

SKRIPSI
ANALISIS SENTIMEN KOMENTAR PENGGUNA TIKTOK
TERHADAP OVERCLAIM PRODUK SKINCARE MENGGUNAKAN
REGRESI LOGISTIK



Disusun Oleh :
HALBILA SUSILO
42421012

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PERADABAN
2025

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Halbila Susilo
NIM : 42421012
Jurusan : Informatika
Judul : Analisis Sentimen Komentar Pengguna TikTok
terhadap *Overclaim* Produk *Skincare*
menggunakan Regresi Logistik

Dengan ini saya menyatakan bahwa Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Ahli Madya/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.

Bumiayu, 27 April 2025



LEMBAR PERSETUJUAN SKRIPSI

JUDUL : ANALISIS SENTIMEN KOMENTAR PENGGUNA
TIKTOK TERHADAP *OVERCLAIM* PRODUK
SKINCARE MENGGUNAKAN REGRESI
LOGISTIK

NAMA : HALBILA SUSILO

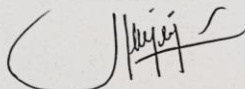
NIM : 42421012

Skripsi ini telah diperiksa serta disetujui oleh Dosen Pembimbing Tugas Akhir sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S1) pada Fakultas Sains dan Teknologi, Universitas Peradaban.

Bumiayu, 24 Agustus 2025

Menyetujui,

Dosen Pembimbing I,



Fathulloh, S.T., M.Kom
NIDN. 0623048102

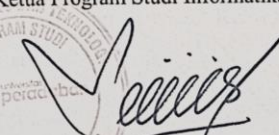
Dosen Pembimbing II,



Tezhar Rayendra TPN, M. Kom
NIDN. 0619019201

Mengetahui,

Ketua Program Studi Informatika


Nurul Mega Saraswati, M.Kom
NIDN. 0606069102

LEMBAR PENGESAHAN SKRIPSI

JUDUL : ANALISIS SENTIMEN KOMENTAR PENGGUNA
TIKTOK TERHADAP *OVERCLAIM* PRODUK
SKINCARE MENGGUNAKAN REGRESI
LOGISTIK

NAMA : HALBILA SUSILO

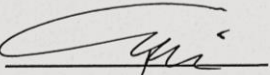
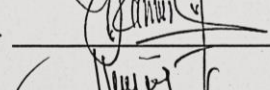
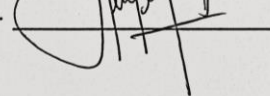
NIM : 42421012

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal 07 September 2025 dan dinyatakan memenuhi syarat.

Nama Penguji

Tanda Tangan

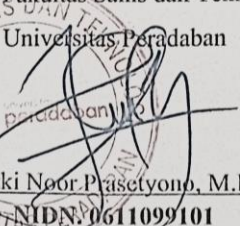
1. Khurotul Aeni, M.Kom
NIDN. 0618098802
2. Sorikhi, M.Kom
NIDN. 0608087902
3. Tezhar Revendra TPN, M.Kom
NIDN. 0619019201
4. Fathulloh, S.T.M.Kom
NIDN. 0623048102

1. 
2. 
3. 
4. 

Bumiayu, 10 September 2025

Dekan Fakultas Sains dan Teknologi

Universitas Peradaban


Rizki Noor Prasetyono, M.Pd
NIDN. 0611099101

Ketua Program Studi Informatika


Nurul Mega Saraswati, M.Kom
NIDN. 0606069102

ABSTRACT

The skincare industry in Indonesia has experienced rapid growth, triggering intense competition among producers, one of which involves the practice of overclaim exaggerated product claims without sufficient scientific evidence. This phenomenon has drawn significant attention on TikTok, particularly after the educational account “Dokter Detektif” revealed inconsistencies between actual product content and the claims displayed on packaging. This study aims to analyze TikTok users’ sentiment toward skincare products suspected of engaging in overclaim practices using a Natural Language Processing (NLP) approach. The research data consisted of 4,000 TikTok comments collected using the Apify scraping tool. The analysis process included text preprocessing (cleansing, normalization, tokenization, stopwords removal, and stemming), automated sentiment labeling with the RoBERTa model, and feature representation using the TF-IDF method. The classified data were then modeled using the Logistic Regression algorithm, with evaluation based on accuracy, precision, recall, and F1-score. The results indicate that Logistic Regression achieved an accuracy of 84.21%, with negative sentiment dominating user comments. These findings suggest that overclaim practices significantly contribute to the formation of negative consumer perceptions on social media. This study is expected to serve as a reference for skincare industry players in developing more transparent and responsible marketing communication strategies.

Keywords: Sentiment Analysis, Overclaim, Skincare, TikTok, Logistic Regression, TF-IDF, RoBERTa

ABSTRAK

Industri *skincare* di Indonesia mengalami pertumbuhan pesat yang memicu persaingan ketat, salah satunya melalui praktik *overclaim* atau klaim berlebihan tanpa bukti ilmiah. Fenomena ini mendapat sorotan di media sosial TikTok, khususnya setelah akun “*Dokter Detektif*” mengungkap ketidaksesuaian kandungan produk dengan klaim yang ditampilkan. Penelitian ini bertujuan menganalisis sentimen pengguna TikTok terhadap produk *skincare* yang diduga melakukan *overclaim* menggunakan pendekatan *Natural Language Processing* (NLP). Data penelitian berupa 4.000 komentar TikTok yang dikumpulkan menggunakan *tools scraping Apify*. Proses analisis meliputi pra-pemrosesan teks (*cleansing*, normalisasi, *tokenisasi*, *stopwords removal*, dan *stemming*), pelabelan sentimen otomatis dengan model *RoBERTa*, serta representasi fitur menggunakan metode TF-IDF. Selanjutnya, data diklasifikasikan menggunakan algoritma Regresi Logistik dengan evaluasi berbasis akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Regresi Logistik mencapai akurasi 84,21%, dengan sentimen negatif mendominasi komentar pengguna. Temuan ini menunjukkan bahwa praktik *overclaim* berkontribusi pada terbentuknya persepsi negatif konsumen di media sosial. Penelitian ini diharapkan dapat menjadi referensi bagi pelaku industri *skincare* untuk menerapkan strategi komunikasi pemasaran yang lebih transparan dan bertanggung jawab.

Kata Kunci: Analisis Sentimen, *Overclaim*, *Skincare*, TikTok, Regresi Logistik, TF-IDF, RoBERTa

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-Nya, sehingga penulis dapat menyusun skripsi ini dengan judul “Analisis Sentimen Komentar Pengguna TikTok terhadap *Overclaim* Produk *Skincare* menggunakan Regresi Logistik” sebagai salah satu tahapan dalam penyusunan tugas akhir untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban.

Penyusunan skripsi ini tentu tidak lepas dari berbagai tantangan dan hambatan, baik dalam hal pemilihan topik, pengumpulan referensi, maupun dalam menyusun kerangka berpikir yang sistematis. Namun berkat doa, dukungan, serta bimbingan dari berbagai pihak, penulis mampu melalui proses ini dengan baik. Untuk itu, penulis menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada semua pihak yang telah berkontribusi dalam proses penyusunan skripsi ini, antara lain:

1. Allah SWT, atas segala rahmat dan petunjuk-Nya yang selalu menyertai penulis dalam menjalani proses pembelajaran dan penyusunan skripsi ini.
2. Dr. Muh. Kadarisman, S.H. M.Si., selaku rektor Universitas Peradaban.
3. Bapak Rizki Noor Prasetyono, M.Pd, selaku Dekan Fakultas Sains Dan Teknologi Universitas Peradaban.
4. Ibu Nurul Mega Saraswati, M. Kom., selaku Ketua Program Studi Informatika Universitas Peradaban.
5. Bapak Fathulloh, S.T., M.Kom., selaku Dosen Pembimbing yang telah bersedia meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis dalam proses penyusunan skripsi ini.
6. Bapak Tezhar Rayendra T.P.N., M.Kom., selaku Dosen Pembimbing yang telah memberikan arahan, masukan, serta bimbingan yang sangat berarti selama proses penulisan skripsi ini.

7. Kedua orang tua penulis, Bapak *Agus Susilo* dan Ibu *Nurlela*, yang selalu mendoakan, memberikan dukungan moril maupun materiil, serta menjadi sumber semangat yang tiada henti selama proses penulisan skripsi ini.
8. Kakak penulis, *Nindya Susilo*, dan adik penulis, *Destiana Maulida Susilo*, yang senantiasa memberikan semangat, dukungan, serta menjadi penyemangat dalam setiap langkah penulis.
9. Teman-teman seperjuangan dari Program Studi Informatika semester 8 yang telah memberikan kerja sama, semangat, dan dukungan selama menjalani proses perkuliahan dan penyusunan skripsi, khususnya *Amara Putri Khaerunnisa*, *Auliya Fitra Sabila*, *Dita Putri Lestari*, *Ikrimatul A'la*, *Lilis Suryani*, dan *Pipit Purwanti*, yang senantiasa menjadi penyemangat dalam perjalanan akademik penulis.
10. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, dukungan, dan perhatian selama proses penyusunan skripsi ini.

Penulis menyadari bahwa proposal ini masih jauh dari kesempurnaan, baik dari segi isi maupun penyusunan. Oleh karena itu, penulis sangat terbuka terhadap saran dan masukan yang membangun demi perbaikan skripsi ini ke depannya. Harapan penulis, skripsi ini dapat menjadi dasar yang kuat untuk melanjutkan proses penyusunan skripsi dan memberi kontribusi positif bagi pengembangan ilmu pengetahuan, khususnya di bidang informatika.

Bumiayu, 27 April 2025

Halbila Susilo
NIM.42421012

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN	ii
LEMBAR PERSETUJUAN SKRIPSI.....	iii
LEMBAR PENGESAHAN SKRIPSI	iv
<i>ABSTRACT</i>	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Penelitian Terdahulu.....	8
2.2 Landasan Teori	13
2.2.1 <i>Overclaim</i>	13
2.2.2 Analisis Sentimen	14
2.2.3 RoBERTa	15
2.2.4 Regresi Logistik	16
2.2.5 Evaluasi Performa	18
2.2.6 Tiktok	20
2.3 Kerangka Pemikiran	21
BAB III METODOLOGI PENELITIAN.....	24
3.1 Tahap Penelitian	24

3.1.1	Identifikasi Masalah	25
3.1.2	Pengumpulan Data	25
3.1.3	Pengolahan Data.....	27
3.1.4	Pemodelan dengan Regresi Logistik.....	29
3.1.5	Evaluasi Model.....	32
3.1.6	Hasil	32
3.2	Alat dan Bahan	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Hasil Pengumpulan Data	34
4.2	Hasil Pengolahan Data	37
4.3	Pemodelan Regresi Logistik.....	41
4.4	Evaluasi Model.....	45
4.5	Perhitungan Manual	49
4.6	Tampilan Streamlit	49
BAB V KESIMPULAN.....		57
5.1	Kesimpulan.....	57
5.2	Saran	57
DAFTAR PUSTAKA		59
LAMPIRAN.....		62

DAFTAR GAMBAR

Gambar 2.1 Arsitektur RoBERTa	16
Gambar 2.2 Arsitektur Regresi Logistik	16
Gambar 2.3 <i>Confusion Matrix</i>	19
Gambar 2.4 Kerangka Pemikiran	23
Gambar 3.1 Tahapan Penelitian	24
Gambar 4.1 Tampilan <i>TikTok Comments Scraper</i> pada <i>Apify</i>	34
Gambar 4.2 Tampilan hasil <i>scraping</i> komentar TikTok	35
Gambar 4.3 Distribusi Sentimen Tiktok	40
Gambar 4.4 <i>Heatmap confusion matrix</i>	46
Gambar 4.5 <i>Visualisasi ROC Curve dan Nilai AUC</i>	48
Gambar 4.6 Tampilan Awal <i>Input</i> Komentar Manual	50
Gambar 4.7 Tampilan Hasil <i>Input</i> Komentar Manual	50
Gambar 4.8 Tampilan Analisis Data Unggah <i>Dataset</i> Mentah	51
Gambar 4.9 Tampilan Analisis Data <i>Preview Dataset</i> Mentah	51
Gambar 4.10 Tampilan Analisis Data Hasil Pelabelan	52
Gambar 4.11 Tampilan Evaluasi Model Unggah <i>Dataset</i>	52
Gambar 4.12 Tampilan Evaluasi Model <i>classification report</i>	53
Gambar 4.13 Tampilan Evaluasi Model <i>Confusion Matrix</i>	53
Gambar 4.14 Tampilan Evaluasi Model <i>ROC-AUC Curve</i>	54
Gambar 4.15 Tampilan Visualisasi Distribusi Sentimen	54
Gambar 4.16 Tampilan Visualisasi <i>Wordcloud</i> Negatif	55
Gambar 4.17 Tampilan Visualisasi <i>Wordcloud</i> Netral	55
Gambar 4.18 Tampilan Visualisasi <i>Wordcloud</i> Positif	55
Gambar 4.19 Tampilan Visualisasi 15 Kata Terbanyak pada Komentar	56
Gambar 4.20 Tampilan Laporan	56

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	11
Tabel 2.2 Kategori nilai AUC	20
Tabel 3.1 <i>Link</i> Video Tiktok	29
Tabel 3.2 <i>Hyperparameter Tuning</i>	32
Tabel 3.3 Spesifikasi Hardware dan Software Penelitian	33
Tabel 4.1 Data Hasil <i>Scraping</i>	35
Tabel 4.2 Contoh hasil pra-pemrosesan data	38
Tabel 4.3 Contoh Pelabelan Sentimen	39
Tabel 4.4 Distribusi Sentimen Hasil Pelabelan Awal	38
Tabel 4.5 Distribusi Label Sentimen.....	38
Tabel 4.6 Distribusi Label Setelah <i>Splitting</i>	42
Tabel 4.7 Konfigurasi parameter TF-IDF	42
Tabel 4.8 <i>Term</i> dengan Nilai TF-IDF Tertinggi pada Data <i>Training</i>	44
Tabel 4.9 Parameter dan Nilai yang Diuji dalam <i>Grid Search</i>	45
Tabel 4.10 hasil <i>classification report</i>	47

DAFTAR LAMPIRAN

Lampiran I Dokumentasi Hasil Cek *Similarity*