PERBANDINGAN KINERJA KLASIFIKASI PESAN GEJALA COVID-19 DARI PESAN SOSIAL MEDIA DENGAN BERT DAN INDOBERT

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh
ASTINA FARIDHAH
NIM. 1911016120003

PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

JUNI 2023

SKRIPSI

PERBANDINGAN KINERJA KLASIFIKASI PESAN GEJALA COVID-19 DARI PESAN SOSIAL MEDIA DENGAN BERT DAN INDOBERT

Oleh:

ASTINA FARIDHAH

NIM. 1911016120003

Telah dipertahankan di depan Dosen Penguji pada tanggal 19 Juni 2023.

Susunan Dosen Penguji:

Pembimbing I	**Dosen Penguji I**
M. Reza Faisal., S.T., M.T., Ph.D	Muhammad Itqan Mazdadi, S.Kom., M.Kom.
NIP. 197612202008121001	NIP. 199006122019031013
Pembimbing II	**Dosen Penguji II**
Andi Farmadi, S.Si., M.T	Triando Hamonangan Saragih, S.Kom., M.Kom.
NIP. 1973072520080110006	NIP. 199308242019031012

Banjarbaru, 23 Juni 2022

Koordinator Program Studi Ilmu Komputer

Irwan Budiman, S.T., M.Kom

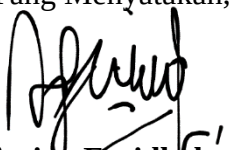
NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juni 2023

Yang Menyatakan,



Astina Faridhah

NIM. 1911016120003

ABSTRAK

PERBANDINGAN KINERJA KLASIFIKASI PESAN GEJALA COVID-19 DARI PESAN SOSIAL MEDIA DENGAN BERT DAN INDOBERT
(Oleh : Astina Faridhah; Pembimbing: Mohammad Reza Faisal, S.T., M.T. Ph.D dan Andi Farmadi, S.Si., M.T; 2023; 55 halaman)

Covid-19 merupakan jenis penyakit yang disebabkan oleh virus Corona yang muncul pada akhir tahun 2019 di Wuhan kemudian diumumkan pertama kali kasus Covid-19 di Indonesia pada Maret 2020. Selama pandemi ini, media sosial, khususnya twitter, telah menjadi salah satu *platform* yang paling banyak digunakan untuk berbagi informasi tentang Covid-19. Twitter merupakan salah satu wadah yang digunakan masyarakat untuk dapat saling berbagi informasi mengenai tanda-tanda maupun dampak dari virus Covid-19. Twitter menjadi *platform* populer untuk berbagi informasi tentang Covid-19, termasuk laporan mandiri dari individu yang mengalami gejala Covid-19 atau telah terinfeksi virus. Model *pre-trained Bidirectional Encoder Representations from Transformers* (BERT) merupakan model yang sedang popular di kalangan peneliti. Pada penelitian ini pesan gejala Covid-19 dari sosial media dianalisis menggunakan *pre-trained model* BERT dan IndoBERT. Diperoleh hasil akurasi BERT dengan dataset asli (tanpa *preprocessing*) sebesar 81,5%, akurasi BERT dengan dataset hasil *preprocessing* sebesar 82,0%, dan IndoBERT menggunakan dataset asli (tanpa *preprocessing*) akurasi sebesar 89,5%, serta IndoBERT menggunakan dataset hasil *preprocessing* akurasi sebesar 94,0%.

Kata kunci: Klasifikasi Teks, Gejala Covid-19, Twitter, BERT, IndoBERT

ABSTRACT

PERFORMANCE COMPARISON OF COVID-19 SYMPTOM MESSAGE CLASSIFICATION FROM SOCIAL MEDIA MESSAGES WITH BERT AND INDOBERT

(By: Astina Faridhah; Supervisor: Mohammad Reza Faisal, S.T., M.T., Ph.D and Andi Farmadi, S.Si., M.T;2023;55 pages)

Covid-19 is a type of disease caused by the Corona virus that emerged at the end of 2019 in Wuhan and then the first announced Covid-19 case in Indonesia in March 2020. During this pandemic, social media, especially twitter, has become one of the most used platforms to share information about Covid-19. Twitter is one of the platforms used by the community to share information about the signs and impacts of the Covid-19 virus. Twitter has become a popular platform for sharing information about Covid-19, including self-reports from individuals who experience symptoms of Covid-19 or have been infected with the virus. The pre-trained Bidirectional Encoder Representations from Transformers (BERT) model is a model that is currently popular among researchers. In this study, Covid-19 symptom messages from social media were analyzed using the pre-trained BERT and IndoBERT models. The BERT accuracy result with the original dataset (without preprocessing) was 81.5%, BERT accuracy with the preprocessed dataset was 82.0%, and IndoBERT using the original dataset (without preprocessing) accuracy was 89.5%, and IndoBERT using the preprocessed dataset accuracy was 94.0%.

Keywords: Text Classification, Covid-19 Symptoms, Twitter, BERT, IndoBERT

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **PERBANDINGAN KINERJA KLASIFIKASI PESAN GEJALA COVID-19 DARI PESAN SOSIAL MEDIA DENGAN BERT DAN INDOBERT** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Mohammad Reza Faisal, S.T., M.T. Ph.D selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Andi Farmadi, S.Si., M.T selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2023

Astina Faridhah

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN