

**ANALISIS SENTIMEN PADA ULASAN APLIKASI DI
GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA
ROBERTA-BILSTM
(STUDI KASUS PADA APLIKASI MYPERTAMINA)**

TUGAS AKHIR

Oleh:

SEAN RICHARD SINDUNATA

NIM.1910817210020



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS LAMBUNG MANGKURAT
BANJARMASIN**

2023

**ANALISIS SENTIMEN PADA ULASAN APLIKASI DI
GOOGLE PLAYSTORE MENGGUNAKAN ALGORITMA
ROBERTA-BILSTM
(STUDI KASUS PADA APLIKASI MYPERTAMINA)**

Tugas Akhir

Diajukan Untuk Memenuhi Salah Satu Syarat
Sarjana Strata-1 Teknologi Informasi

Oleh:

SEAN RICHARD SINDUNATA

NIM.1910817210020



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS LAMBUNG MANGKURAT
BANJARMASIN, JULI 2023**

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Sean Richard Sindunata
NIM : 1910817210020
Fakultas : Teknik
Program Studi : Teknologi Informasi
Judul Tugas Akhir : Analisis Sentimen pada Ulasan Aplikasi di Google Playstore Menggunakan Algoritma RoBERTa-BiLSTM (Studi Kasus pada Aplikasi MyPertamina)
Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom.

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, Juli 2023



Sean Richard Sindunata

NIM. 1910817210020

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI

Analisis Sentimen pada Ulasan Aplikasi di *Google Play Store*

Menggunakan Algoritma RoBERTa-BiLSTM

(Studi Kasus pada Aplikasi MyPertamina)

oleh

Sean Richard Sindunata (1910817210020)

Telah dipertahankan di depan Tim Penguji pada 26 Juni 2023 dan dinyatakan

LULUS

Komite Penguji :

Ketua : Eka Setya Wijaya S.T., M.Kom
NIP 198205082008011010

Anggota 1 : Dr. Ir. Yuslena Sari, S.Kom., M.Kom
NIP 198411202015042002

Anggota 2 : Andry Fajar Zulkarnain, S.ST., M.T
NIP 199007272019031018

Pembimbing : Andreyan Rizky Baskara, S.Kom., M.Kom
Utama NIP 199307032019031011

Banjarbaru, 4 Juni 2023
diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik
Fakultas Teknik ULM,

Koordinator Program Studi
S-1 Teknologi Informasi,



Dr. S.T., M.T.
NIP 1998021001

Dr. Ir. Yuslena Sari, S.Kom., M.Kom
NIP 198411202015042002

ABSTRAK

Dalam evaluasi *usability*, tahap analisis sentimen diperlukan agar data dari pengguna dapat dipahami dengan lebih baik. Analisis sentimen digunakan untuk menemukan informasi berharga yang dibutuhkan dari data yang tidak terstruktur, seperti pada data ulasan pengguna. Konsep *Transformers* telah menjadi salah satu konsep model NLP yang sering digunakan pada analisis sentimen dan memiliki tingkat akurasi yang cukup tinggi. Salah satu model yang menerapkan konsep ini adalah model RoBERTa. RoBERTa merupakan model yang terinspirasi dari BERT yang dioptimisasi. Agar bisa meningkatkan akurasi dari model RoBERTa, penelitian ini mengusulkan tambahan layer BiLSTM pada model RoBERTa karena model RoBERTa bisa memberikan pemahaman bahasa yang lebih mendalam dan tambahan layer BiLSTM dapat menangkap informasi kontekstual yang lebih kaya dalam sebuah teks, karena model BiLSTM menggunakan 2 layer yang bergerak maju dan mundur (*Forward* dan *Backward*). Dataset yang digunakan pada penelitian ini adalah ulasan aplikasi MyPertamina pada *Google Play Store*. Dalam penelitian ini dilakukan eksperimen terhadap beberapa *hyperparameter* pada model untuk mendapatkan model yang optimal dalam analisis sentimen data ulasan berbahasa Indonesia. Hasil akhir eksperimen menunjukkan bahwa dengan tambahan layer BiLSTM, model yang diusulkan berhasil meningkatkan akurasi model sebesar 1,76% dibanding model RoBERTa biasa. Berdasarkan hasil laporan klasifikasi model ini juga mendapatkan akurasi sebesar 91,02%, dengan rata – rata berbobot pada *precision* 91,06%, *recall* 91,02%, dan *F1-Score* 91,03%.

Kata Kunci : Analisis Sentimen, BiLSTM, RoBERTa, Transformers, Ulasan Bahasa Indonesia

ABSTRACT

In usability evaluation, a sentiment analysis stage is required to better understand the data from users. Sentiment analysis is used to find valuable information needed from unstructured data, such as user review data. The Transformers concept has become one of the NLP model concepts that is often used in sentiment analysis and has a fairly high accuracy rate. One model that applies this concept is the RoBERTa model. RoBERTa is a BERT-inspired model that is optimized. In order to improve the accuracy of the RoBERTa model, this research proposes an additional BiLSTM layer in the RoBERTa model because the RoBERTa model can provide a deeper understanding of language and the additional BiLSTM layer can capture richer contextual information in a text, because the BiLSTM model uses 2 layers that move forward and backward (Forward and Backward). The dataset used in this research is the MyPertamina application review on the Google Play Store. In this study, experiments were conducted on several hyperparameters in the model to get the optimal model in sentiment analysis of Indonesian language review data. The final results of the experiment showed that with the addition of the BiLSTM layer, the proposed model succeeded in increasing the accuracy of the model by 1.76% compared to the regular RoBERTa model. Based on the classification report results, this model also gets an accuracy of 91.02%, with a weighted average of precision 91.06%, recall 91.02%, and F1-Score 91.03%.

Keywords: *BiLSTM, Indonesian Review, RoBERTa, Sentiment Analysis, Transformers.*

KATA PENGANTAR

Dengan mengucapkan puji dan syukur sembari berdoa kepada kehadiran Tuhan Yang Maha Esa atas segala kerendahan hati dan keikhlasan serta ketulusan atas Rahmat dan Anugerah-Nya lah sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul: **“Analisis Sentimen pada Ulasan Aplikasi di Google Playstore Menggunakan Algoritma RoBERTa-BiLSTM (Studi Kasus pada Aplikasi MyPertamina)”**.

Selama pengerjaan laporan ini, penulis mendapatkan banyak arahan, bantuan, saran, dan kritik sehingga penyusunan laporan ini dapat berjalan dengan lancar sampai pada tahap penyelesaian. Laporan ini disusun dalam rangka memenuhi salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknologi Informasi di Fakultas Teknik Universitas Lambung Mangkurat, Banjarmasin. Dalam kesempatan ini penulis menyampaikan terimakasih yang sebesar-besarnya kepada:

1. Ibu Dr. Ir. Yuslena Sari, S.Kom., M.Kom selaku Koordinator Program Studi Teknologi Informasi yang telah memberikan arahan dan solusi dalam penyelesaian Tugas Akhir.
2. Bapak Andreyan Rizky Baskara, S.Kom., M.Kom. selaku Dosen Pembimbing Utama yang telah banyak membantu dan memberikan bimbingan kepada peneliti selama melaksanakan penelitian dalam proses penyelesaian Tugas Akhir.
3. Dosen-dosen beserta staff di Program Studi Teknologi Informasi yang telah mengarahkan dan teman-teman yang membantu dalam proses penyelesaian Tugas Akhir.
4. Orang tua, keluarga, dan teman – teman yang telah membantu dan memberikan motivasi demi mendukung kelancaran penyelesaian Tugas Akhir ini.

Akhir kata, penulis menyampaikan terimakasih kepada semua pihak yang turut membantu dalam penyelesaian laporan Tugas Akhir ini. Harapan yang paling besar dari penyusunan laporan ini adalah, semoga apa yang penulis susun penuh manfaat, baik untuk pribadi, teman-teman, serta pembaca. Penulis juga mengharapakan saran dan kritik demi perbaikan dan penyempurnaan laporan ini.

Semoga laporan ini dapat bermanfaat bagi para pembaca dan semua pihak yang membutuhkan.

Banjarmasin, Juli 2023

Penulis,

A handwritten signature in black ink, consisting of a large, stylized 'S' followed by a horizontal line that tapers to the right.

Sean Richard Sindunata

DAFTAR ISI

HALAMAN SAMPUL	i
SAMPUL DALAM.....	ii
LEMBAR PERNYATAAN	iii
LEMBAR PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xv
DAFTAR RIWAYAT HIDUP.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.5.1 Bagi peneliti	5
1.5.2 Bagi pengembang aplikasi	5
1.5.3 Bagi peneliti berikutnya	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Landasan Teori.....	6
2.1.1 Analisis Sentimen.....	6
2.1.2 MyPertamina.....	6
2.1.3 <i>Natural Language Processing</i> (NLP)	7
2.1.4 <i>Transformer</i>	7
2.1.5 <i>Transfer learning</i>	9
2.1.6 Model <i>Pre-trained</i>	9
2.1.7 <i>Text Pre-processing</i>	9
2.1.8 <i>Bidirectional Encoder Representation from Transformers</i> (BERT)...	10

2.1.9	<i>Robustly Optimized BERT Pretraining Approach (RoBERTa)</i>	11
2.1.10	<i>Long Short-Term Memory (LSTM)</i>	12
2.1.11	<i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	13
2.1.12	<i>Weighted Average</i>	15
2.2	Penelitian Terkait	15
2.2.1	<i>Transformer-based deep learning models for the sentiment analysis of social media data</i>	15
2.2.2	<i>Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model</i>	16
2.2.3	<i>A Comparative Study on COVID-19 Fake News Detection Using Different Transformer Based Models</i>	17
2.2.4	<i>BERT- and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords</i>	17
2.2.5	Analisis Sentimen Pendapat Masyarakat Mengenai Vaksin Covid-19 Pada Media Sosial Twitter dengan <i>Robustly Optimized BERT Pretraining Approach</i>	18
2.2.6	<i>Sentiment Analysis With Ensemble Hybrid Deep Learning Model</i>	19
2.3	Kerangka Pemikiran.....	26
2.3.1	<i>Indicators</i>	26
2.3.2	<i>Proposed Method</i>	26
2.3.3	<i>Objectives</i>	26
2.3.4	<i>Measurement</i>	26
BAB III METODOLOGI PENELITIAN.....		27
3.1	Alat dan Bahan.....	27
3.1.1	Alat Penelitian.....	27
3.1.2	Bahan Penelitian.....	27
3.2	Alur Penelitian	28
3.2.1	Identifikasi Masalah.....	28
3.2.2	Studi Literatur	29
3.2.3	Pengumpulan Data	29
3.2.4	Implementasi Model.....	30
3.2.4.1	Text Pre-processing.....	32

3.2.4.2	Input Data.....	33
3.2.4.3	Pembagian Data	33
3.2.4.4	Pembuatan Model.....	34
3.2.4.5	Pengujian Model	34
3.2.4.6	Pengukuran performa model.....	35
3.2.5	Pengujian.....	35
3.2.6	Pembagian data menurut sentimen.....	35
3.2.7	Pengukuran Performa.....	35
3.2.8	Perbandingan Model	36
3.2.9	Implementasi Sistem	36
BAB IV	HASIL DAN PEMBAHASAN	37
4.1	Hasil	37
4.1.1	Dataset.....	37
4.1.2	Pengumpulan dan Pemilahan Data.....	37
4.1.3	<i>Preprocessing Text</i> dan Pelabelan Data.....	38
4.1.4	Implementasi Model.....	39
4.1.4.1	Konfigurasi awal RoBERTa-BiLSTM.....	39
4.1.4.2	Eksperimen Batch Size	39
4.1.4.3	Eksperimen Dropout	42
4.1.4.4	Eksperimen jumlah unit LSTM.....	44
4.1.4.5	Eksperimen Epoch	47
4.1.5	Evaluasi Model.....	49
4.2	Pembahasan.....	51
4.3	Implementasi Sistem	53
BAB V	KESIMPULAN DAN SARAN.....	57
5.1	Kesimpulan	57
5.2	Saran.....	57
DAFTAR	PUSTAKA	58
LAMPIRAN	62

DAFTAR TABEL

Tabel 2.1 Hasil pengukuran performa model BERT-BiLSTM.....	18
Tabel 2.2 Hasil evaluasi pada dataset IMDB	20
Tabel 2.3 Hasil evaluasi pada dataset maskapai penerbangan AS	20
Tabel 2.4 Hasil evaluasi pada dataset Sentiment140	21
Tabel 2.5 Ringkasan penelitian terkait.....	22
Tabel 3.1 Alat Penelitian.....	27
Tabel 3.2 Contoh data	29
Tabel 3.3 Proses Case Folding.....	32
Tabel 3.4 Proses Filtering	32
Tabel 3.5 Proses Tokenization	33
Tabel 3.6 Proses Lemmatization	33
Tabel 3.7 Contoh pembagian sentimen.....	35
Tabel 4.1 Contoh hasil pemilahan data	37
Tabel 4.2 Contoh data hasil preprocessing dan labeling.....	38
Tabel 4.3 Jumlah data tiap kelas label	38
Tabel 4.4 Pengaturan konfigurasi awal sebelum eksperimen	39
Tabel 4.5 Hasil performa model berdasarkan nilai Batch Size	42
Tabel 4.6 Hasil performa model berdasarkan nilai dropout.....	44
Tabel 4.7 Hasil performa model berdasarkan jumlah unit LSTM	47
Tabel 4.8 Hasil performa model berdasarkan jumlah epoch.....	49
Tabel 4.9 Konfigurasi model paling optimal berdasarkan eksperimen.....	49
Tabel 4.10 Penilaian model RoBERTa	50
Tabel 4.11 Penilaian model RoBERTa-BiLSTM	50
Tabel 4.12 Hasil Rincian Prediksi Data	51
Tabel 4.13 Contoh data prediksi	52
Tabel 4.14 Kata yang sering muncul pada data training.....	52

DAFTAR GAMBAR

Gambar 2.1 Arsitektur BERT [17].....	10
Gambar 2.2 Arsitektur LSTM [19]	12
Gambar 2.3 Arsitektur BiLSTM [20].....	14
Gambar 2.4 Kerangka Pemikiran.....	26
Gambar 3.1 Alur Penelitian.....	28
Gambar 3.2 Alur proses implementasi model.....	31
Gambar 3.3 Penerapan layer BiLSTM pada RoBERTa.....	34
Gambar 3.4 Rancangan prototype sistem.....	36
Gambar 4.1 Word cloud pada label positif (atas), netral (kiri), dan negatif (kanan)	39
Gambar 4.2 Grafik validation accuracy dan validation loss pada batchsize = 4 ..	40
Gambar 4.3 Grafik validation accuracy dan validation loss pada batchsize = 8 ..	40
Gambar 4.4 Grafik validation accuracy dan validation loss pada batchsize = 16	41
Gambar 4.5 Grafik validation accuracy dan validation loss pada batchsize = 32	41
Gambar 4.6 Grafik validation accuracy dan validation loss pada batchsize = 64	41
Gambar 4.7 Grafik validation loss (kiri) dan validation accuracy (kanan) tanpa layer dropout	42
Gambar 4.8 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan dropout = 0.2	43
Gambar 4.9 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan dropout = 0.4	43
Gambar 4.10 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan dropout = 0.6	43
Gambar 4.11 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan dropout = 0.8	44
Gambar 4.12 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 64 LSTM unit.....	45
Gambar 4.13 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 128 LSTM unit.....	45
Gambar 4.14 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 256 LSTM unit.....	46

Gambar 4.15 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 512 LSTM unit.....	46
Gambar 4.16 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 1024 LSTM unit.....	46
Gambar 4.17 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 2048 LSTM unit.....	47
Gambar 4.18 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 5 epoch	48
Gambar 4.19 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 10 epoch	48
Gambar 4.20 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 15 epoch	48
Gambar 4.21 Grafik validation loss (kiri) dan validation accuracy (kanan) dengan 20 epoch	49
Gambar 4.22 Struktur model RoBERTa-BiLSTM yang dibangun.....	51
Gambar 4.23 Tampilan sistem analisis sentimen.....	53
Gambar 4.24 Tampilan sistem saat prediksi bernilai “Positive”.....	54
Gambar 4.25 Tampilan sistem saat prediksi bernilai “Negative”	55
Gambar 4.26 Tampilan sistem saat prediksi bernilai “Neutral”	55

DAFTAR LAMPIRAN

Lampiran 1. Lembar Pernyataan Kesediaan Membimbing Tugas Akhir	63
Lampiran 2. Lembar Konsultasi.....	64
Lampiran 3. Kode model RoBERTa-BiLSTM	65
Lampiran 4. Kode Preprocessing Text.....	69