

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA
TERHADAP PENGUNGSI ROHINGYA MENGGUNAKAN
ALGORITMA FINE TUNING BERT**

SKRIPSI

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



disusun oleh

ATHARIQ RAFFI FAHREZI EDININGTYAS

20.11.3411

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA
TERHADAP PENGUNGSI ROHINGYA MENGGUNAKAN
ALGORITMA FINE TUNING BERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ATHARIQ RAFFI FAHREZI EDININGTYAS

20.11.3411

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP
PENGUNGS ROHINGYA MENGGUNAKAN ALGORITMA FINE
TUNING BERT**

yang disusun dan diajukan oleh

ATHARIQ RAFFI FAHREZI EDININGTYAS
20.11.3411

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Agustus 2024

Dosen Pembimbing,



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

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP
PENGUNGSI ROHINGYA MENGGUNAKAN ALGORITMA FINE
TUNING BERT**

yang disusun dan diajukan oleh

ATHARIQ RAFFI FAHREZI EDININGTYAS

20.11.3411

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Agustus 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Balta, M. Kom.
NIK. 190302290

Ikmah, M.Kom.
NIK. 190302282

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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Agustus 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 : **ATHARIQ RAFFI FAHREZI EDININGTYAS**
NIM : **20.11.3411**

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP
PENGUNGSI ROHINGYA MENGGUNAKAN ALGORITMA FINE
TUNING 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, 12 Agustus 2024

Yang Menyatakan,



ATHARIQ RAFFI FAHREZI EDININGTYAS

HALAMAN PERSEMBAHAN

Puji dan syukur ke hadirat Allah Subhanahu wa Ta'ala atas segala nikmat yang telah menyertai saya dalam menyelesaikan tugas akhir yang berjudul “ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA TERHADAP PENGUNGSI ROHINGYA MENGGUNAKAN ALGORITMA FINE TUNING BERT”. Dengan kerendahan hati dan penuh rasa syukur, saya mempersembahkan tugas akhir ini kepada:

1. Allah SWT, karena berkat rahmat dan hidayah-Mu, saya masih diberi kesehatan dan kenikmatan sehingga dapat menyelesaikan tugas akhir ini.
2. Kedua orang tua, kedua saudara dan seluruh keluarga, yang selalu memberikan motivasi, apresiasi, dukungan dan doa selama penulis mengerjakan tugas akhir ini.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng., selaku dosen pembimbing yang telah membimbing penulis selama pengerjaan. Terima kasih atas ilmu, waktu, dan dorongan yang telah diberikan kepada penulis selama proses pengerjaan tugas akhir ini.
4. Seluruh teman-teman IF-02, kos rahmad, dan *future theory* yang selalu memberikan dukungan dan masukan kepada penulis.

KATA PENGANTAR

Segala puji dan syukur kita panjatkan kepada Allah SWT atas berkat dan rahmat-Nya. Berkat karunia-Nya, penulis berhasil menyelesaikan skripsi yang berjudul “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Pengungsi Rohingya Menggunakan Algoritma Fine Tuning BERT”.

Penulis menyadari bahwa banyak pihak yang turut membantu dan mendukung dalam penulisan skripsi ini. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

- Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
- Ibu Dwi Nurani, M. Kom. selaku dosen wali.
- Bapak Theopilus Bayu Sasongko, S. Kom, M.Eng., selaku dosen pembimbing.
- Seluruh teman-teman yang telah mendukung dan membantu penelitian ini.

Akhir kata, penulis berharap dengan penelitian ini dapat memberikan manfaat yang besar bagi para pembaca dan peneliti. Semoga penelitian ini dapat terus dikembangkan dan memberikan manfaat bagi seluruh masyarakat.

Yogyakarta, 04 Agustus 2024

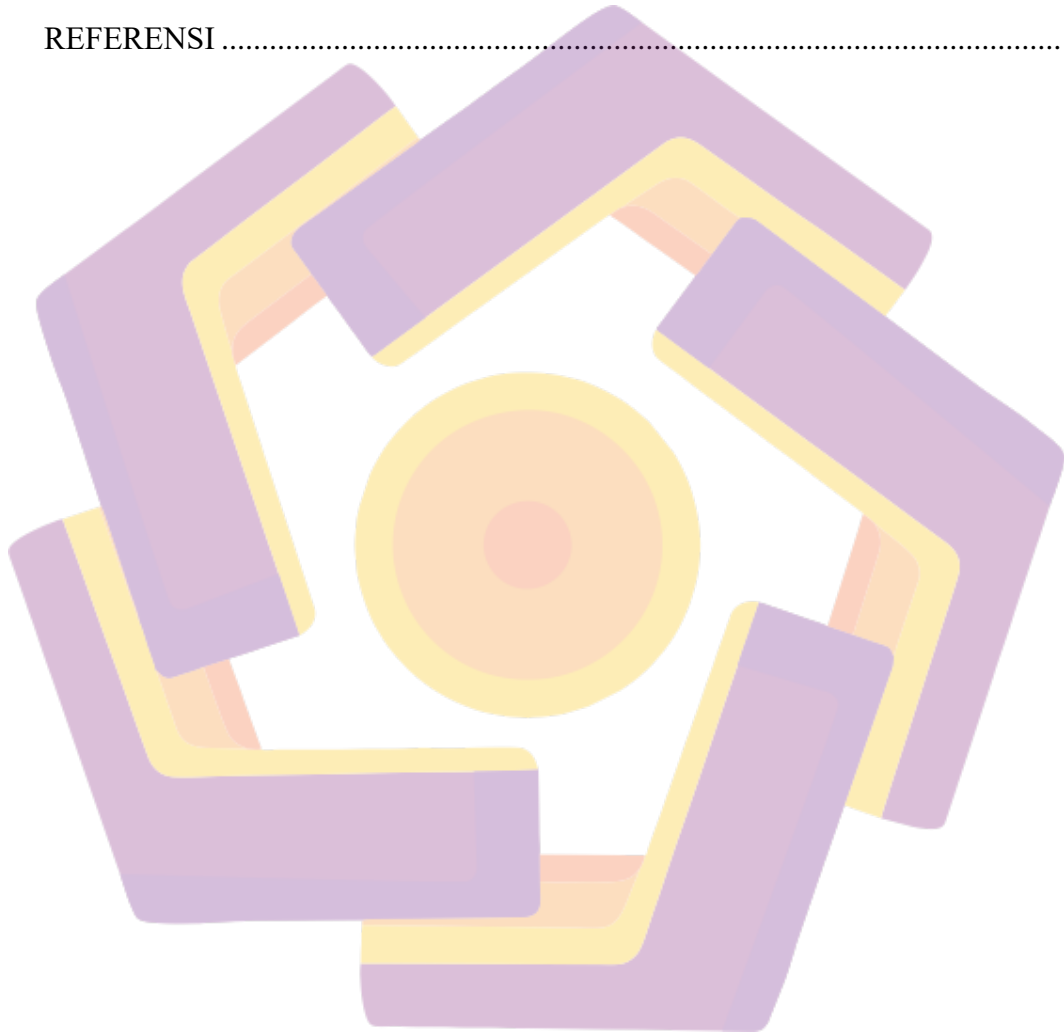
Penulis

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
1.6. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Studi Literatur	6
2.2. Dasar Teori	11
2.2.1. Pengungsi Rohingya	11
2.2.2. <i>Knowledge Discovery in Database (KDD)</i>	11
2.2.3. <i>Data Mining</i>	13
2.2.4. <i>Text Mining</i>	14
2.2.5. Analisis Sentimen	16
2.2.6. <i>Deep Learning</i>	16

2.2.7.	BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	18
2.2.8.	Evaluasi Model	27
BAB III METODE PENELITIAN.....		30
3.1.	Objek Penelitian	30
3.2.	Alur Penelitian	30
3.2.1.	<i>Dataset</i>	31
3.2.2.	<i>Data Preprocessing</i>	31
3.2.3.	EDA (<i>Exploratory Data Analysis</i>)	36
3.2.4.	<i>Data Preparation</i>	36
3.2.5.	Model BERT	38
3.2.6.	Prediksi dan Evaluasi Model.....	39
3.3.	Alat dan Bahan.....	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1.	Dataset.....	41
4.2.	<i>Data Preprocessing</i>	42
4.2.1.	<i>Case Folding</i>	42
4.2.2.	<i>Data Cleaning</i>	43
4.2.3.	<i>Stopwords Removal</i>	44
4.2.4.	<i>Tokenizing</i>	44
4.2.5.	<i>Stemming</i>	45
4.2.6.	<i>Labeling</i>	46
4.3.	EDA (<i>Exploratory Data Analysis</i>)	47
4.3.1.	Visualisasi Persentase	48
4.3.2.	Visualisasi <i>WordCloud</i>	48
4.4.	<i>Data Preparation</i>	50
4.4.1.	<i>Tokenizing dan Encoding BERT</i>	50
4.4.2.	<i>Data Splitting</i>	51
4.4.3.	<i>DataLoader</i>	52
4.5.	Model BERT	52
4.5.1.	<i>Pre-trained BERT</i>	52
4.5.2.	<i>Optimizer dan Scheduler</i>	53

4.5.3. Fine-tuning BERT	54
4.5.4. Training Model.....	55
4.6. Prediksi dan Evaluasi Model.....	57
BAB V PENUTUP.....	62
5.1. Kesimpulan	62
5.2. Saran.....	62
REFERENSI	63



DAFTAR TABEL

Tabel 2.1 Tabel Keaslian Penelitian	8
Tabel 2.2 Tabel Confusion Matrix.....	28
Tabel 3.1 Hasil memuat dataset	31
Tabel 4.1 Dataset.....	41
Tabel 4.2 Hasil <i>Case Folding</i>	42
Tabel 4.3 Hasil <i>Data Cleaning</i>	43
Tabel 4.4 Hasil <i>Stopwords Removal</i>	44
Tabel 4.5 Hasil <i>Tokenizing</i>	45
Tabel 4.6 Hasil <i>Stemming</i>	45
Tabel 4.7 Hasil <i>Labeling</i>	46
Tabel 4.8 Jumlah Setiap Kelas Sentimen	47
Tabel 4.9 Hasil <i>Splitting</i>	51
Tabel 4.10 Tabel Informasi <i>Model Fine-tuning</i> Selama <i>Training</i>	56
Tabel 4.11 Tabel Informasi <i>Model BERT</i> Selama <i>Training</i>	57
Tabel 4.12 Tabel <i>Classification Report Fine-tuning BERT</i>	59
Tabel 4.13 Tabel <i>Classification Report BERT</i>	60
Tabel 4.14 Tabel Perbandingan Hasil <i>Fine-tuning</i> dan <i>BERT Biasa</i>	61

DAFTAR GAMBAR

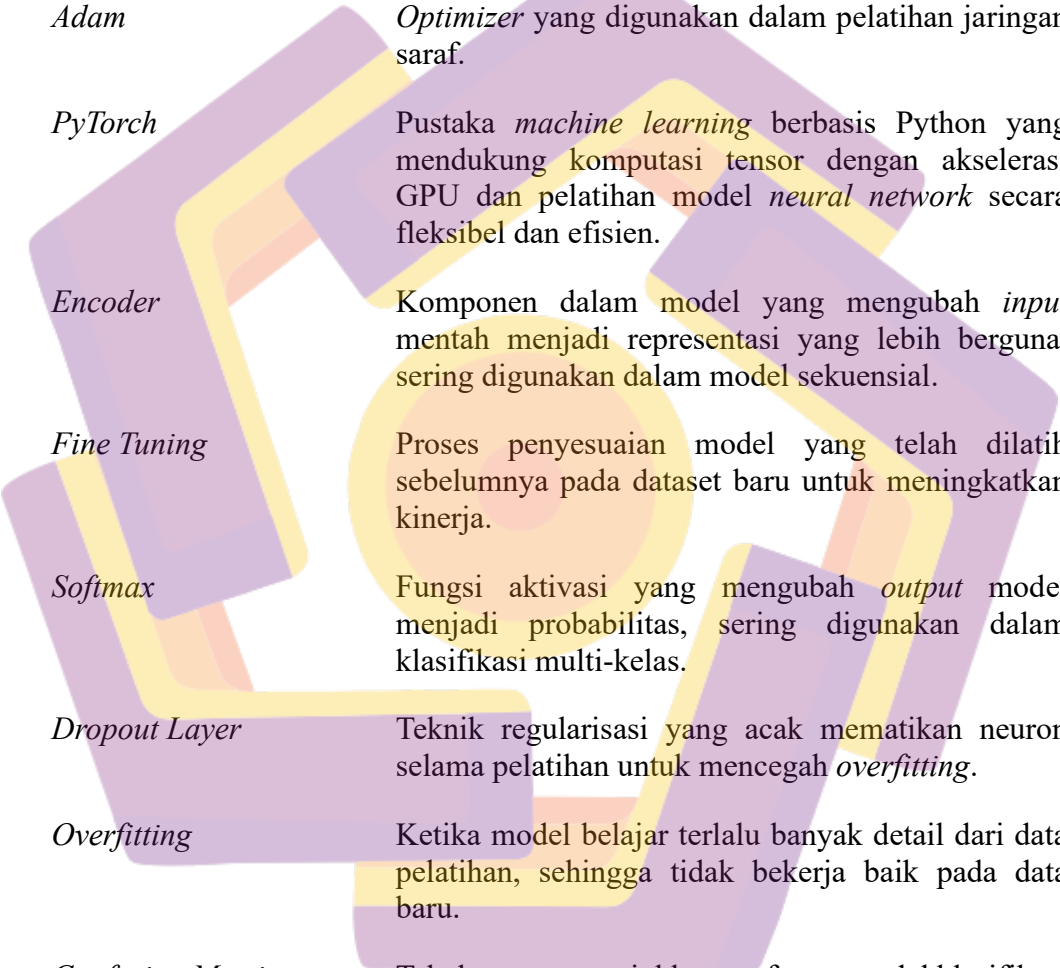
Gambar 2.1. Tahapan dalam KDD.....	12
Gambar 2.2. Tahapan dalam Text Mining.....	16
Gambar 2.3. Ilustrasi Arsitektur Deep Learning.....	17
Gambar 2.4 Ilustrasi proses pada <i>encoder</i> BERT.....	19
Gambar 2.5 Ilustrasi MLM dan NSP.....	24
Gambar 2.6 Ilustrasi fine-tuning BERT.....	25
Gambar 2.7 Ilustrasi token embeddings.....	27
Gambar 3.1. Alur Penelitian.....	30
Gambar 3.2 Alur <i>case folding</i>	32
Gambar 4.1 Hasil distribusi sentimen.....	47
Gambar 4.2 Hasil visualisasi persentase.....	48
Gambar 4.3 Hasil visualisasi positif.....	49
Gambar 4.4 Hasil visualisasi negatif.....	49
Gambar 4.5 Hasil visualisasi netral.....	50
Gambar 4.6 Hasil pengecekan panjang teks.....	51
Gambar 4.7 Hasil <i>tokenizing</i> dan <i>encoding</i> BERT.....	51
Gambar 4.8 Arsitektur <i>Pre-trained</i> Model.....	53
Gambar 4.9 Arsitektur Fine-tuning BERT.....	55
Gambar 4.10 Grafik <i>Training Acc</i> dan <i>Loss Fine-tuning</i> BERT.....	56
Gambar 4.11 Grafik <i>Training Acc</i> dan <i>Loss Model</i> BERT.....	57
Gambar 4.12 Hasil <i>Confusion Matrix Fine-tuning</i> BERT.....	58
Gambar 4.13 Hasil <i>Confusion Matrix Model</i> BERT.....	59

DAFTAR LAMBANG DAN SINGKATAN



BERT	<i>Bidirectional Encoder Representation from Transformers</i>
ANN	<i>Artificial Neural Network</i>
SVM	<i>Support Vector Machine</i>
PUBG	<i>Player Unknown Battle Ground</i>
RCNN	<i>Recurrent Convolutional Neural Network</i>
KDD	<i>Knowledge Data Discovery</i>
NLP	<i>Natural Language Processing</i>
EDA	<i>Explaratory 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



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

INTISARI

Krisis kemanusiaan yang melibatkan pengungsi Rohingya telah menciptakan dampak sosial dan politik yang signifikan di berbagai negara. Krisis kemanusiaan ini memiliki akar yang kompleks dan bersejarah. Opini dan persepsi masyarakat Indonesia akan kedatangan para pengungsi Rohingya menjadi sangat penting, karena hal ini dapat memengaruhi sikap negara, kebijakan pemerintah, serta respons masyarakat terhadap kehadiran dan kebutuhan pengungsi Rohingya. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap pengungsi Rohingya menggunakan *fine-tuning* BERT, dengan klasifikasi sentimen positif, netral, dan negatif, serta mengevaluasi pengaruh *fine-tuning* pada kinerja BERT. Hasil dari penelitian ini menunjukkan bahwa *fine-tuning* BERT lebih unggul dalam klasifikasi sentimen dibandingkan BERT standar. Dengan dataset 5902 data, rasio pembagian 80:10:10 untuk latihan, validasi, dan uji, serta *hyperparameter* batch size 32, *optimizer* AdamW dengan *learning rate* $2e-5$, dan 4 *epoch*, model *fine-tuning* BERT menghasilkan akurasi sebesar 81% sedangkan BERT standar hanya mencapai akurasi sebesar 76%. *Fine-tuning* BERT juga unggul dalam mengklasifikasikan sentimen negatif dengan *recall*, *precision*, dan *f1-score* 0.88, serta performa yang sedikit lebih baik untuk kelas netral dan positif dengan *f1-score* masing-masing 0.71 dan 0.65.

Kata kunci: Rohingya, Pengungsi Rohingya, BERT, Sentimen Analisis, *Deep Learning*, *Neural Network*.

ABSTRACT

The humanitarian crisis involving Rohingya refugees has created significant social and political impacts in various countries. This crisis has complex and historical roots. The opinions and perceptions of the Indonesian public regarding the arrival of Rohingya refugees are crucial as they can influence the country's attitude, government policies, and societal response to the presence and needs of these refugees. This study aims to analyze the sentiment of the Indonesian public towards Rohingya refugees using fine-tuning BERT, classifying sentiments into positive, neutral, and negative categories, and evaluating the impact of fine-tuning on BERT's performance. The results show that fine-tuning BERT outperforms the standard BERT in sentiment classification. Using a dataset of 5902 data points, split into 80:10:10 ratios for training, validation, and testing, with hyperparameters including a batch size of 32, AdamW optimizer with a learning rate of $2e-5$, and 4 epochs, the fine-tuned BERT model achieved an accuracy of 81% (compared to the standard BERT's 76%). Fine-tuning BERT also excelled in classifying negative sentiments with a recall, precision, and f1-score of 0.88, and performed slightly better for neutral and positive classes with f1-scores of 0.71 and 0.65, respectively.

Keywords: *Rohingya, Rohingya Refugees, BERT, Sentiment Analysis, Deep Learning, Neural Network.*