

**ANALISIS SENTIMEN KOMENTAR PRODUK MOISTURIZER DARI
BRAND SKINCARE "SKINTIFIC" DI APLIKASI TIKTOK DENGAN
BIDIRECTIONAL ENCODER REPRESENTATION from
TRANSFORMERS (BERT)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FRISCHA PRISILIA LUMENO

19.61.0162

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISIS SENTIMEN KOMENTAR PRODUK MOISTURIZER
DARI BRAND SKINCARE "SKINTIFIC" DI APLIKASI TIKTOK
DENGAN BIDIRECTIONAL ENCODER REPRESENTATION from
TRANSFORMERS (BERT)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FRISCHA PRISILIA LUMENO

19.61.0162

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR PRODUK MOISTURIZER DARI
BRAND SKINCARE "SKINTIFIC" DI APLIKASI TIKTOK DENGAN
BIDIRECTIONAL ENCODER REPRESENTATION from
TRANSFORMERS (BERT**

yang disusun dan diajukan oleh

Frischa Prisilia Lumeno

19.61.0162

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Juli 2023

Dosen Pembimbing,



Kusnawi, S.Kom, M.Eng
NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR PRODUK MOISTURIZER DARI
BRAND SKINCARE “SKINTIFIC” DI APLIKASI TIKTOK DENGAN
BIDIRECTIONAL ENCODER REPRESENTATION from
TRANSFORMERS (BERT)**

yang disusun dan diajukan oleh

Frischa Prisilia Lumeno

19.61.0162

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Juli 2023

Susunan Dewan Penguji

Nama Penguji

Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276

Moch Farid Fauzi, M.Kom
NIK. 190302284

Kusnawi, S.Kom, M.Eng
NIK. 190302112

Tanda Tangan



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Juli 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Frischa Prisilia Lumeno
NIM : 19.61.0162

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Komentar Produk Moisturizer Dari Brand Skincare
"Skintific" Di Aplikasi TikTok Dengan Bidirectional Encoder
Representation from Transformers (BERT)**

Dosen Pembimbing : Kusnawi, S.Kom, M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 24 Juli 2023

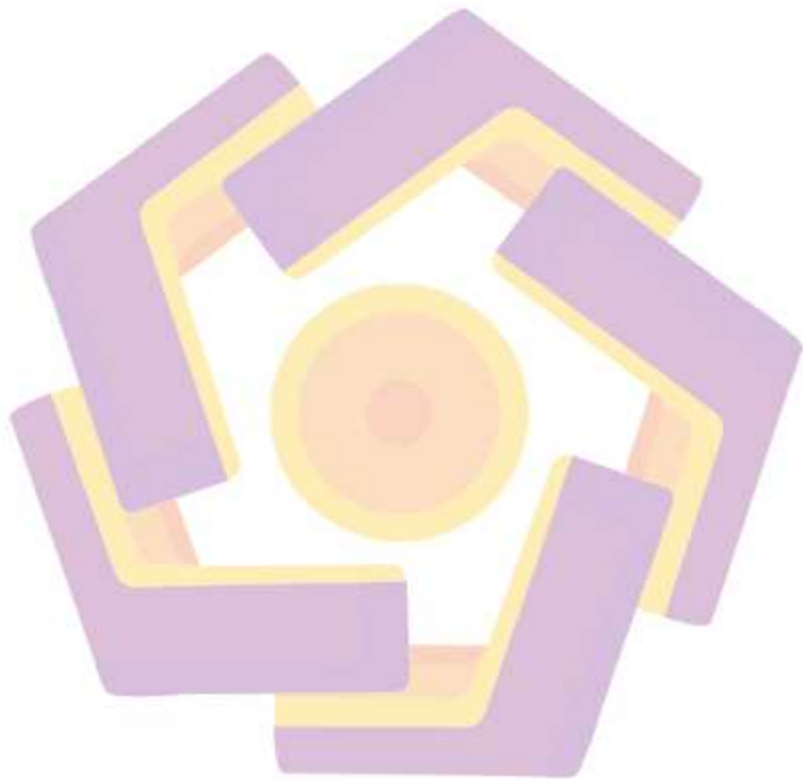
Yang Menyatakan,



Frischa Prisilia Lumeno

HALAMAN PERSEMBAHAN

Karya Ilmiah ini saya persembahkan kepada kedua orang tua penulis, Bapak Wilson F.Lumeno dan Ibu Nona Maku, serta saudari yang selalu memberikan doa, semangat, dan motivasi sampai penulis menyelesaikan karya ilmiah ini.



KATA PENGANTAR

Puji syukur penulis ucapkan kehadiran Allah SWT yang telah melimpahkan berkah, rahmat, kesehatan, serta hidayahnya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Analisis Sentimen Komentar Produk Moisturizer Dari Brand Skincare “Skintific” Di Aplikasi TikTok Dengan Bidirectional Encoder From Transformers (BERT)”. Tugas Akhir berupa skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Bachelor of Informatics di Universitas Amikom Yogyakarta.

Selama penyusunan tugas akhir ini, banyak rintangan yang dihadapi penulis, namun berkat dukungan dan bantuan baik berupa doa, tenaga, ilmu dan waktu, serta dorongan dari berbagai pihak, tugas akhir ini dapat terselesaikan dengan baik. Dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih atas kerja sama dan dukungan dari berbagai pihak selama proses penelitian hingga penyusunan tugas akhir ini kepada :

1. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi Informatika di Universitas Amikom Yogyakarta.
3. Bapak Kusnawi, S.Kom, M.Eng. selaku dosen pembimbing yang selalu meluangkan waktunya untuk membimbing penulis dan selalu memberikan support dan motivasi kepada penulis.
4. Seluruh civitas Universitas Amikom Yogyakarta yang telah memberikan pelayanan terbaik kepada penulis.
5. Bapak Wilson F. Lumeno dan Ibu Nona Maku, kedua orang tua penulis. Serta, Fanesa Olivia Lumeno selaku saudari penulis yang tidak henti-hentinya memberikan dukungan dan doa yang tulus sampai penulis menyelesaikan skripsi ini.
6. Kepada Prananda Putra Adam, yang menjadi tempat berkeluh-kesah dan support system terbesar di hidup saya. Terima kasih karena selalu memotivasi, menyemangati, mengedukasi, serta mengingatkan saya agar selalu bersungguh-

sungguh dalam menuntut ilmu. Saya persembahkan karya tulis sederhana ini untukmu.

7. Terima kasih kepada Ananda Anggun, Elvin Munsifana, dan Nurainun'nisa, yang tiada henti-hentinya memberikan semangat dan dukungan yang sangat tulus kepada saya dalam mengerjakan skripsi ini hingga selesai, dan juga terima kasih telah menjadi rumah kedua bagi saya dalam menjalani kehidupan di perantauan.
8. Teman-teman sekaligus sahabat penulis, Jihan Angraini, Sri Wahyuni, Nur Rahma, Ahmad Nur Fauzy, Alvina Malindra, Cahyarini, Charel Adi, Danny Nurakhmad, Hasna Nirfya, Leo Agista, Labib Siraj dan Raffi Putra. Terima kasih telah membantu dan menyemangati hingga karya tulis ini selesai.
9. Teman-teman EXO, Taylor Swift, dan Day6 khususnya Kim Jongin dan Kang Young-hyun yang selalu menyemangati dan menemani saya secara tidak langsung melalui karya-karya musiknya yang indah.
10. Serta semua pihak yang telah membantu dalam penyelesaian Tugas Akhir ini yang tidak dapat penulis sebutkan satu per satu.

Yogyakarta, 24 Juli 2023

Penulis

DAFTAR ISI

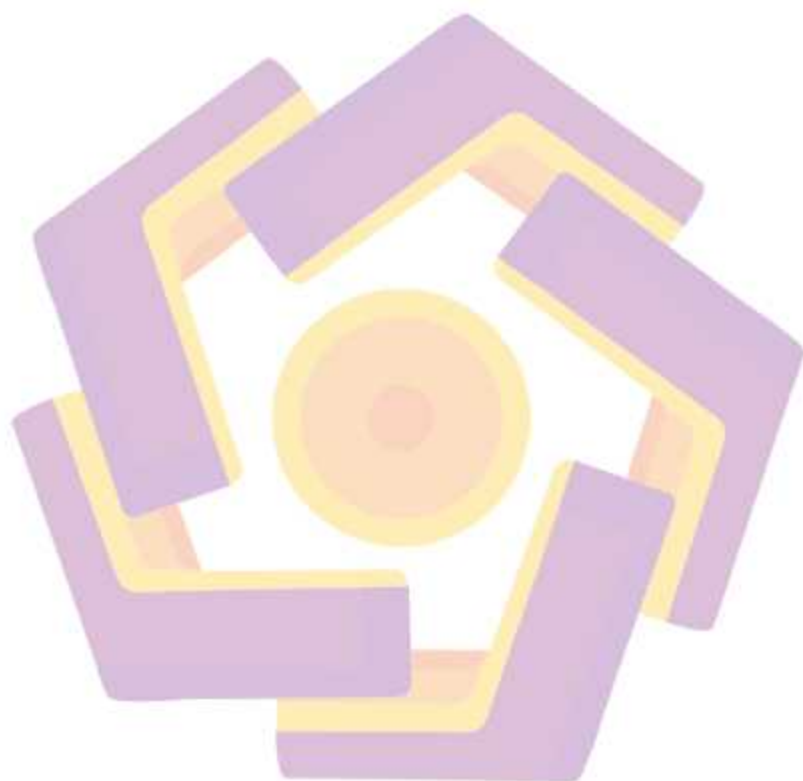
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xvi
ABSTRACT	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	20
BAB III METODE PENELITIAN	32

3.1. Objek Penelitian	32
3.2. Alat dan Bahan	32
3.2. Alur Penelitian	34
BAB IV HASIL DAN PEMBAHASAN	43
4.1. Pengumpulan Data	43
4.2. Preprocessing Data	44
4.3. Labelling Dataset	58
4.4. Exploratory Data Analysis	60
4.5. Splitting Dataset	62
4.6. BERT Model	64
BAB V PENUTUP	92
5.1. Kesimpulan	92
5.2. Saran	92
REFERENSI	93

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 3. 1 Confusion Matrix.....	41
Tabel 4. 1 Casefolding.....	44
Tabel 4. 2 Codes for Casefolding	44
Tabel 4. 3 Cleansing	45
Tabel 4. 4 Codes for Cleansing.....	45
Tabel 4. 5 Tokenizing	48
Tabel 4. 6 Codes for Tokenizing	48
Tabel 4. 7 Normalization	49
Tabel 4. 8 Codes for Normalization.....	49
Tabel 4. 9 Negation Handling.....	51
Tabel 4. 10 Codes for Negation Handling	51
Tabel 4. 11 Stemming	52
Tabel 4. 12 Codes for Stemming	53
Tabel 4. 13 Stopword Removal	54
Tabel 4. 14 Codes for Stopword Removal.....	55
Tabel 4. 15 Hasil Pre-Processing	56
Tabel 4. 16 Labelling	58
Tabel 4. 17 Codes for Labelling	58
Tabel 4. 18 Codes for EDA	61
Tabel 4. 19 Codes for Splitting Dataset.....	62
Tabel 4. 20 Tokenizer Code.....	65
Tabel 4. 21 Codes Membuat DataLoaders.....	68
Tabel 4. 22 Codes for BertForSequenceClassification Model	70
Tabel 4. 23 Code Untuk Set-Up Pre-Trained Model	72
Tabel 4. 24 Codes for Optimizer & Learning Rate Scheduler.....	73
Tabel 4. 25 Training and Validation Code	75
Tabel 4. 26 Code untuk membuat kurva dari Training history.....	84
Tabel 4. 27 Classification Report	85
Tabel 4. 28 Code for Classification report.....	86

Tabel 4. 29 Hasil prediksi dengan teks baru 88

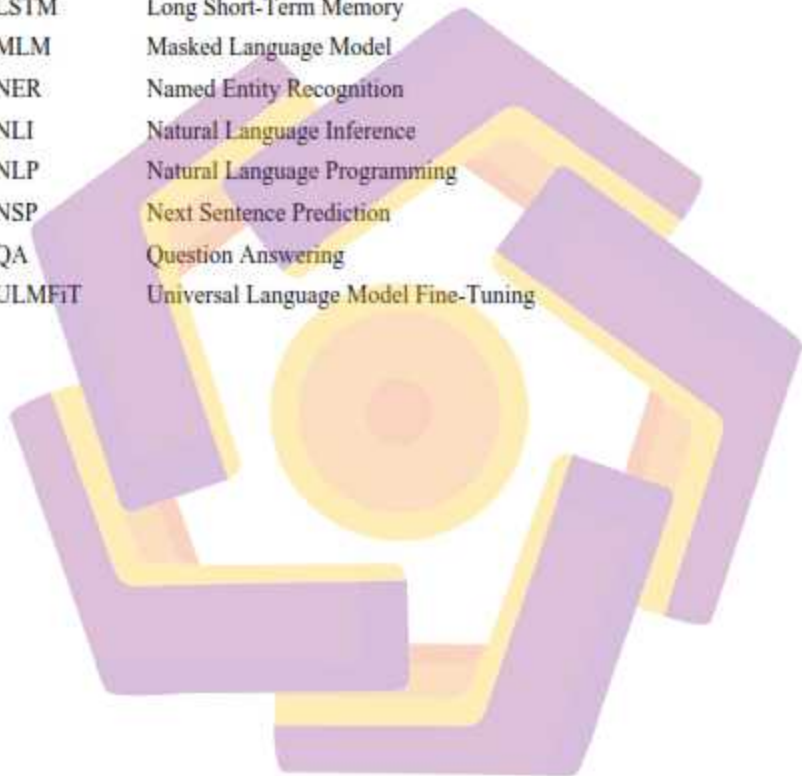


DAFTAR GAMBAR

Gambar 2. 1 Top rank pelembab wajah 1-15 maret 2022	20
Gambar 2. 2 Data pengguna aktif TikTok April 2022	21
Gambar 2. 3 Differences in pre-training model architecture	27
Gambar 3. 1 Flowchart Penelitian.....	34
Gambar 3. 2 Flowchart Pre-processing.....	37
Gambar 3. 3 Flowchart BERT	40
Gambar 4. 1 Scraping data with APIFY.....	43
Gambar 4. 2 Hasil Scraping Dataset	43
Gambar 4. 3 TikTok sentiment Pie Chart	60
Gambar 4. 4 Polarity Counts.....	61
Gambar 4. 5 Hasil Split Dataset.....	62
Gambar 4. 6 Hasil Tokenizer in Training Data	64
Gambar 4. 7 Hasil Tokenizer in Testing Data	65
Gambar 4. 8 BertForSequenceClassification Model	70
Gambar 4. 9 Hasil set-up pre-trained model.....	71
Gambar 4. 10 Training Statistic DataFrame	74
Gambar 4. 11 Kurva Hasil Performa Training dan Validasi	84
Gambar 4. 12 Plot Confusion Matrix.....	86

DAFTAR LAMBANG DAN SINGKATAN

APJII	Asosiasi Penyelenggara Jasa Internet Indonesia
BERT	Bidirectional Encoder Representation from Transformer
EDA	Exploratory Data Analysis
ELMo	Embeddings from Language Models
LSTM	Long Short-Term Memory
MLM	Masked Language Model
NER	Named Entity Recognition
NLI	Natural Language Inference
NLP	Natural Language Programming
NSP	Next Sentence Prediction
QA	Question Answering
ULMFIT	Universal Language Model Fine-Tuning

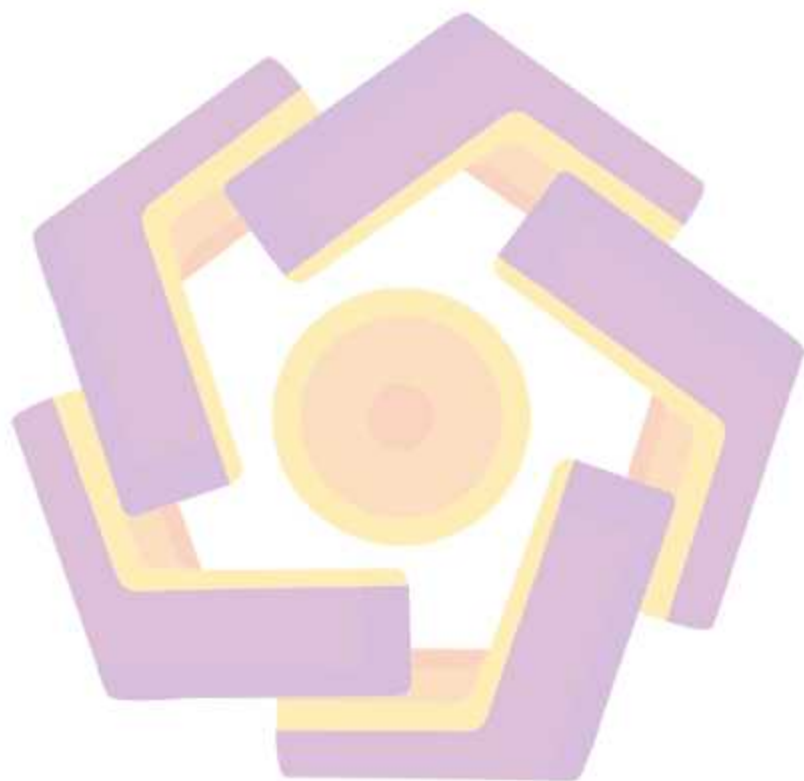


DAFTAR ISTILAH

E-Commerce	Perdagangan atau transaksi bisnis yang dilakukan secara elektronik di internet.
Fine-Tuning	Proses mengoptimalkan model yang sudah dilatih sebelumnya dengan menggunakan dataset khusus atau tugas tertentu.
Hyperparameter	Parameter yang tidak diubah selama proses pembelajaran mesin dan harus ditentukan sebelum pelatihan model.
Influencer	Merujuk pada individu yang memiliki pengaruh signifikan dalam komunitas online atau dalam industri tertentu.
Lexicon	Daftar kata-kata atau istilah yang digunakan dalam analisis teks atau analisis sentimen.
Pre-Trained	Model atau algoritma yang sudah dilatih sebelumnya pada dataset besar atau tugas tertentu.
PyTorch	Sebuah library atau perpustakaan open-source yang digunakan untuk komputasi numerik dan pengolahan data pada deep learning.
Scraping	Proses ekstraksi informasi atau data dari halaman web secara otomatis.
Self-Branding	Proses membangun dan memperkuat citra atau identitas pribadi sebagai merek yang unik dan konsisten.
Skincare	Merupakan istilah yang mengacu pada perawatan kulit.
Softmax	Fungsi matematika yang digunakan dalam pembelajaran mesin untuk menghasilkan distribusi probabilitas dari beberapa kelas eksklusif.
Tensor Flow	Platform atau perpustakaan open source yang digunakan untuk pengembangan dan pelatihan model machine learning dan deep learning.
Transfer Learning	Pendekatan dalam pembelajaran mesin dimana pengetahuan yang telah dipelajari dari suatu tugas atau

Transformer

domain digunakan dalam tugas atau domain lainnya. Arsitektur jaringan saraf yang menggabungkan perhatian (attention) untuk memodelkan hubungan antara kata-kata dalam teks.



INTISARI

Media sosial sangat penting dalam keberlangsungan bisnis saat ini. Platform media sosial seperti TikTok yang sedang populer sepanjang tahun 2020 ini menyediakan beragam konten yang informatif dan menarik perhatian banyak *influencer* kecantikan dan pengguna TikTok yang memberikan ulasan tentang penggunaannya. Penelitian ini bertujuan untuk menganalisis sentimen terhadap sebuah produk moisturizer dari brand Skintific yang menjadi viral di TikTok dan juga menarik perhatian banyak *influencer* kecantikan serta pengguna tiktok untuk memberikan ulasan tentang penggunaannya melalui sebuah konten video maupun kolom komentar. Model dan metode yang digunakan dalam penelitian ini adalah *Bidirectional Encoder Representation from Transformers* (BERT) yang dapat diklasifikasikan menjadi tiga kelas sentimen yaitu positif, negatif, dan netral. Penggunaan model BERT dalam menganalisis sentimen dinilai dapat menghasilkan akurasi yang lebih baik dari algoritma lainnya. Tahap *Pre-processing* dilakukan dengan menggunakan pendekatan *Natural Language Processing* (NLP). Jumlah dataset hasil scraping sebanyak 5.768 data komentar dari 20 top konten video di TikTok. Dataset hasil-scraping yang telah melalui *pre-processing* adalah 4.192. Penelitian ini menggunakan proporsi data latih, uji, dan validation 80:10:10, dengan pemilihan *hyperparameter* yaitu batch size 32, learning rate $2e-5$, dan dilatih sebanyak 10 epoch. Performa algoritma BERT bekerja dengan baik dalam mengklasifikasikan komentar tentang produk skintific moisturizer di platform TikTok dan memiliki nilai akurasi sebesar 0,869 atau 87%.

Kata kunci: Analisis sentimen, BERT, TikTok, Skincare, Skintific

ABSTRACT

Social media is very important for today's business continuity. Social media platforms like TikTok, which are currently popular throughout 2020, provide a variety of informative content and attract the attention of many beauty influencers and TikTok users who give reviews on their use. This study aims to analyze sentiment towards a moisturizing product from the Skintific brand, which has gone viral on TikTok and also attracted the attention of many beauty influencers and TikTok users to provide reviews about its use through video content or comment columns. The model and method used in this study are the Bidirectional Encoder Representation from Transformers (BERT), which can be classified into three sentiment classes, namely positive, negative, and neutral. Using the BERT model in analyzing sentiment is considered to produce better accuracy than other algorithms. The pre-processing stage is carried out using the Natural Language Processing (NLP) approach. The total scraped dataset is 5,768 comment data from TikTok's top 20 video content. The scraping result dataset that has gone through pre-processing is 4,192. This study uses the proportion of training, test, and validation data 80:10:10, with the selection of hyperparameters, namely batch size 32, learning rate 2e-5, and as many as 10 epochs. The performance of the BERT algorithm works well in classifying comments about skintific moisturizer products on the TikTok platform and has an accuracy value of 0.869, or 87%.

Keyword: *Sentiment analysis, BERT, TikTok, Skincare, Skintific*