IMPLEMENTASI ALGORITMA BERT UNTUK KLASIFIKASI TOPIK PENELITIAN PADA UNIVERSITAS MULTIMEDIA NUSANTARA

SKRIPSI

Charlie Frederico
00000043442

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

IMPLEMENTASI ALGORITMA BERT UNTUK KLASIFIKASI TOPIK PENELITIAN PADA UNIVERSITAS MULTIMEDIA NUSANTARA

SKRIPSI

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

Charlie Frederico
00000043442

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Charlie Frederico
Nomor Induk Mahasiswa	:	00000043442
Program Studi	:	Informatika

Skripsi dengan judul:
Implementasi Algoritma BERT untuk Klasifikasi Topik Penelitian di Universitas Multimedia Nusantara

merupakan hasil karya saya sendiri bukan plagiat dari karya 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 Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 22 Mei 2024

(Charlie Frederico)

ii
Implementasi Algoritma BERT..., Charlie Frederico, Universitas Multimedia Nusantara

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI ALGORITMA BERT UNTUK KLASIFIKASI TOPIK PENELITIAN PADA UNIVERSITAS MULTIMEDIA NUSANTARA

oleh

Nama	:	Charlie Frederico
NIM	:	00000043442
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 04 Juni 2024
Pukul 10.00 s/s 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

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

Penguji

(Aditiyawan, S.Komp., M.Si)
NIDN: 8994550022

Pembimbing I

(Eunike Endariahna Surbakti, S.Kom., M.T.I)
NIDN: 0322099401

Pembimbing II

(Fenina Adline Twince Tobing, S.Kom., M.Kom)
NIDN: 0406058802

Ps Ketua Program Studi Informatika,

(Dr. Eng. Niki Prastomo, S.T., M.Sc.)
NIDN: 0419128203

iii
Implementasi Algoritma BERT..., Charlie Frederico, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Charlie Frederico
NIM	:	00000043442
Program Studi	:	Informatika
Jenjang	:	S1
Jenis Karya	:	Skripsi

Menyatakan dengan sesungguhnya bahwa:

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 22 Mei 2024

Yang menyatakan

Charlie Frederico

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

”Success is not permanent and failure is not final. Never stop working after success and never stop trying after failure”

Priyanka Ghediya

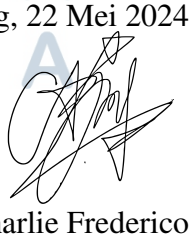
KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma BERT untuk Klasifikasi Topik Penelitian di Universitas Multimedia Nusantara dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, 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 Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Eunike Endariahna Surbakti, S.Kom., M.T.I, sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Ibu Fenina Adline Twince Tobing, S.Kom., M.Kom, sebagai Pembimbing kedua yang telah banyak membantu dan memberikan bimbingan atas terselesainya Skripsi ini.
6. Orang Tua yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 22 Mei 2024



Charlie Frederico

IMPLEMENTASI ALGORITMA BERT UNTUK KLASIFIKASI TOPIK PENELITIAN PADA UNIVERSITAS MULTIMEDIA NUSANTARA

Charlie Frederico

ABSTRAK

Terdapat lebih dari 3000 jurnal akademik dan artikel penelitian dari Universitas Multimedia Nusantara yang diterbitkan pada tahun 2018 hingga 2023. Untuk memudahkan pengkategorian, dibangun sistem klasifikasi teks yang menggunakan model *Bidirectional Encoder Representations from Transformers* (BERT) untuk mengkategorikan judul atau abstrak jurnal ke dalam 17 kategori UN SDG. UN SDG merupakan parameter utama yang digunakan untuk meningkatkan akreditasi fakultas dan program studi oleh Universitas Multimedia Nusantara. Terdapat 76.958 dataset berbahasa Inggris yang digunakan selama *training*. Model terbaik diperoleh dengan menggunakan *preprocessing* dengan *library NLTK* tanpa metode *sampling* yang menggunakan 70% *data training* dan 30% *data testing*. Parameter model mencakup 4 *epochs*, *learning rate* 2e-5, dan *batch size* 32. Model ini mendapatkan nilai *precision* 0.99, *recall* 0.82, *f1-score* 0.87, dan *akurasi* 90.68%. Model ini dipilih untuk pengklasifikasian teks karena mendapatkan hasil terbaik dibandingkan model lainnya dan telah di demonstrasikan kepada pihak LPPM UMN.

Kata kunci: *Bidirectional Encoder Representations from Transformers*, *Deep Learning*, Jurnal Akademik, Klasifikasi Teks, *Natural Language Processing*

Implementation of BERT Algorithm for Classification of Research Topics at Multimedia Nusantara University

Charlie Frederico

ABSTRACT

There are more than 3,000 academic journals and research articles from Universitas Multimedia Nusantara, published between 2018 and 2023. To facilitate categorization, a text classification system was built using the Bidirectional Encoder Representations from Transformers (BERT) model to categorize the titles or abstracts of the journals into 17 UN SDG categories. The UN SDGs are crucial for enhancing the accreditation of faculties and study programs at the university. A dataset of 76,958 English-language records was used for training. The best model, obtained using NLTK preprocessing without sampling methods, utilized 70% training data and 30% testing data. Model parameters included 4 epochs, a learning rate of 2e-5, and a batch size of 32. It achieved a precision of 0.99, recall of 0.82, f1-score of 0.87, and accuracy of 90.68%. This model is chosen for text classification because it outperforms other models and has been demonstrated to the LPPM UMN.

Keywords: Academic Journal, Bidirectional Encoder Representations from Transformers, Deep Learning, Natural Language Processing, Text Classification

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE

DAFTAR LAMPIRAN