

**ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE
MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN
*INDOBERT***

SKRIPSI

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



disusun oleh

M. BAYUN NURPRASETYO

21.11.3922

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE
MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN
INDOBERT**

SKRIPSI

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



disusun oleh

M. BAYUN NURPRASETYO

21.11.3922

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE
MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN
*INDOBERT***

yang disusun dan diajukan oleh

M. BAYUN NURPRASETYO

21.11.3922

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Juli 2025

Dosen Pembimbing,



Ferian Fauzi Abdullah, S.Kom., M.Kom.
NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE
MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN
INDOBERT**

yang disusun dan diajukan oleh

M. Bayun Nurprasetyo

21.11.3922

Telah dipertahankan di depan Dewan Penguji
pada tanggal 29 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

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

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 29 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : M. BAYUN NURPRASETYO
NIM : 21.11.3922

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE
MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN
INDOBERT**

Dosen Pembimbing : Ferian Fauzi Abdulloh, 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, 29 Juli 2025

Yang Menyatakan,



METERAI TEMPEK
66110AMX319222928

M. Bayun Nurprasetyo

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan bahagia, saya panjatkan puji syukur ke hadirat Allah SWT atas segala nikmat, kekuatan, dan kemudahan yang telah diberikan hingga saya dapat menyelesaikan tugas akhir ini. Skripsi ini menjadi bukti dari setiap perjuangan, harapan, dan doa yang mengiringi langkah saya selama menempuh pendidikan. Karya ini saya persembahkan kepada:

1. Allah SWT Zat Yang Maha Mengetahui, tempat saya memohon kekuatan dan keikhlasan dalam menjalani setiap proses.
2. Kedua orang tua kandung saya tercinta Nuriadianto dan Siti Aminah, terima kasih atas cinta, doa, dan pengorbanan yang tak pernah lelah kalian berikan. Bersama dengan saudara-saudaraku Raditya Dwi Nayaka, Satria Nur Fitriansyah, dan Wahyuni Nur Sholehah yang selalu memberi dukungan dan semangat.
3. Orang tua kedua saya Pakde Agung Budiarto dan Bude Nuristiani, terima kasih telah membesarkan saya sejak Kecil hingga saat ini dengan kasih sayang yang tulus, serta selalu mengingatkan untuk tidak melupakan orang tua kandung saya. Kalian adalah teladan hidup saya. Dan juga, mbak Sekar Jingga Citranuari dan mbak Bunga Azaria, sudah saya anggap seperti kakak sendiri yang selalu menemani dan mendukung.
4. Keluarga besar Bantul Family Terima kasih atas doa, perhatian, dan dorongan selama masa studi. Khususnya kepada Muhammad Affidianto Arusda yang banyak memberi semangat di masa-masa sulit.
5. Teman-teman Bantul Warung-Pedak khususnya kak Fransiska Adiningtyas dan rekan-rekan lainnya yang menjadi tempat saya bercerita dan berbagi selama menyusun skripsi ini.
6. Teman-teman tongkrongan Bima, Amalia, dan yang lainnya yang selalu menyebarkan semangat positif serta mengingatkan pentingnya progres.

7. Teman-teman IF-02 Eker, Ocan, vika dan yang lainnya, terima kasih atas semangat, bantuan, dan kebersamaan selama menyusun skripsi dari awal hingga akhir. serta penghuni Kontrakan Ceria Fakhri, Febri, dan Aska, terima kasih telah mengizinkan saya hadir dan merasa diterima di tengah-tengah kalian.
8. Seluruh teman dan sahabat lainnya yang tidak bisa saya sebutkan satu per satu, tetapi telah memberikan dukungan, doa, candaan, dan semangat dalam berbagai bentuk. Saya sangat menghargainya.

Sebagai penutup dari persembahan ini, izinkan saya mengutip sepenggal lirik yang begitu mewakili perjalanan ini:

“Dan biarkan hanya cinta yang dapat tumbuh disini.” — Rebellion Rose

Semoga skripsi ini tumbuh dari cinta, perjuangan, dan doa yang tulus, serta menjadi langkah awal untuk berkarya lebih luas dan memberi makna lebih dalam kehidupan.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul "ANALISIS SENTIMEN KOMENTAR PENGGUNA YOUTUBE MENGENAI PEMAIN TIMNAS DIASPORA MENGGUNAKAN INDOBERT". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penyusunan skripsi ini tidak lepas dari bantuan, dukungan, serta bimbingan berbagai pihak. Dengan hormat, penulis menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusri, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta beserta seluruh jajarannya.
4. Bapak Ferian Fauzi Abdulloh, S.Kom., M.Kom., selaku dosen pembimbing yang telah meluangkan waktu, memberikan saran, kritik, motivasi, dan bimbingan dengan penuh perhatian selama proses penyusunan skripsi ini.
5. Kedua orang tua saya yang telah memberikan dukungan, doa, motivasi, dan pendidikan sehingga dapat menempuh pendidikan hingga jenjang perguruan tinggi.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang pemrosesan bahasa alami (Natural Language Processing) dan analisis sentimen pada media sosial.

Yogyakarta, 31 Juli 2025



M. Bayun Nurprasetyo

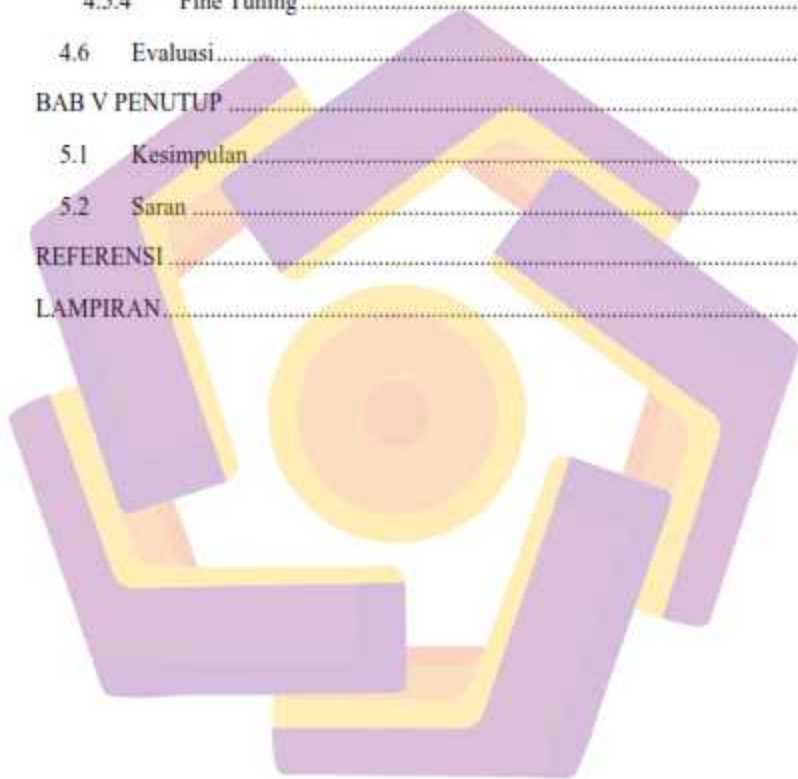
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
DAFTAR LAMBANG DAN SINGKATAN	xvi
DAFTAR ISTILAH.....	xvii
INTISARI	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	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	7
2.1 Studi Literatur	7

2.2	Dasar Teori.....	16
2.2.1	Youtube.....	16
2.2.2	Opini Publik.....	16
2.2.3	Tim Nasional Indonesia.....	16
2.2.4	Pemain Diaspora.....	17
2.2.5	Analisis Sentimen.....	17
2.2.6	<i>Scraping Data</i>	18
2.2.7	<i>Natural Language Processing</i>	18
2.2.8	<i>Deep Learning</i>	18
2.2.9	<i>Bidirectional Encoder Representations from Transformers</i>	19
2.2.10	<i>IndoBERT</i>	22
2.2.11	<i>K-Fold Cross Validation</i>	22
2.2.12	<i>Confusion Matrix</i>	23
BAB III METODE PENELITIAN.....		26
3.1	Objek Penelitian.....	26
3.2	Alur Penelitian.....	27
3.2.1	<i>Scraping Data</i>	28
3.2.2	<i>Labeling Data</i>	28
3.2.3	<i>Exploratory Data Analysis</i>	28
3.2.4	<i>Preprocessing Data</i>	29
3.2.4.1.	Data Cleaning.....	29
3.2.4.2.	Case Folding.....	29
3.2.4.3.	Normalisasi.....	29
3.2.5	<i>Wordcloud</i>	30
3.2.6	Pra Pemodelan.....	30

3.2.6.1	<i>Input Formating</i>	30
3.2.6.2	Splitting Data	31
3.2.7	Implementasi Pemodelan	33
3.2.7.1	Data Loader	33
3.2.7.2	<i>Load Pre-Trained IndoBERT</i>	33
3.2.7.3	Optimizer	34
3.2.7.4	<i>Fine Tuning</i>	34
3.2.8	Evaluasi	35
3.3	Alat dan Bahan	35
3.3.1	Alat Instrumen	35
3.3.2	Bahan Penelitian	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Pengumpulan Dataset	37
4.1.1	<i>Scraping Data</i>	37
4.1.2	Labeling Data	38
4.1.3	EDA (<i>EXPLORATORY DATA ANALYSIS</i>)	39
4.2	Preprocessing Data	40
4.2.1	Data Cleaning	40
4.2.2	Case Folding	41
4.2.3	Normalize	41
4.3	Word Could	42
4.4	Pra Pemodelan	43
4.4.1	Input Formating	43
4.4.2	Splitting Data	46
4.4.2.1	Pembagian Train-Test (80:20)	46
4.4.2.2	K-Fold Cross Validation	47

4.5	Implementasi Modeling	48
4.5.1	Data Loader.....	48
4.5.2	Load Pre Trained <i>IndoBERT</i>	49
4.5.3	Optimizer	49
4.5.4	Fine Tuning.....	50
4.6	Evaluasi.....	52
BAB V PENUTUP		57
5.1	Kesimpulan	57
5.2	Saran	57
REFERENSI		59
LAMPIRAN.....		63



DAFTAR TABEL

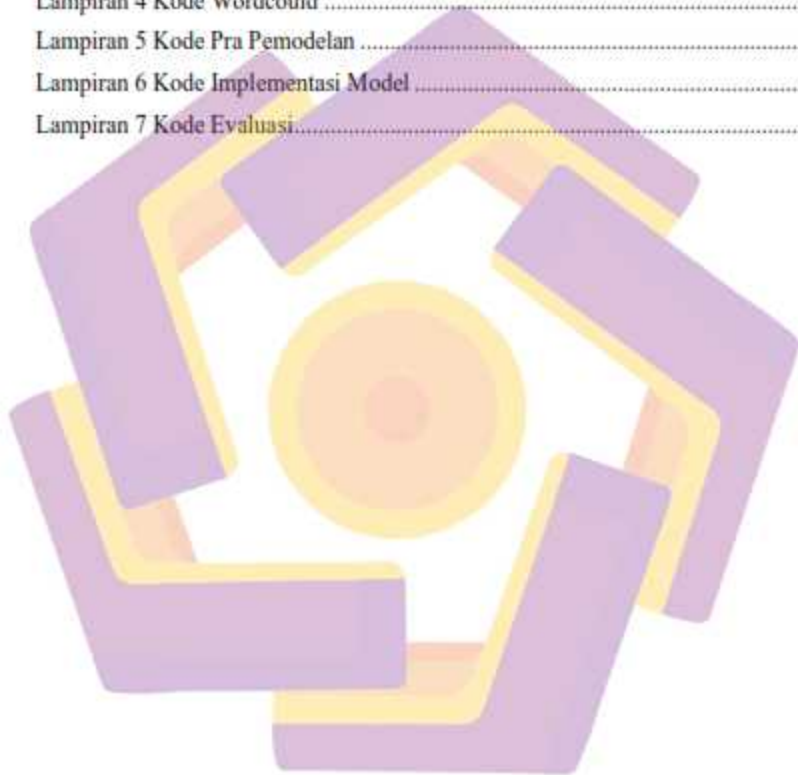
Tabel 2. 1 Keaslian Penelitian	10
Tabel 2. 2 Pengertian Confusion Matrix	24
Tabel 3. 1 Informasi Data Objek Penelitian.....	26
Tabel 3. 2 Tiga Data Komentar teratas	36
Tabel 4. 1 Hasil Scraping	37
Tabel 4. 2 Hasil Labeling.....	38
Tabel 4. 3 Jumlah Kelas Positif dan Negatif.....	39
Tabel 4. 4 Hasil Data Cleaning	41
Tabel 4. 5 Hasil Case Folding.....	41
Tabel 4. 6 Hasil Normalisasi.....	42
Tabel 4. 7 Hasil Tokenisasi.....	44
Tabel 4. 8 Hasil input_ids	45
Tabel 4. 9 Hasil Attention Mask	45
Tabel 4. 10 Hasil Tuning terbaik dari tiap fold.....	51

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Deep Learning	18
Gambar 2. 2 Alur Kerja BERT	20
Gambar 2. 3 Ilustrasi Proses K-Fold Cross Validation	22
Gambar 3. 1 Alur Penelitian	27
Gambar 3. 2 Skema Input Formating	31
Gambar 3. 3 Skema Fine Tuning	35
Gambar 4. 1 Diagram Pie	40
Gambar 4. 2 Visualisasi Komentar Positif	42
Gambar 4. 3 Visualisasi Komentar Negatif	43
Gambar 4. 4 Visualisasi Penentuan Nilai Max_length	44
Gambar 4. 5 Distribusi data train dan test	46
Gambar 4. 6 Distribusi data train dan validasi pada (5-fold CV)	47
Gambar 4. 7 Jumlah Batch tiap data train, validation, dan test	49
Gambar 4. 8 Grafik Loss dan Accuracy tiap Epoch pada ke 5 fold	52
Gambar 4. 9 Confusion matrix pada model <i>IndoBERT</i> dengan 5 fold	53
Gambar 4. 10 Diagram Akurasi tiap fold data train dan data test (unseen data)	54
Gambar 4. 11 Confusion Matrix data test pada fold 4	55

DAFTAR LAMPIRAN

Lampiran 1 Kode Scraping Data.....	63
Lampiran 2 Kode EDA	65
Lampiran 3 Kode Preprocessing	66
Lampiran 4 Kode Wordcoudl	69
Lampiran 5 Kode Pra Pemodelan	71
Lampiran 6 Kode Implementasi Model	76
Lampiran 7 Kode Evaluasi.....	84



DAFTAR LAMBANG DAN SINGKATAN

<i>MLM</i>	Masked Language Modeling
<i>NSP</i>	Next Sentence Prediction
<i>[MASK]</i>	Token khusus untuk menyembunyikan kata
<i>[CLS]</i>	Token awal untuk representasi seluruh kalimat.
<i>[SEP]</i>	Token pemisah antara dua segmen teks atau kalimat
$P(w_i context)$	Probabilitas munculnya kata
h_i	Representasi vektor dari kata ke- i dalam BERT.
W	Matriks bobot model
h_{CLS}	Vektor representasi dari token [CLS]
h_{sep}	Vektor representasi dari token [SEP]
W_s	Matriks bobot khusus untuk prediksi NSP.
<i>NSP</i>	Next Sentence Prediction
<i>MSE</i>	Mean Squared Error
$CV(k)$	Cross Validation – Rata-rata MSE dari K iterasi.
k	Jumlah fold dalam K-Fold Cross Validation.
$\hat{y}_i$	Nilai prediksi dari model untuk data ke- i .
y_i	Nilai aktual dari data ke- i .
<i>torch.tensor()</i>	Fungsi dari PyTorch yang mengubah array menjadi tensor.

DAFTAR ISTILAH



<i>Vektor</i>	besaran yang mempunyai arah
<i>Sentimen</i>	Pandangan atau opini yang terdapat dalam teks
<i>Tokenisasi</i>	Proses membagi teks menjadi unit-unit kecil
<i>Input Formatting</i>	Proses menyiapkan input teks
<i>Tensor</i>	Representasi array multidimensi
<i>Fine-Tuning</i>	Pelatihan tambahan pada model pra-latih pretrained
<i>Epoch</i>	Satu siklus lengkap pelatihan
<i>Precision</i>	Rasio antara prediksi benar dan total prediksi positif.
<i>Recall</i>	Rasio antara prediksi benar dan total data aktual positif.
<i>F1-score</i>	Rata-rata harmonis dari precision dan recall.
<i>Attention Mask</i>	Penanda token yang perlu diperhatikan oleh model
<i>IndoBERT</i>	Model BERT dalam bahasa Indonesia
<i>Unseen Data</i>	Data uji yang tidak dilibatkan saat pelatihan

INTISARI

Penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat terhadap keberadaan pemain diaspora dalam konteks sepak bola, khususnya di Tim Nasional Indonesia, dengan menggunakan model *IndoBERT* pada komentar YouTube. Isu naturalisasi pemain diaspora menimbulkan berbagai opini publik, yang tercermin dalam komentar di media sosial, terutama di YouTube.

Komentar-komentar tersebut dikumpulkan secara otomatis dari video YouTube yang membahas pemain diaspora di Timnas Indonesia. Selanjutnya, data tersebut diberi label sentimen secara manual ke dalam dua kategori: positif dan negatif. Tahapan prapemodelan melibatkan proses *preprocessing* yang mencakup *cleaning*, *case folding*, dan *normalisasi*. Setelah itu, data dikonversi menggunakan *tokenizer IndoBERT* dan diformat ke dalam tensor. Proses pemodelan dilakukan dengan metode *K-Fold Cross Validation* (5-Fold) dan *fine-tuning* model *IndoBERT* menggunakan *optimizer AdamW* dan *scheduler linear warmup*. Selain itu, 20% dari data dipisahkan sebagai data uji (*test set*) untuk menguji performa model terhadap data yang belum pernah dilihat sebelumnya.

Hasil penelitian menunjukkan bahwa *IndoBERT* mampu melakukan klasifikasi sentimen dengan akurasi terbaik sebesar 81,95% pada fold ke-4 dan 83,98% pada data uji terpisah. Temuan ini mengindikasikan bahwa model *IndoBERT* efektif untuk analisis sentimen komentar masyarakat mengenai keberadaan pemain diaspora di Timnas Indonesia. Penelitian ini dapat dimanfaatkan oleh peneliti, analis media sosial, dan federasi sepak bola nasional untuk memahami opini publik secara lebih sistematis.

Kata kunci: analisis sentimen, *IndoBERT*, pemain diaspora, Timnas Indonesia, opini publik.

ABSTRACT

This study aims to conduct sentiment analysis of public opinion regarding the presence of diaspora players in the context of football, particularly in the Indonesian National Team, using the IndoBERT model on YouTube comments. The issue of naturalized diaspora players has sparked various public reactions, reflected in social media comments, especially on YouTube.

These comments were automatically collected from YouTube videos discussing diaspora players in the Indonesian national team. The data were manually labeled into two sentiment categories: positive and negative. The pre-modeling stage involved preprocessing steps including cleaning, case folding, and normalization. The data were then tokenized using IndoBERT and formatted into tensors. The modeling process employed 5-Fold Cross Validation and fine-tuning of the IndoBERT model using the AdamW optimizer and linear warmup scheduler. Additionally, 20% of the data were set aside as a test set to evaluate model performance on unseen data.

The results of this study indicate that IndoBERT is capable of performing sentiment classification with a best accuracy of 81.95% on the fourth fold and 83.98% on the separate test set. These findings suggest that the IndoBERT model is effective for sentiment analysis of public comments regarding the presence of diaspora players in the Indonesian national football team. This research can be utilized by researchers, social media analysts, and the national football federation to better understand public opinion in a more systematic manner.

Keyword: *sentiment analysis, IndoBERT, diaspora players, Indonesian National Team, public opinion.*