

SKRIPSI
IMPLEMENTASI ALGORITMA BERT UNTUK ANALISIS SENTIMEN
ULASAN PENGGUNA APLIKASI PEDULILINDUNGI



SISKA RAHMAWATI

NIM : 185410089

PROGRAM STUDI INFORMATIKA
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2023

SKRIPSI
IMPLEMENTASI ALGORITMA BERT UNTUK ANALISIS SENTIMEN
ULASAN PENGGUNA APLIKASI PEDULILINDUNGI

Diajukan sebagai salah satu syarat untuk menyelesaikan studi



Program Sarjana
Program Studi Informatika
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Disusun Oleh
SISKA RAHMAWATI
NIM : 185410089

PROGRAM STUDI INFORMATIKA
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2023

PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 03 Maret 2023



Siska Rahmawati

NIM: 185410089

HALAMAN PERSEMBAHAN

Puji syukur kehadiran Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya yang dilimpahkan kepada kita semua. Dengan rasa syukur yang luar biasa, saya persembahkan skripsi ini kepada:

1. Kedua orang tua tercinta yang senantiasa memberikan semangat, kasih sayang, dan doa terbaik sehingga dapat diberi kemudahan dalam menyelesaikan skripsi ini.
2. Keluarga yang selalu memberikan dukungan dalam pengerjaan skripsi ini.
3. Ibu Maria Mediatrix Sebatubun S.Kom., M.Eng. selaku dosen pembimbing yang dengan sabar memberikan bimbingan terhadap saya.
4. Bapak Dr., Bambang P.D.P., S.E., Akt., S.Kom., MMSI. selaku dosen pembimbing akademik yang selalu memberi semangat dan membuat saya yakin dalam mengambil mata kuliah setiap semester.
5. Serta teman-teman yang selalu memberikan dukungan.

HALAMAN MOTTO

Remember to always believe in yourself and your abilities. Keep pushing yourself to achieve your goals, and don't let obstacles stand in your way. With determination and hard work, you can achieve anything you set your mind to.

Best of luck!

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Skripsi ini yang berjudul **“Implementasi Algoritma BERT Untuk Analisis Sentimen Ulasan Pengguna Aplikasi PeduliLindungi”**. Skripsi ini ditujukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer bagi mahasiswa program S1 program studi Informatika di Universitas Teknologi Digital Indonesia.

Penulis menyadari bahwa dalam menyelesaikan Skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan banyak terimakasih kepada:

1. Kedua orang tua dan keluarga yang selalu memberikan motivasi dan dukungan dalam menyelesaikan Skripsi ini.
2. Bapak Ir. Totok Suprawoto M.M., M.T. selaku Rektor Universitas Teknologi Digital Indonesia.
3. Bapak Ir. Muhammad Guntara, M.T. selaku Dekan Fakultas Teknologi Informasi.
4. Ibu Dini Fakta Sari S.T., M.T. selaku Ketua Prodi Informatika Universitas Teknologi Digital Indonesia.
5. Bapak Dr., Bambang P.D.P., S.E, Akt., S.Kom, MMSI. selaku dosen pembimbing akademik.
6. Ibu Maria Mediatrice Sebatubun S.Kom., M.Eng. selaku dosen pembimbing skripsi yang telah banyak membimbing, memberikan pengarahan, serta saran-saran dalam penyusunan tugas akhir ini..

7. Seluruh dosen dan staff karyawan Universitas Teknologi Digital Indonesia.
8. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung yang tidak bisa saya sebutkan satu persatu.

Penulis menyadari sepenuhnya, bahwa penulisan tugas akhir ini masih banyak kekurangan, baik dalam hal penulisan maupun cara penyajian materi. Untuk itu, kritik dan saran yang membangun sangatlah penulis harapkan guna menambah wawasan dan pengembangan ilmu yang telah penulis peroleh selama ini. Semoga hasil dari karya tulis ini dapat bermanfaat bagi semua pihak.

Yogyakarta, 03 Maret 2023



Siska Rahmawati

NIM : 185410089

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6

2.1 Tinjauan Pustaka	6
2.2 Dasar Teori.....	8
2.2.1 PeduliLindungi.....	8
2.2.2 Google Play Store	9
2.2.3 Analisis Sentimen	10
2.2.4 Text Mining.....	11
2.2.5 Machine Learning	12
2.2.6 VADER.....	13
2.2.7 BERT (Bidirectional Encoder Representation from Transformers)	13
2.2.8 BERT-Base Multilingual	18
BAB III METODE PENELITIAN.....	20
3.1 Bahan / Data.....	20
3.1.1 Kebutuhan Input.....	20
3.1.2 Kebutuhan Proses.....	20
3.1.3 Kebutuhan Output	21
3.2 Peralatan.....	21
3.2.1 Kebutuhan Perangkat Lunak	21
3.2.2 Kebutuhan Perangkat Keras.....	21
3.3 Rancangan Sistem	22
3.3.1 Pengumpulan Data	23
3.3.2 Text Preprocessing.....	23
3.3.3 Pelabelan Data.....	26
3.3.4 Split Dataset	27
3.3.5 Modelling	27
3.3.6 Evaluation	28
BAB IV IMPLEMENTASI DAN PEMBAHASAN SISTEM	30
4.1 Dataset.....	30
4.2 Implementasi Model BERT base multilingual.....	32

4.3 Uji Coba dan Pembahasan Sistem.....	33
4.3.1 Hasil Uji Coba.....	34
4.4 Implementasi Web	35
BAB V.....	38
5.1 Kesimpulan	38
5.2 Saran.....	38
DAFTAR PUSTAKA	40
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Transformer (Vaswani et al., 2017)	14
Gambar 2. 2 Representasi Input BERT (Devlin et al., 2019)	17
Gambar 2. 3 Paradigma Pelatihan BERT (Devlin et al., 2019)	18
Gambar 3. 1 Diagram Blok Sistem	22
Gambar 4. 1 Sebaran Data Berdasarkan Rating	30
Gambar 4. 2 Sebaran Label	31
Gambar 4. 3 Jumlah Dataset Pelatihan	32
Gambar 4. 4 Jumlah Dataset Pelatihan Setelah Random Oversampling	32
Gambar 4. 5 Grafik akurasi training dan validation	33
Gambar 4. 6 Performa Model	33
Gambar 4. 7 Diagram Confusion Matrix	34
Gambar 4. 8 Input Teks dan Hasil Klasifikasi Netral	35
Gambar 4. 9 Input Teks dan Hasil Klasifikasi Positif	36
Gambar 4. 10 Input Teks dan Hasil Klasifikasi Negatif	36
Gambar 4. 11 Tampilan Grafik Akurasi	37
Gambar 4. 12 Tampilan Hasil Prediksi Dataset Testing	37

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	7
Tabel 3. 1 Hasil Lowercase	24
Tabel 3. 2 Hasil Data Cleaning	24
Tabel 3. 3 Hasil Tokenisasi	25
Tabel 3. 4 Hasil Stopword Removal	25
Tabel 3. 5 Hasil Stemming	25
Tabel 3. 6 Hasil Normalisasi	26
Tabel 3. 7 Hyperparameter Fine Tuning Model bert-base-multilingual-cased.....	27
Tabel 3. 8 Confusion Matrix 3x3	28
Tabel 4. 2 Perbandingan Akurasi	32
Tabel 4. 3 Hasil Prediksi Model Pada Data Acak.....	34

INTISARI

Pandemi COVID19 yang terjadi di Indonesia mengharuskan masyarakat dan pemerintah menerapkan kebiasaan baru. Salah satu upaya pemerintah dalam menangani penyebaran COVID19 di Indonesia yaitu mengembangkan aplikasi PeduliLindungi yang dapat diunduh melalui Google Play. Pada tahun 2020, aplikasi ini telah diunduh lebih dari 1 juta kali dengan jumlah ulasan lebih dari 43 ribu. Ulasan merupakan salah satu aspek penting dalam pengembangan aplikasi, karena pengguna dapat memberikan kritik dan saran tentang pengalaman menggunakan aplikasi dan membantu pengembang untuk mengembangkan layanan sesuai harapan pengguna sehingga dapat memberikan pengalaman yang lebih baik pada masa yang akan datang.

Analisis sentimen dapat digunakan untuk mengkategorikan ulasan pada suatu polaritas dimana ulasan dapat bernilai positif, negatif, atau netral. Dengan berkembangnya teknologi, analisis sentimen dapat dilakukan secara otomatis. Penelitian ini menggunakan algoritma BERT (*Bidirectional Encoder Representations from Transformers*) untuk membuat analisis sentimen pada ulasan aplikasi PeduliLindungi. Pemodelan analisis sentimen menggunakan BERT yaitu pre-trained BERT *base-multilingual-cased*.

Jumlah data yang digunakan sebanyak 13.574 data berbahasa Indonesia dan bahasa Inggris. Data dibagi menggunakan rasio 80:10:10 dimana 80% untuk data latih yaitu 10589 data, 10% untuk data validasi yaitu 1357 data, dan 10% untuk data uji yaitu 1358 data. Penelitian ini menghasilkan tingkat akurasi sebesar 89% dengan nilai presisi pada sentimen negatif yaitu 70%, netral 91%, dan positif sebesar 91%.

Kata kunci : analisis sentimen, bert, google play, ulasan aplikasi pedulilindungi

ABSTRACT

The COVID19 pandemic that occurred in Indonesia forced the community and government to adopt new habits. One of the government's efforts to deal with the spread of COVID-19 in Indonesia is to develop the PeduliLindungi application which can be downloaded via Google Play. In 2020, this app has been downloaded more than 1 million times with more than 43 thousand reviews. Reviews are an important aspect of application development, because they can provide criticism and suggestions about the experience of using the application and help developers to develop services according to user expectations so that they can provide a better experience in the future.

Sentiment analysis can be used to categorize reviews on a polarity where reviews can be positive, negative or neutral. With the development of sentiment analysis technology can be done automatically. This study uses the BERT (Bidirectional Encoder Representations from Transformers) algorithm to make a sentiment analysis on the PeduliLindungi application review. Sentiment analysis modeling uses BERT, namely BERT base-multilingual-cased which has been trained before.

The amount of data used is 13,574 data in Indonesian and English. The data is divided using an 80:10:10 ratio where 80% for training data is 10589, 10% for data validation is 1357 data, and 10% for test data is 1358 data. This study produces an accuracy rate of 89% with precision score to negative sentiment is 70%, neutral is 91%, and positive 91%.

Keywords:, bert, google play, pedulilindungi app reviews, sentiment analysis