IMPLEMENTASI NAMED ENTITY RECOGNITION UNTUK IDENTIFIKASI DIAGNOSIS MEDIS MENGGUNAKAN MODEL MULTILINGUAL BERT

SKRIPSI

STEVE CHRISTIAN Y. P.
00000058797

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

IMPLEMENTASI NAMED ENTITY RECOGNITION UNTUK IDENTIFIKASI DIAGNOSIS MEDIS MENGGUNAKAN MODEL MULTILINGUAL BERT

SKRIPSI

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

STEVE CHRISTIAN Y. P.
00000058797

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Steve Christian Y. P.
Nomor Induk Mahasiswa	:	00000058797
Program Studi	:	Informatika

Skripsi dengan judul:

Implementasi Named Entity Recognition untuk Identifikasi Diagnosis Medis pada Teks Berbahasa Indonesia Menggunakan Multilingual BERT Berbasis BIO Tagging dan FHIR

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, 4 Juli 2025

(Steve Christian Y. P.)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI NAMED ENTITY RECOGNITION UNTUK IDENTIFIKASI DIAGNOSIS MEDIS MENGGUNAKAN MODEL MULTILINGUAL BERT

oleh

Nama	:	Steve Christian Y. P.
NIM	:	00000058797
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 14 Juli 2025
Pukul 08.00 s/d 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Januar Wahjudi, S.Kom., M.Sc.)
NIDN: 0330017201

Penguji

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

Pembimbing

(Dr. Ivransa Zuhdi Pane, B. Eng., M. Eng.)
NIDN: 8812520016

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	:	Steve Christian Y. P.
NIM	:	00000058797
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Implementasi Named Entity Recognition untuk Identifikasi Diagnosis Medis pada Teks Berbahasa Indonesia Menggunakan Multilingual BERT Berbasis BIO Tagging dan FHIR

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, 4 Juli 2025

Yang menyatakan

Steve Christian Y. P.

HALAMAN PERSEMBAHAN / MOTTO

”A good name is to be more desired than great wealth, Favor is better than silver and gold.”

Proverbs 22:1 (NASB)

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Named Entity Recognition Untuk Identifikasi Diagnosis Medis menggunakan Model Multilingual BERT 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 laporan Skripsiini, sangatlah sulit bagi saya untuk menyelesaikan laporan Skripsiini. Oleh karena itu, saya mengucapkan terima kasih kepada:
Mengucapkan terima kasih

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. Bapak Dr. Ivransa Zuhdi Pane, B. Eng., M. Eng., 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.

Semoga karya ilmiah ini dapat memberikan manfaat dan kontribusi positif, baik bagi penulis sendiri, dunia pendidikan, maupun pihak-pihak lain yang membutuhkan informasi terkait. Penulis juga berharap laporan ini dapat menjadi referensi serta inspirasi untuk penelitian selanjutnya dalam bidang yang relevan.

Tangerang, 4 Juli 2025

Steve Christian Y. P.

IMPLEMENTASI NAMED ENTITY RECOGNITION UNTUK IDENTIFIKASI DIAGNOSIS MEDIS MENGGUNAKAN MODEL MULTILINGUAL BERT

Steve Christian Y. P.

ABSTRAK

Penelitian ini mengembangkan sebuah *pipeline* otomatis berbasis `Multilingual BERT` dan skema `BIO tagging` untuk ekstraksi dan identifikasi diagnosis medis dari teks berbahasa Indonesia. Proses *pipeline* mencakup *preprocessing* data, pelabelan `BIO`, pelatihan model `NER` menggunakan `mBERT`, pemetaan ke kode `SNOMED CT`, dan konversi ke *resource Fast Healthcare Interoperability Resources* (FHIR). *Dataset* terdiri dari 9.000 kalimat yang mengandung diagnosis, dengan distribusi label yang beragam. Pengujian pada data nyata dari *website* kesehatan seperti *Halodoc* dan *Alodokter* menunjukkan model mampu mencapai akurasi sebesar 93.1%, precision 91.8%, recall 92.3%, dan F1-score 92% untuk label `Condition`. *Pipeline* juga diimplementasikan sebagai layanan *API* berbasis `Flask`. Tantangan utama meliputi keterbatasan kamus `SNOMED CT` lokal dan variasi istilah diagnosis. *Pipeline* ini berpotensi mendukung digitalisasi dan interoperabilitas data klinis di Indonesia.

Kata kunci: *BIO tagging*, *Fast Healthcare Interoperability Resources*, *Multilingual BERT*, *NER*, *SNOMED CT*

IMPLEMENTATION OF NAMED ENTITY RECOGNITION FOR MEDICAL DIAGNOSIS IDENTIFICATION USING MULTILINGUAL BERT MODEL

Steve Christian Y. P.

ABSTRACT

This study develops an automatic pipeline based on Multilingual BERT and the BIO tagging scheme for extracting and identifying medical diagnoses from Indonesian-language texts. The pipeline covers data preprocessing, BIO labeling, NER model training using mBERT, mapping to SNOMED CT codes, and conversion to FHIR resources. The dataset consists of 9,000 diagnosis sentences with diverse label distribution. Testing on real-world data from health websites such as Halodoc and Alodokter shows that the model achieves 93.1% accuracy, 91.8% precision, 92.3% recall, and 92% F1-score for the Condition label. The pipeline is also implemented as an API service using Flask. Major challenges include the limited availability of a local SNOMED CT dictionary and the wide variety of diagnosis terms. This pipeline has the potential to support the digitalization and interoperability of clinical data in Indonesia.

Keywords: *BIO tagging*, *Fast Healthcare Interoperability Resources*, *Multilingual BERT*, *NER*, *SNOMED CT*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR LAMPIRAN