

**Komparasi Model *Support Vector Machine* dan *IndoBERT* dalam
Analisis Sentimen dan Emosi pada Ulasan Produk Tokopedia**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer**



Aldrian Aliv

1212002014

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
JAKARTA
2025**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Aldrian Aliv

NIM : 1212002014

Tanda Tangan : 

Tanggal : 6 Agustus 2025

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Aldrian Aliv

NIM : 1212002014

Program Studi : Sistem Informasi

Fakultas : Fakultas Teknik dan Ilmu Komputer

Judul Skripsi : Komparasi Model *Support Vector Machine* dan *IndoBERT* dalam Analisis Sentimen dan Emosi pada Ulasan Produk Tokopedia

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

Dewan Penguji

Pembimbing 1 : Zakiul Fahmi Jailani, S.Kom.,MSc.

()

Pembimbing 2 : Dita Nurmadewi, S.Kom., M.Kom.

()

Penguji 1 : Prof. Dr. Siti Rohajawati, S.Kom., M.Kom.

()

Penguji 2 : Dr. Elin Cahyaningsih, S.Kom., M.Msi.

()

Ditetapkan di : Jakarta

Tanggal : 6 Agustus 2025

UNGKAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkah rahmat-Nya serta segala limpahan karunia-Nya yang telah memberikan kemudahan dan kelancaran dalam penulisan dan pengerjaan tugas akhir ini dengan judul “Komparasi Model *Support Vector Machine* dan *IndoBERT* dalam Analisis Sentimen dan Emosi pada Ulasan Produk Tokopedia”. Tujuan penelitian ini dilaksanakan dalam rangka memenuhi syarat dalam mendapatkan gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer di Universitas Bakrie.

Penulis menyadari bahwa penyusunan penelitian ini disertai bantuan, doa dan bimbingan dari berbagai pihak. Dalam kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Bapak Zakiul Fahmi Jailani, S.Kom., MSc. selaku Dosen Pembimbing 1 yang telah bersedia meluangkan waktu dan tenaga untuk membimbing dan mendukung penulis dalam penyusunan tugas akhir ini.
2. Ibu Dita Nurmadewi, S.Kom., M.Kom. selaku Dosen Pembimbing 2 yang telah bersedia meluangkan waktu dan tenaga untuk membimbing dan mendukung penulis dalam penyusunan tugas akhir ini.
3. Ibu Prof. Dr. Siti Rohajawati, S.Kom., M.Kom. selaku Dosen Penguji yang telah bersedia membahas, menguji dan memberikan penilaian serta masukan kepada penulis dalam penyusunan tugas akhir ini.
4. Ibu Dr. Elin Cahyaningsih, S.Kom., M.Msi. selaku Dosen Penguji yang telah bersedia membahas, menguji dan memberikan penilaian serta masukan kepada penulis dalam penyusunan tugas akhir ini.
5. Bapak Brian Arnesto Sitorus, S.Kom., MSc. selaku dosen yang memberikan dukungan dan bimbingan seputar topik tugas akhir penulis.
6. Para Dosen Sistem Informasi Universitas Bakrie yang telah memberikan ilmu bermanfaat selama penulis melaksanakan kegiatan perkuliahan.
7. Orang tua penulis yang senantiasa memberikan dukungan dan doa sehingga penulis dapat sampai pada jenjang ini.
8. Raka Dimas Saputra, Raden Bambang Syumanjaya, Refal Fahriza Ramanda, Muhammad Zahran Shahizidan Sudirman, Muhammad

Haikal Raihansyah, Ni Kadek Srimanik, dan Yusran Rizky Mulyadi selaku teman seperjuangan penulis dalam masa perkuliahan, dan senantiasa memberikan semangat dan dukungan.

9. Mentor dan teman magang yang telah memberikan ilmu dan dukungan selama penulis magang.

Dengan adanya penyusunan tugas akhir ini dapat memberikan manfaat dan wawasan kepada pembaca. Penulis menyadari bahwa dalam penyusunan memiliki kekurangan sehingga penulis sangat menerima kritik dan saran yang membangun untuk penyusunan tugas akhir.

Jakarta, 6 Agustus 2025



Aldrian Aliv

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan dibawah ini:

Nama : Aldrian Aliv
NIM : 1212002014
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik dan Ilmu Komputer
Jenis Tugas Akhir : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah saya yang berjudul:

Komparasi Model *Support Vector Machine* dan *IndoBERT* dalam Analisis Sentimen dan Emosi pada Ulasan Produk Tokopedia.

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 6 Agustus 2025

Yang menyatakan,



Aldrian Aliv

Komparasi Model *Support Vector Machine* dan *IndoBERT* dalam Analisis Sentimen dan Emosi pada Ulasan Produk Tokopedia

Aldrian Aliv

ABSTRAK

Perkembangan teknologi telah mendorong pertumbuhan *e-commerce* di Indonesia, seiring dengan meningkatnya penetrasi internet yang mencapai 79,5% pada tahun 2024. Namun, dinamika *e-commerce* juga memunculkan tantangan berupa kebutuhan untuk mengolah informasi dari ulasan produk secara efisien. Ulasan produk memainkan peran penting dalam memengaruhi keputusan pembelian, tetapi analisis manual terhadap ulasan memerlukan waktu yang lama dan usaha yang besar. Oleh karena itu, analisis sentimen dan emosi berbasis teknologi *machine learning* dan *deep learning* menjadi solusi potensial untuk mengatasi permasalahan tersebut. Penelitian ini bertujuan untuk membandingkan kinerja model *IndoBERT* dan *Support Vector Machine* (SVM) dalam menganalisis sentimen dan emosi pada ulasan produk *e-commerce* menggunakan dataset PRDECT-ID. Hasil menunjukkan bahwa model SVM dengan fitur *n-gram* dan parameter terbaik ($C=10$, $\text{kernel}='rbf'$, $\text{gamma}=1$ untuk sentimen, dan $C=10$, $\text{kernel}='rbf'$, $\text{gamma}=0.1$ untuk emosi) menghasilkan akurasi uji sebesar 95,00% untuk sentimen dan 60,57% untuk emosi. Di sisi lain, model *IndoBERT* memberikan performa terbaik dengan akurasi 97% untuk sentimen ($\text{learning rate } 2e-5$) dan 67% untuk emosi ($\text{learning rate } 5e-5$). Penelitian ini menyimpulkan bahwa model *IndoBERT* unggul dalam hal akurasi untuk klasifikasi sentimen maupun emosi. Namun, model SVM tetap menunjukkan performa yang relatif sama dengan *IndoBERT*, terutama pada klasifikasi sentimen dengan representasi fitur *n-gram*.

Kata kunci: Analisis Sentimen, Analisis Emosi, *Support Vector Machine*, *IndoBERT*, *Machine Learning*.

Comparison of Support Vector Machine and IndoBERT Models in Sentiment and Emotion Analysis of Tokopedia Product Reviews

Aldrian Aliv

ABSTRACT

Technological advancements have driven the growth of e-commerce in Indonesia, alongside an increasing internet penetration rate that reached 79.5% in 2024. However, the dynamics of e-commerce also present challenges, particularly the need to efficiently process information from product reviews. Product reviews play a crucial role in influencing purchasing decisions, but manually analyzing them requires significant time and effort. Therefore, sentiment and emotion analysis powered by machine learning and deep learning technologies offers a promising solution to address this issue. This study aims to compare the performance of the IndoBERT and Support Vector Machine (SVM) models in analyzing sentiment and emotion in e-commerce product reviews using the PRDECT-ID dataset. The results show that the SVM model, with n-gram features and optimal parameters ($C=10$, $\text{kernel}='rbf'$, $\text{gamma}=1$ for sentiment, and $C=10$, $\text{kernel}='rbf'$, $\text{gamma}=0.1$ for emotion), achieved a test accuracy of 95.00% for sentiment and 60.57% for emotion. On the other hand, the IndoBERT model delivered the best performance with an accuracy of 97% for sentiment classification (learning rate of $2e-5$) and 67% for emotion classification (learning rate of $5e-5$). This study concludes that IndoBERT outperforms in terms of accuracy for both sentiment and emotion classification. Nevertheless, the SVM model still shows comparable performance to IndoBERT, especially in sentiment classification with n-gram feature representation.

Keywords: Sentiment Analysis, Emotion Analysis, Support Vector Machine, IndoBERT, Machine Learning.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN.....	iii
UNGKAPAN TERIMA KASIH	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR SKRIP PROGRAM	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penelitian	4
BAB II TINJAUAN PUSTAKA	6
2.1 Analisis Sentimen	6
2.2 Analisis Emosi	7
2.3 <i>Machine Learning</i>	7
2.4 <i>Deep Learning</i>	8
2.5 BERT (<i>Bidirectional Encoding Representations from Transformers</i>)	9
2.6 TF-IDF (<i>Term Frequency – Inverse Document Frequency</i>)	11
2.7 SVM (<i>Support Vector Machine</i>)	14
2.8 <i>Hugging Face</i>	16
2.9 Evaluasi Model.....	17

2.10	Penelitian Terdahulu.....	19
BAB III METODE PENELITIAN.....		25
3.1	Kerangka Penelitian	25
3.2	Metode Pengumpulan Data	31
3.3	Objek Penelitian	32
3.4	Instrumen Penelitian.....	33
3.5	Alokasi Waktu Penelitian	34
BAB IV HASIL DAN PEMBAHASAN.....		35
4.1	Pengumpulan dan <i>Input Data</i>	35
4.2	Hasil <i>Preprocessing</i>	36
4.3	Hasil <i>Exploratory Data Analysis</i>	42
4.4	Hasil Pelatihan Model SVM	47
4.5	Hasil <i>Fine-Tuning</i> Model <i>IndoBERT</i>	49
4.6	Evaluasi Hasil Model	50
4.7	Iterasi dan Optimisasi.....	51
4.8	Repositori GitHub	51
BAB V KESIMPULAN DAN SARAN.....		52
5.1	Kesimpulan	52
5.2	Saran.....	52
DAFTAR PUSTAKA		54
LAMPIRAN		58

DAFTAR GAMBAR

Gambar 2. 1 Perbedaan <i>machine learning</i> dengan <i>deep learning</i>	8
Gambar 2. 2 Proses <i>Single Sentence Task</i>	9
Gambar 2. 3 Prosedur <i>pre-training</i> dan <i>fine-tuning</i>	10
Gambar 2. 4 Ilustrasi SVM	15
Gambar 2. 5 Metode <i>machine learning</i> pada penelitian sebelumnya	16
Gambar 3. 1 Kerangka Penelitian	25
Gambar 4. 1 Cuplikan awal data	35
Gambar 4. 2 Kolom dataset yang digunakan dalam penelitian.....	36
Gambar 4. 3 Ulasan sebelum dilakukan <i>case folding</i>	37
Gambar 4. 4 Ulasan setelah dilakukan <i>case folding</i>	38
Gambar 4. 5 Label sentimen dan emosi sebelum proses <i>encoding</i>	39
Gambar 4. 6 Label sentimen dan emosi setelah proses <i>encoding</i>	39
Gambar 4. 7 Hasil vektorisasi dengan TF-IDF	40
Gambar 4. 8 Statistik deskriptif	42
Gambar 4. 9 Pengecekan nilai <i>null</i>	43
Gambar 4. 10 Persebaran label sentimen	44
Gambar 4. 11 Persebaran label emosi	45
Gambar 4. 12 Persebaran kategori	45
Gambar 4. 13 Matriks korelasi antar fitur.	46
Gambar 4. 14 Hasil <i>Cross-Validation</i>	47
Gambar 4. 15 Perbandingan hasil <i>cross-validation</i> dan hasil uji.....	48

DAFTAR TABEL

Tabel 2. 1 Perbandingan Performa <i>IndoBERT</i>	11
Tabel 2. 2 Tabel Dokumen untuk TF-IDF	13
Tabel 2. 3 Contoh pembobotan dengan TF-IDF	13
Tabel 2. 4 Seleksi Fitur TF-IDF	14
Tabel 2. 5 Penelitian Terdahulu	19
Tabel 3. 1 Dataset Yang Digunakan Dalam Penelitian.....	32
Tabel 3. 2 Alokasi Waktu Penelitian	34
Tabel 4. 1 Tokenisasi dan <i>Encoding</i> BERT	41
Tabel 4. 2 Hasil Akurasi <i>IndoBERT sentiment</i>	49
Tabel 4. 3 Hasil Akurasi <i>IndoBERT emotion</i>	49
Tabel 4. 4 Evaluasi Hasil Model	50

DAFTAR SKRIP PROGRAM

Skrip 4. 1 Penginputan dataset	35
Skrip 4. 2 <i>Case Folding</i> , Penghapusan Karakter Spesial dan Simbol	37
Skrip 4. 3 <i>Data Partition</i> dengan <i>Scikit-learn</i>	41
Skrip 4. 4 Pengecekan Duplikasi Data	43

DAFTAR LAMPIRAN

Lampiran 1 Dataset PRDECT-ID.....	58
Lampiran 2 Repositori GitHub	60
Lampiran 3 Hasil Iterasi Model	61