask, Indonesian Language.