ANALISIS SENTIMEN KOMENTAR YOUTUBE BERBAHASA INDONESIA TERKAIT ISU WARGA NEGARA INDONESIA DI JEPANG MENGGUNAKAN MODEL HYBRID INDOBERT DAN LSTM

SKRIPSI

VANESS FEBRAIO
00000055310

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

ANALISIS SENTIMEN KOMENTAR YOUTUBE BERBAHASA INDONESIA TERKAIT ISU WARGA NEGARA INDONESIA DI JEPANG MENGGUNAKAN MODEL HYBRID INDOBERT DAN LSTM

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

VANESS FEBRAIO
00000055310

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Vaness Febraio
Nomor Induk Mahasiswa	:	00000055310
Program Studi	:	Informatika

Skripsi dengan judul:

Analisis Sentimen Komentar YouTube Berbahasa Indonesia terkait Isu Warga Negara Indonesia di Jepang Menggunakan Model Hybrid IndoBERT dan LSTM

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 22 Desember 2025

(Vaness Febraio)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN KOMENTAR YOUTUBE BERBAHASA INDONESIA TERKAIT ISU WARGA NEGARA INDONESIA DI JEPANG MENGGUNAKAN MODEL HYBRID INDOBERT DAN LSTM

oleh

Nama	:	Vaness Febraio
NIM	:	00000055310
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 15 Januari 2026
Pukul 08.00 s/s 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(David Agustriawan, S.Kom., M.Sc., Ph.D.)
NIDN: 0525088601

Penguji

(SY. Yuliani Yakub, S.Kom., M.T., PhD., CEH)
NIDN: 0411037904

Pembimbing

(Marlinda Vasty Overbeek, S.Kom., M.Kom)
NIDN: 0818038501

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Vaness Febraio
NIM	:	00000055310
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Analisis Sentimen Komentar YouTube Berbahasa Indonesia terkait Isu Warga Negara Indonesia di Jepang Menggunakan Model Hybrid IndoBERT dan LSTM

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
- Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 22 Desember 2025

Yang menyatakan

Vaness Febraio

HALAMAN PERSEMBAHAN / MOTTO

”I strive with all my strength, and entrust the final result to Buddha”

Vaness Febraio

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas berkat, rahmat, dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul "Analisis Sentimen Komentar YouTube Berbahasa Indonesia terkait Isu Warga Negara Indonesia di Jepang Menggunakan Model Hybrid IndoBERT dan LSTM" dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Kepada Muhammad Rajja Farabi telah mendukung penulis untuk mengerjakan skripsi ini hingga selesai.

Penulis menyadari bahwa laporan skripsi ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan untuk perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan menjadi referensi bagi pembaca.

Tangerang, 22 Desember 2025

Vaness Febraio

ANALISIS SENTIMEN KOMENTAR YOUTUBE BERBAHASA INDONESIA TERKAIT ISU WARGA NEGARA INDONESIA DI JEPANG MENGGUNAKAN MODEL HYBRID INDOBERT DAN LSTM

Vaness Febraio

ABSTRAK

Perkembangan pesat media sosial telah mengubah cara masyarakat mengekspresikan opini dan emosi terhadap berbagai isu sosial, dengan YouTube menjadi salah satu platform utama dalam diskursus publik. Salah satu isu yang menarik perhatian luas adalah keterlibatan Warga Negara Indonesia (WNI) dalam berbagai insiden di Jepang yang memicu beragam reaksi dari masyarakat Indonesia. Reaksi tersebut tercermin dalam komentar *YouTube* berbahasa Indonesia yang mengandung sentimen positif, negatif, dan netral. Penelitian ini bertujuan untuk menganalisis sentimen komentar *YouTube* terkait isu WNI di Jepang menggunakan model *hybrid IndoBERT–LSTM*. *IndoBERT*dimanfaatkan untuk menghasilkan *embedding kontekstual* yang mampu merepresentasikan makna semantik teks secara akurat, sedangkan *Long Short-Term Memory (LSTM)* digunakan untuk memodelkan urutan kata serta dependensi jangka panjang dalam kalimat. Dataset dikumpulkan dari platform YouTube dan melalui beberapa tahap praproses, meliputi *pembersihan teks*, *normalisasi*, *tokenisasi*, penghapusan *stopword*, dan *stemming*. Proses pelabelan sentimen dilakukan menggunakan *InSet Lexicon*, kemudian data dibagi menjadi data latih dan data uji dengan rasio 80:20. Hasil eksperimen menunjukkan bahwa model *IndoBERT–LSTM* mencapai akurasi sebesar 77%, dengan nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,77. Secara per kelas, sentimen negatif memperoleh *F1-score* 0,79, sentimen netral 0,71, dan sentimen positif 0,80. Selain itu, hasil analisis menunjukkan bahwa sentimen negatif mendominasi komentar, yang mencerminkan kuatnya respons emosional masyarakat terhadap isu tersebut serta menegaskan efektivitas model yang diusulkan dalam analisis sentimen teks media sosial berbahasa Indonesia.

Kata kunci: Analisis sentimen, *IndoBERT*, *LSTM*, Hibrida, *YouTube*

SENTIMENT ANALYSIS OF INDONESIAN-LANGUAGE YOUTUBE COMMENTS ON THE ISSUE OF INDONESIAN CITIZENS IN JAPAN USING THE INDOBERT AND LSTM HYBRID MODEL

Vaness Febraio

ABSTRACT

The rapid development of social media has changed the way people express opinions and emotions toward various social issues, with YouTube becoming one of the main platforms for public discourse. One issue that has attracted widespread attention is the involvement of Indonesian citizens (WNI) in various incidents in Japan, which has triggered diverse reactions from Indonesian society. These reactions are reflected in Indonesian-language YouTube comments containing positive, negative, and neutral sentiments. This study aims to analyze the sentiment of YouTube comments related to the issue of Indonesian citizens in Japan using a hybrid IndoBERT–LSTM model. IndoBERT is utilized to generate contextualized embeddings that accurately represent the semantic meaning of text, while Long Short-Term Memory (LSTM) is employed to model word sequences and long-term dependencies within sentences. The dataset was collected from YouTube and processed through several preprocessing stages, including text cleaning, normalization, tokenization, stopword removal, and stemming. Sentiment labeling was conducted using the InSet Lexicon, and the data were divided into training and testing sets with an 80:20 ratio. The experimental results show that the IndoBERT–LSTM model achieves an accuracy of 77%, with precision, recall, and F1-score values of 0.77. Class-wise analysis indicates that the negative sentiment class obtains an F1-score of 0.79, the neutral class 0.71, and the positive class 0.80. Furthermore, the results indicate that negative sentiment dominates the comments, reflecting strong emotional responses from the public and confirming the effectiveness of the proposed model for sentiment analysis of Indonesian-language social media text.

Keywords: *Sentiment analysis, IndoBERT, LSTM, Hybrid, Youtube*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR LAMPIRAN