

HALAMAN MOTTO

"Kegagalan hanya terjadi bila kita menyerah."

"Belajarlal mengucap syukur dari hal-hal baik di hidupmu. Belajarlal menjadi kuat dari hal-hal buruk di hidupmu."

(BJ Habibie)



HALAMAN PERSEMBAHAN

Dengan mengucapkan puji syukur kehadiran Allah Swt. Atas segala rahmat, nikmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini. Dengan penuh rasa syukur dan hormat, karya ini penulis persembahkan kepada:

1. Allah Yang Maha Kuasa, atas segala kemudahan dan petunjuk-Nya dalam setiap kata dan baris yang ditulis penyusun.
2. Kedua orangtua tercinta, Ayah Rasdi, dan Mamah Sri Lestari, dan keluarga besar penulis. Keringat, perjuangan, dan kasih sayang yang telah diberikan tidak akan pernah mampu penulis balas sepenuhnya.
3. Diri saya sendiri, Radita Prinadi, sebagai penulis mengapresiasi upaya yang telah dilakukan dalam menghadapi berbagai tantangan selama proses penyusunan skripsi ini.
4. Seseorang yang selalu menemani penulis dan sebagai tempat untuk bercerita Saudari Nida Fadilah.
5. Seluruh civitas akademika Universitas Putra Bangsa, yang telah menjadi bagian dari perjalanan penulis dalam menimba ilmu hingga penyelesaian skripsi ini.
6. Teman-teman seperjuangan S1 Sains Data angkatan 2022 yang telah menjadi keluarga penulis.

ABSTRAK

Perkembangan TikTok sebagai media sosial dan platform perdagangan digital mendorong meningkatnya interaksi pengguna melalui komentar pada konten review produk. Komentar tersebut dapat mengandung berbagai sentimen yang dapat digunakan untuk mengetahui tanggapan pengguna terhadap suatu produk. Penelitian ini bertujuan untuk menganalisis sentimen komentar pada video review produk di TikTok serta membandingkan kinerja metode *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM), dan *IndoBERT*. Data yang digunakan sebanyak 1.670 komentar yang diberi label menggunakan metode *Lexicon-Based* dengan kamus InSet, kemudian diklasifikasikan ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Pengujian dilakukan dengan pembagian data 80:20 menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa sentimen positif merupakan sentimen yang paling dominan dengan 901 komentar (53,95%), diikuti netral sebanyak 422 komentar (25,27%) dan negatif sebanyak 347 komentar (20,78%). Berdasarkan hasil pengujian, *IndoBERT* memperoleh *accuracy* tertinggi sebesar 78,44%, sedangkan SVM memperoleh *accuracy* sebesar 73,35%, LSTM sebesar 77,84%. Dengan demikian, *IndoBERT* memiliki kinerja terbaik berdasarkan nilai *accuracy* pada penelitian ini.

Kata kunci: Analisis Sentimen, TikTok, SVM, LSTM, *IndoBERT*.

ABSTRACT

The development of TikTok as a social media platform and digital commerce platform has led to increased user interaction through comments on product review content. These comments may contain various sentiments that can be used to identify users' responses to a product. This study aims to analyze the sentiment of comments on product review videos on TikTok and compare the performance of the Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and IndoBERT methods. The data consisted of 1,670 comments labeled using the Lexicon-Based method with the InSet lexicon and classified into three sentiment classes: negative, neutral, and positive. The evaluation was conducted using an 80:20 data split with accuracy, precision, recall, and F1-score as evaluation metrics. The results showed that positive sentiment was the most dominant, with 901 comments (53.95%), followed by neutral sentiment with 422 comments (25.27%) and negative sentiment with 347 comments (20.78%). Based on the evaluation results, IndoBERT achieved the highest accuracy of 78.44%, while SVM achieved an accuracy of 73.35% and LSTM achieved 77.84%. Therefore, IndoBERT demonstrated the best performance based on accuracy in this study.

Keywords: Sentiment Analysis, TikTok, SVM, LSTM, IndoBERT.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas rahmat, taufik, dan hidayah-Nya, penulis dapat menyelesaikan skripsi dengan judul “Sentimen Analisis Komentar Video *Review* Produk Tiktok Dengan Perbandingan Kinerja SVM, LSTM, Dan *IndoBERT*” dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi yang penulis tempuh di Universitas Putra Bangsa.

Dalam proses penyusunannya, penulis menghadapi berbagai tantangan yang menjadi bagian dari proses pembelajaran. Melalui usaha, kesabaran, serta dukungan dari berbagai pihak, skripsi ini akhirnya dapat diselesaikan. Penulis berharap hasil penelitian ini dapat memberikan manfaat serta menjadi referensi bagi pengembangan ilmu pengetahuan. Skripsi ini masih jauh dari kata sempurna.

Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan penyelesaian skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Dr. Gunarso Wiwoho, S.E., M.M. selaku Rektor Universitas Putra Bangsa.
2. Bapak Yulianto, S.Kom., M.Kom., selaku Dekan Fakultas Sains dan Teknologi Universitas Putra Bangsa.
3. Bapak Sarjimin, S.Kom., M.Kom. selaku Ketua Prodi Sains Data Universitas Putra Bangsa.

4. Ibu Lolanda Hamim Annisa, S.Kom., M.Kom. selaku dosen pembimbing, terima kasih yang telah memberi arahan, bimbingan, kesabaran, waktu yang diluangkan, dan ilmu yang telah diberikan selama proses penyusunan skripsi ini.
5. Segenap dosen Fakultas Sains dan Teknologi Universitas Putra Bangsa, khususnya dosen Program Studi Sains Data, yang telah memberikan ilmu, bimbingan, dan pengalaman selama penulis menempuh pendidikan. Semoga setiap ilmu yang telah diberikan menjadi amal jariyah yang terus mengalir, dan Allah SWT membalas segala jasa serta kebaikan dengan balasan terbaik.
6. Kepada semua pihak yang tidak dapat disebutkan satu per satu, penulis mengucapkan terima kasih atas segala doa, dukungan, dan kebaikan yang telah diberikan. Semoga karya sederhana ini dapat memberikan manfaat, menambah wawasan, serta menjadi langkah awal dalam meraih masa depan yang lebih baik.

DAFTAR ISI

HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN BEBAS PLAGIARISME	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	7
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	9
1.5 Manfaat Penelitian	10
1.5.1 Manfaat Teoritis.....	10
1.5.2 Manfaat Praktis	10
BAB II KAJIAN PUSTAKA.....	11
2.1 Tinjauan Teori.....	11
2.1.1 Analisis Sentimen	11
2.1.2 Media Sosial.....	12
2.1.3 <i>Text Mining</i>	13
2.1.4 Google Colab	13

2.1.5 <i>Natural Language Processing</i>	14
2.1.6 <i>Support Vector Machine</i>	15
2.1.7 <i>Long Short-Term Memory</i>	17
2.1.8 <i>IndoBERT</i>	21
2.2 Penelitian Terdahulu	24
2.3 Kerangka Pemikiran.....	30
2.4 Keaslian Penelitian.....	33
BAB III METODOLOGI PENELITIAN.....	38
3.1 Tahapan Penelitian.....	38
3.2 Objek dan Subjek Penelitian	42
3.2.1 Objek Penelitian.....	42
3.2.2 Subjek Penelitian	42
3.3 Metode Penelitian	43
3.4 Metode Pengumpulan Data	44
3.5 Metode Analisis Data.....	47
3.5.1 <i>Preprocessing</i>	47
3.5.2 <i>Labeling</i>	49
3.5.3 Transformasi Data.....	51
3.5.4 Pembagian <i>Dataset (Training dan Testing)</i>	53
3.5.5 Implementasi Metode.....	54
3.5.6 Visualisasi <i>Word Cloud</i>	57
3.5.7 Evaluasi Model <i>Confusion Matrix</i>	57
3.6 Alat dan Bahan.....	60
3.6.1 Perangkat Lunak (<i>Software</i>)	60
3.6.2 Perangkat Keras (<i>Hardware</i>)	60

3.7 Prosedur Pengujian	61
3.7.1 Skenario split data.....	61
3.7.2 Skenario Perbandingan Metode	62
BAB IV ANALISIS DAN PEMBAHASAN	63
4.1 Pengumpulan Data	63
4.2 Implementasi <i>Preprocessing</i>	64
4.3 Labeling	74
4.4 Transformasi Data.....	78
4.5 Pembagian Data Latih dan Data Uji	81
4.6 Implementasi Metode SVM, LSTM dan IndoBERT.....	81
4.7 Evaluasi Model Confusion Matrix.....	90
4.8 Perbandingan Metode	94
4.9 Visualisasi <i>Word Cloud</i>	96
4.10 Pembahasan.....	98
4.11 Rekomendasi.....	101
BAB V KESIMPULAN.....	103
5.1 Kesimpulan	103
5.2 Saran.....	105
DAFTAR PUSTAKA	106
LAMPIRAN.....	110

DAFTAR TABEL

Tabel II-1 Tabel Penelitian Terdahulu	25
Tabel II-2 Perbandingan Penggunaan Metode dan alat	34
Tabel III-1 Tabel Akun TikTok	45
Tabel III-2 Tahapan <i>Preprocessing</i>	49
Tabel III-3 Contoh Kamus <i>Lexicon</i>	50
Tabel III-4 Tabel Parameter SVM	54
Tabel III-5 Tabel Parameter LSTM	55
Tabel III-6 Tabel Parameter IndoBERT	56
Tabel III-7 Tabel <i>Confusion Matrix</i>	58
Tabel III-8 Perangkat Lunak	60
Tabel III-9 Tabel Perangkat Keras	61
Tabel IV-1 Hasil <i>Text Cleaning</i>	66
Tabel IV-2 Hasil <i>Case Folding</i>	67
Tabel IV-3 Hasil <i>Normalization</i>	69
Tabel IV-4 Hasil <i>Tokenize</i>	71
Tabel IV-5 Hasil <i>Stopword Removal</i>	72
Tabel IV- 6 Hasil <i>Stemming Data</i>	74
Tabel IV-7 Hasil <i>Labeling Data</i>	77
Tabel IV-8 Hasil Implementasi Analisis SVM	83
Tabel IV-9 Hasil Implementasi Analisis LSTM	86
Tabel IV-10 Hasil Implementasi Analisis IndoBERT	89
Tabel IV-11 Perbandingan Metode	95

DAFTAR GAMBAR

Gambar II-1 Arsitektur LSTM.....	17
Gambar II-2 Arsitektur IndoBERT.....	22
Gambar II-3 Kerangka Pemikiran.....	31
Gambar III-1 Tahapan Penelitian.....	38
Gambar III-2 Tahap <i>Preprocessing</i>	47
Gambar III-3 <i>WordCloud</i>	57
Gambar IV-1 <i>Scraping Apify</i>	63
Gambar IV-2 <i>Hasil Scraping</i>	64
Gambar IV-3 Proses <i>Cleaning</i> di Google Colab.....	65
Gambar IV-4 Proses <i>Case Folding</i> di Google Colab.....	67
Gambar IV-5 Proses <i>Normalization</i> di Google Colab	69
Gambar IV-6 Proses <i>Tokenize</i> di Google Colab	70
Gambar IV-7 Proses <i>Stopword Removal</i> di Google Colab	72
Gambar IV-8 Proses <i>Stemming</i> di Google Colab.....	73
Gambar IV-9 Proses <i>Labeling</i> di Google Colab	76
Gambar IV-10 Jumlah Analisis Sentimen Metode <i>Lexicon Based</i>	78
Gambar IV-11 Proses TF-IDF SVM di Google Colab	79
Gambar IV-12 Proses <i>Tokenizer LSTM</i> di Google Colab.....	79
Gambar IV-13 Proses <i>Tokenizer IndoBERT</i> di Google Colab.....	80
Gambar IV-14 Proses SVM di Google Colab.....	82
Gambar IV-15 Proses LSTM di Google Colab.....	85
Gambar IV-16 Proses IndoBERT di Google Colab.....	88
Gambar IV-17 <i>Confusion Matrix SVM</i>	91
Gambar IV-18 <i>Confusion Matrix LSTM</i>	92
Gambar IV-19 <i>Confusion Matrix IndoBERT</i>	93
Gambar IV-20 <i>Word Cloud</i> Sentimen Positif.....	96
Gambar IV-21 <i>Word Cloud</i> Sentimen Netral	97
Gambar IV-22 <i>Word Cloud</i> Sentimen Negatif	97

DAFTAR LAMPIRAN

Lampiran 1 Data Review Komentar Asli.....	111
Lampiran 2 Data Hasil Labeling.....	116
Lampiran 3 Kode Pemrosesan	120
Lampiran 4 Hasil Pengujian <i>Confusion Matrix</i>	136
Lampiran 5 Kartu Bimbingan	148
Lampiran 6 Kartu Tanda Peserta Seminar Proposal Skripsi.....	149

