

**SENTIMEN ANALISIS PENGGUNA APLIKASI BIBIT
MENGUNAKAN ALGORITMA BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS
(BERT)**

SKRIPSI

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



disusun oleh

ASIF ULINNUHA

20.11.3656

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**SENTIMEN ANALISIS PENGGUNA APLIKASI BIBIT
MENGUNAKAN ALGORITMA BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS
(BERT)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
ASIF ULINNUHA
20.11.3656

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

**SENTIMEN ANALISIS APLIKASI BIBIT MENGGUNAKAN
ALGORITMA BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS (BERT)**

yang disusun dan diajukan oleh

Asif Ulinnuha

20.11.3656

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Juli 2024

Dosen Pembimbing,



Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

HALAMAN PENGESAHAN
SKRIPSI
SENTIMEN ANALISIS APLIKASI BIBIT MENGGUNAKAN
ALGORITMA BIDIRECTIONAL ENCODER REPRESENTATIONS
FROM TRANSFORMERS (BERT)

yang disusun dan diajukan oleh

Asif Ulinnuha

20.11.3656

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Ike Verawati, M.Kom
NIK. 190302237

Robert Marco, M.T., Ph.D
NIK. 190302228

Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Juli 2024

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 : Asif Ulinnuha
NIM : 20.11.3656

Menyatakan bahwa Skripsi dengan judul berikut:

**SENTIMEN ANALISIS APLIKASI BIBIT REKSADANA
MENGUNAKAN ALGORITMA BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS (BERT)**

Dosen Pembimbing : Yoga Pristyanto, 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 Juli 2024

Yang Menyatakan,


Asif Ulinnuha

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil akamin

“Dengan segala puji syukur ke hadirat Allah SWT yang telah melimpahkan karunia-Nya, sehingga saya bisa menyelesaikan skripsi ini.

Tugas akhir ini saya persembahkan kepada orang-orang tercinta ayah, ibu, dan adik. Terima kasih atas segala dukungan dan doanya.”

“Justru karena dunia ini kejam, kita jadi sadar dan menikmati bahwa hal-hal kecil yang membuat bahagia adalah hal-hal yang berharga” -Fiony Alveria

Tantri

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Allah SWT, yang telah memberikan petunjuk dan kekuatan dalam menyelesaikan Skripsi yang berjudul “Sentimen Analisis Aplikasi Bibit Menggunakan Algoritma Bidirectional Encoder Representations from Transformers (BERT)”. Skripsi ini merupakan syarat terakhir yang harus ditempuh untuk menyelesaikan pendidikan pada jenjang Strata Satu (S1) Program Studi Informatika Universitas Amikom Yogyakarta. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT yang senantiasa memberikan kesehatan, petunjuk dan kemudahan selama menyelesaikan tugas akhir.
2. Ayah dan Ibu yang telah sabar untuk selalu mendukung dan mendoakan penulis dalam menyelesaikan tugas akhir ini.
3. Bapak Yoga Pristyanto, S.Kom., M.Eng. selaku dosen pembimbing atas waktu, pengertian, segala bantuan dalam memberikan referensi dan inovasi, pemberian kritik dan saran yang membangun, dan juga pemberian semangat kepada penulis.
4. Seluruh Dosen dan Pegawai jurusan Informatika atas kebaikan, dukungan, dan bantuan yang diberikan kepada penulis dalam menyelesaikan skripsi.
5. Semua rekan yang telah membantu penyusunan tugas akhir.

Penulis menyadari penyusunan tugas akhir ini tentunya tidak lepas dari kekurangan dan masih jauh dari sempurna. Oleh karena itu penulis mengharapkan masukan berupa kritik, saran, dan tanggapan yang bersifat membangun dalam upaya pembelajaran lebih lanjut. Akhir kata semoga tugas akhir ini bermanfaat bagi semua pihak, dan penulis sendiri pada khususnya.

Yogyakarta, 25 Juli 2024

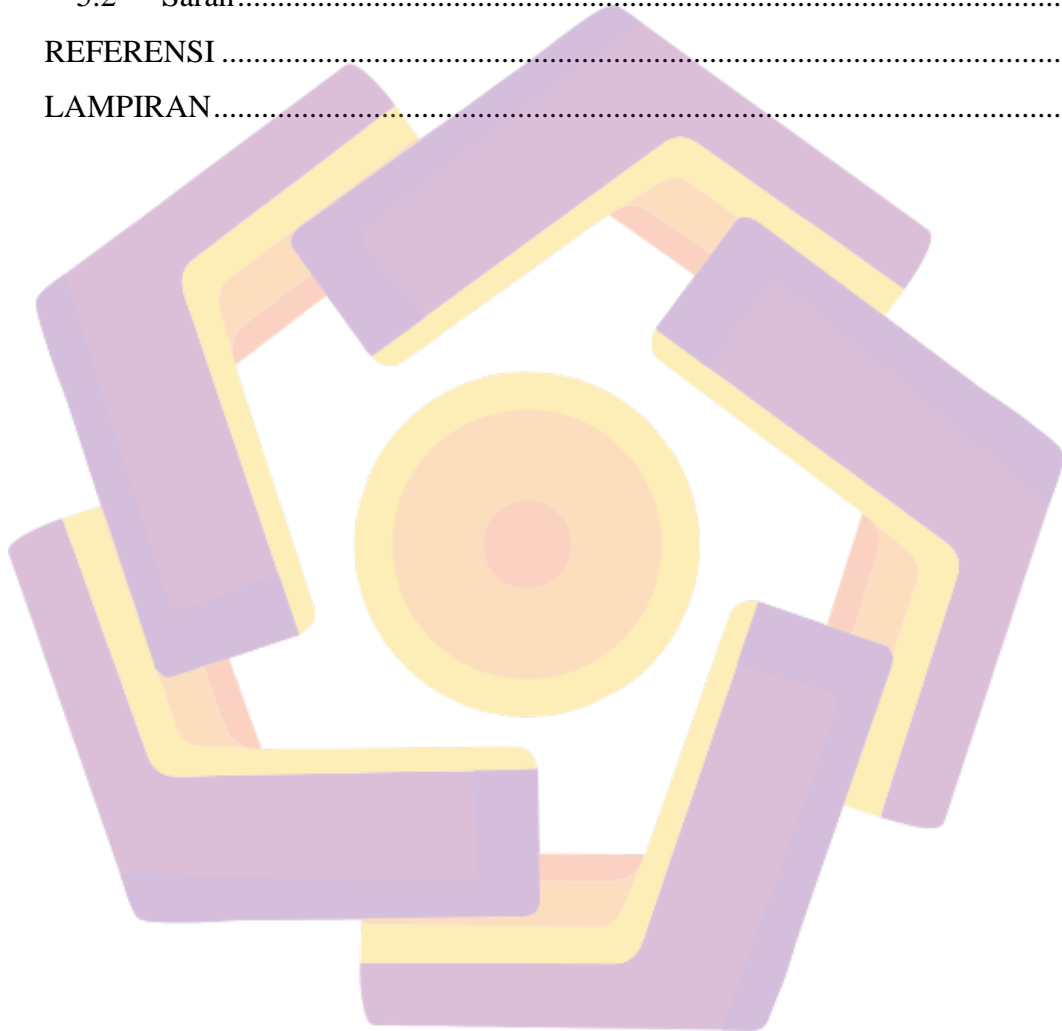
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	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.....	10
2.2.1. Reksa Dana	10
2.2.2. Bibit.....	10
2.2.3. Sentimen Analisis	11
2.2.4. Web Scraping	11
2.2.5. <i>Pre-Processing</i>	11
2.2.6. EDA	13

2.2.7.	Bidirectional Encoder Representations from Transformers (BERT)	
	13	
2.2.9.	Confusion Matrix	18
2.2.10.	<i>Python</i>	19
2.2.11.	<i>Google Colaboratory</i>	20
2.2.12.	Gradio.....	20
BAB III METODE PENELITIAN		22
3.1	Alur Penelitian.....	22
3.2	Pengumpulan Data	24
3.3	Pelabelan Data (<i>Labeling</i>).....	25
3.4	<i>Pre-Processing</i>	26
3.5	<i>Split Data</i>	26
3.6	<i>Encode</i>	26
3.6	Implementasi Model IndoBERT	27
3.7	Evaluasi Model.....	27
3.8	Grafik dan Diagram.....	28
3.9	Alat Bantu Penelitian	28
BAB IV HASIL DAN PEMBAHASAN		29
4.1	Pengumpulan Data	29
4.2	Labeling.....	31
4.3	<i>Pre-Processing Data</i>	33
4.3.1.	<i>Case Folding</i>	33
4.3.2.	<i>Filtering</i>	34
4.3.3.	<i>Tokenizing</i>	35
4.3.4.	<i>Stemming</i>	36
4.4	EDA.....	37
4.5	Tokenizer & Encode.....	42
4.6	<i>Split Data</i>	46
4.7	Membuat <i>Data Loaders</i>	46
4.8	<i>Modelling IndoBERT</i>	47
4.8.1.	Set-up BERT Pretrained Model.....	47

4.8.2. <i>Fine Tuning & Train Model</i>	50
4.9 Evaluasi Model.....	52
4.10 Implementasi Prediksi dengan Gradio.....	61
BAB V PENUTUP	64
5.1 Kesimpulan.....	64
5.2 Saran.....	64
REFERENSI	66
LAMPIRAN.....	71



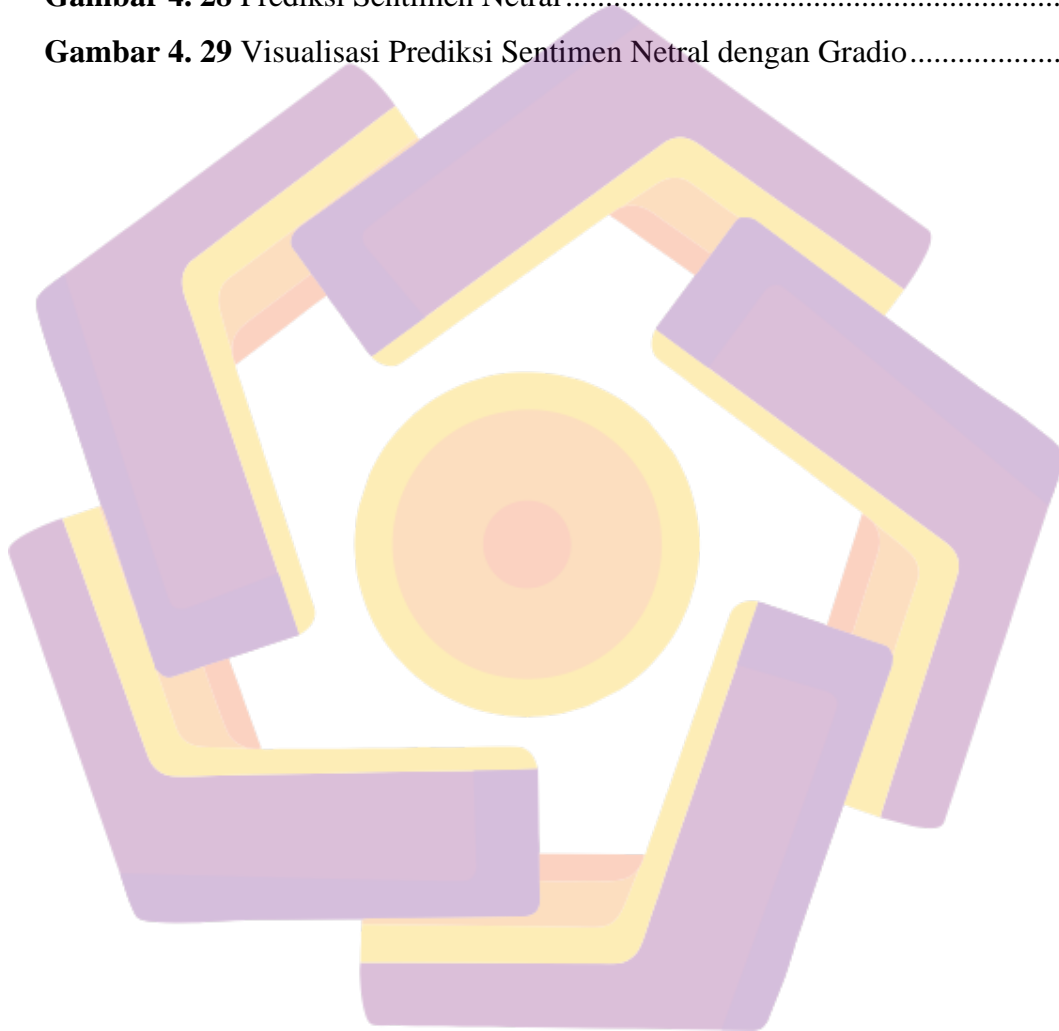
DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	7
Tabel 2. 2 Confusion Matrix multi-class classification.....	18
Tabel 3. 1 Deskripsi Variabel Hasil Scraping	24
Tabel 3. 2 Hyperparameter	27
Tabel 4. 1 Contoh dari Hasil Pengumpulan Data.....	31
Tabel 4. 2 Hasil Labeling	33
Tabel 4. 3 Contoh Hasil Case Folding	34
Tabel 4. 4 Contoh data Hasil Filtering	35
Tabel 4. 5 Contoh data Hasil Tokenizing.....	36
Tabel 4. 6 Contoh Hasil Stemmed	36
Tabel 4. 7 Proses Encode	45
Tabel 4. 8 Hyperparameter Penelitian.....	52
Tabel 4. 9 Hasil Epoch	54
Tabel 4. 10 Hasil Perhitungan Model.....	60

DAFTAR GAMBAR

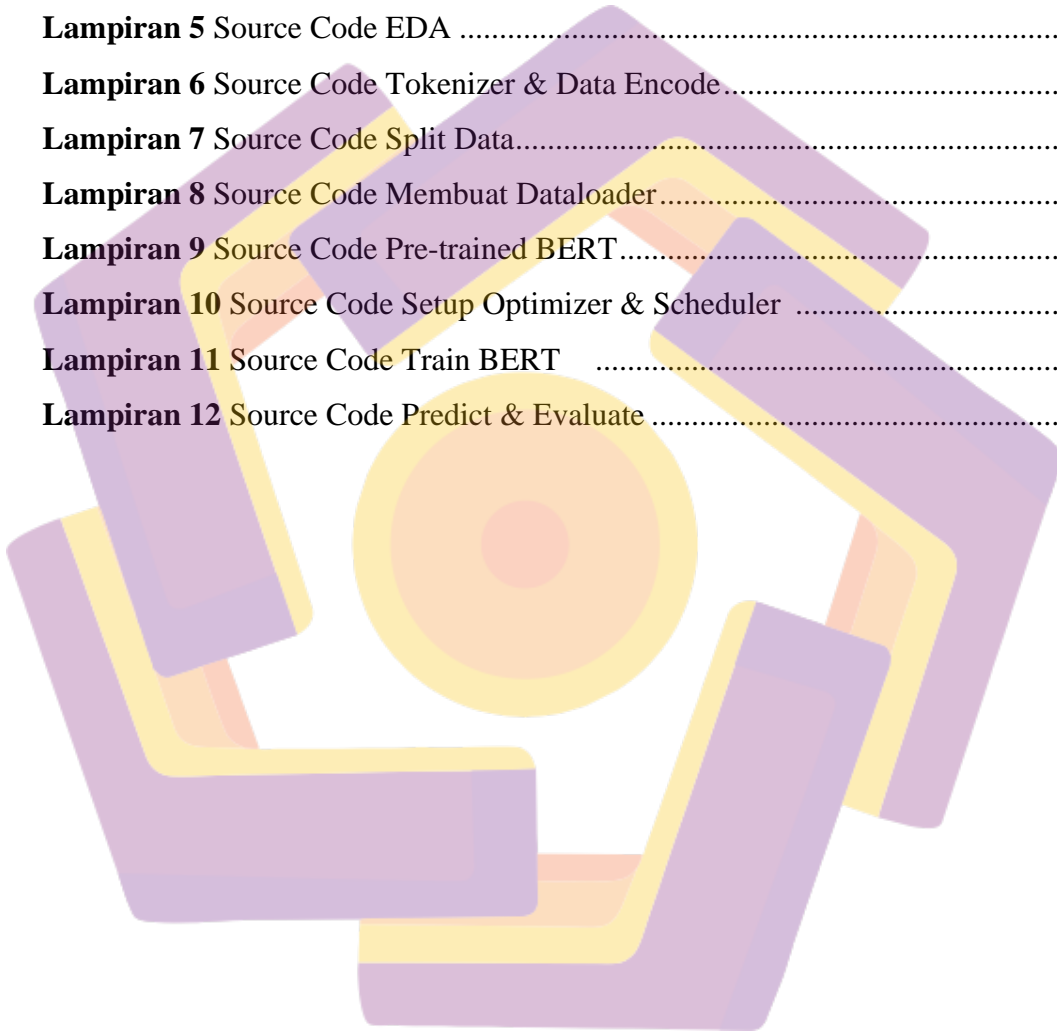
Gambar 2. 1 Arsitektur Transformers	14
Gambar 2. 2 Representasi Input BERT	15
Gambar 2. 3 Diagram Pre-training dan Fine-Tuning	16
Gambar 2. 4 Proses Masked Language Modeling.....	17
Gambar 3. 1 Alur Penelitian.....	22
Gambar 3. 2 Proses Encoder	27
Gambar 4. 1 Ulasan Pengguna Terhadap Aplikasi Bibit di Website Google Play Store	29
Gambar 4. 2 Data Mentah Hasil Scrapping.....	30
Gambar 4. 3 Penyebaran Data Menurut Score	37
Gambar 4. 4 Box Plot Penyebaran Rating.....	38
Gambar 4. 5 Bar Plot Berdasarkan Kategori.....	38
Gambar 4. 6 Pieplot Berdasarkan Kategori	39
Gambar 4. 7 Word Cloud Sentimen Positif.....	40
Gambar 4. 8 Word Cloud Sentimen Negatif	41
Gambar 4. 9 Word Cloud Sentimen Netral	42
Gambar 4. 10 BERT Tokenizer from IndoBERT-base-uncased.....	43
Gambar 4. 11 Proses Tokenisasi	43
Gambar 4. 12 Padding Data	45
Gambar 4. 13 Hasil Split Data.....	46
Gambar 4. 14 Proses Load Model	48
Gambar 4. 15 Srtuktur Parameter model BERT.....	49
Gambar 4. 16 Ilustrasi proses Fine-tuning IndoBERT untuk Klasifikasi	50
Gambar 4. 17 Code Hyperparameter.....	51
Gambar 4. 18 Code Train Model.....	51
Gambar 4. 19 Proses Training	53
Gambar 4. 20 Diagram Training Loss.....	54
Gambar 4. 21 Hasil Akhir Training.....	54
Gambar 4. 22 Diagram Confusion Matrix.....	55

Gambar 4. 23	Diagram Classification Report	60
Gambar 4. 24	Prediksi Sentimen Positif	61
Gambar 4. 25	Visualisasi Prediksi Sentimen Positif dengan Gradio	61
Gambar 4. 26	Prediksi Sentimen Negatif	62
Gambar 4. 27	Visualisasi Prediksi Sentimen Negatif dengan Gradio.....	62
Gambar 4. 28	Prediksi Sentimen Netral	62
Gambar 4. 29	Visualisasi Prediksi Sentimen Netral dengan Gradio.....	63



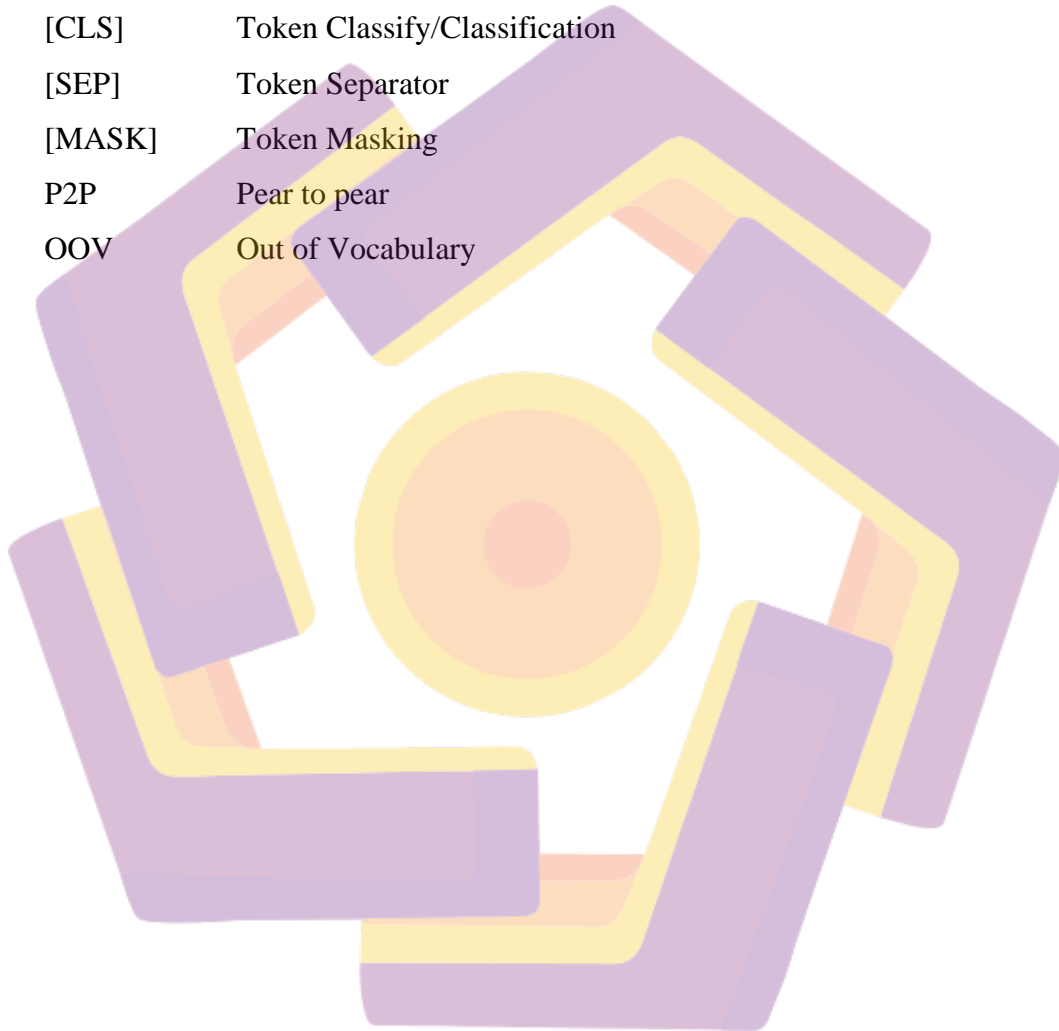
DAFTAR LAMPIRAN

Lampiran 1 Source Code Scraping Data Ulasan Pengguna Aplikasi Bibit	71
Lampiran 2 Import Library dan Load Dataset	72
Lampiran 3 Pre-processing	73
Lampiran 4 Source Code Export Dataset Hasil Preprocessing	75
Lampiran 5 Source Code EDA	76
Lampiran 6 Source Code Tokenizer & Data Encode	77
Lampiran 7 Source Code Split Data	78
Lampiran 8 Source Code Membuat Dataloader	78
Lampiran 9 Source Code Pre-trained BERT	79
Lampiran 10 Source Code Setup Optimizer & Scheduler	79
Lampiran 11 Source Code Train BERT	80
Lampiran 12 Source Code Predict & Evaluate	81



DAFTAR LAMBANG DAN SINGKATAN

BERT	Bidirectional Encoder Representation from Transformers
MLM	Masked Language Modeling
ML	Machine Learning
NSP	Next Sentence Prediction
[CLS]	Token Classify/Classification
[SEP]	Token Separator
[MASK]	Token Masking
P2P	Pear to pear
OOV	Out of Vocabulary



DAFTAR ISTILAH

Algoritma	prosedur yang diterapkan oleh model untuk mempelajari pola dari kata
Sentimen Analisis	proses menggunakan pemrosesan bahasa alami (NLP), analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari sumber teks
Scraping	proses pengambilan data
Akurasi	tingkat ketepatan prediksi model dengan nilai sebenarnya
Data Training	sampel yang digunakan untuk melatih model
Hyperparameter	parameter yang tidak dipelajari dari data pelatihan, tetapi ditentukan sebelum pelatihan dimulai
Epoch	satu siklus penuh di mana algoritma pembelajaran mesin menggunakan seluruh dataset untuk melatih model
Fine-tuning	proses mengadaptasi model BERT yang telah dilatih sebelumnya pada tugas-tugas umum agar dapat berfungsi lebih baik pada tugas tertentu
Data Testing	sampel yang digunakan untuk menguji model
Training loss	tingkat ketidaktepatan hasil prediksi dengan nilai sebenarnya
Dataset	kumpulan data yang digunakan untuk analisis
Presisi	tingkat ketepatan dari prediksi positif
Recall	tingkat kecakupan dalam mengidentifikasi semua instance positif
F1-score	tingkat keseimbangan antara prediksi positif dan kemampuan model mencakup sebagian besar kasus positif sebenarnya

INTISARI

Investasi digital atau online sudah mulai bermunculan sejak pertengahan tahun 2000-an. Investasi online hadir dalam berbagai jenis, di antaranya investasi saham, reksa dana, emas, valas (valuta asing) dan P2P (*Peer to peer*). Salah satu aplikasi yang mendukung investasi berbasis online ini adalah Bibit – Reksadana & Obligasi. Hingga saat ini telah diunduh lebih dari 10 juta orang di platform Google Play Store. Sebuah ulasan pengguna merupakan salah satu aspek penting dalam pengembangan sebuah aplikasi, karena berisi penilaian dan kritik pengguna semasa menggunakan aplikasi tersebut guna sebagai acuan untuk mengembangkan pelayanan yang lebih baik dimasa mendatang.

Analisis sentimen dapat berfungsi sebagai tolak ukur mengategorikan suatu ulasan, dimana dapat bernilai ulasan positif, negatif, atau netral. Seiring perkembangan teknologi, analisis sentimen dapat dilakukan secara otomatis. Penelitian ini akan mengukur performa metode *Bidirectional Encoder Representations from Transformers* (BERT) dalam melakukan klasifikasi sentimen menggunakan *pretrained* model *IndoBERT base uncased* dengan *fine-tuning*.

Jumlah data yang digunakan sebanyak 5000 data. Dan dibagi menggunakan rasio 70 : 30, dimana 70% digunakan sebagai data latih yaitu 3500 data, sedangkan 30% digunakan sebagai data uji sebanyak 1500 data. Penelitian ini menghasilkan tingkat akurasi sebesar 81% dengan pemilihan *hyperparameter*, yaitu *batch size* 16, *learning rate* 2e-5, dan *epoch* 10.

Kata kunci: analisis sentimen, BERT, Bibit, Google Play, *hyperparameter*.

ABSTRACT

Digital or online investments have emerged since the mid-2000s. Online investments come in various forms, including stocks, mutual funds, gold, forex (foreign exchange), and P2P (peer-to-peer) lending. One of the applications that support online-based investments is Bibit – Reksadana & Obligasi. To date, it has been downloaded by over 10 million users on the Google Play Store. User reviews are a crucial aspect of an application's development, as they contain assessments and critiques from users based on their experiences. These reviews serve as a reference for improving services in the future.

Sentiment analysis can serve as a benchmark for categorizing a review, indicating whether it is positive, negative, or neutral. With the advancement of technology, sentiment analysis can now be performed automatically. This research aims to measure the performance of the Bidirectional Encoder Representations from Transformers (BERT) method in sentiment classification using the pretrained model IndoBERT-base-uncased with fine-tuning techniques.

A total of 5000 data points were used in this study. These were divided using a 70:30 ratio, where 70% (3500 data points) were used for training, and 30% (1500 data points) were used for testing. The study achieved an accuracy rate of 81% with the selected hyperparameters: a batch size of 16, a learning rate of $2e-5$, and 10 epochs.

Keyword: *sentiment analysis, BERT, Bibit, Google Play, hyperparameter.*