

TESIS

**EKSTRAKSI ASPEK-ASPEK PERAWATAN KULIT
PADA KOMENTAR TIKTOK DENGAN
PENDEKATAN NAMED-ENTITY RECOGNITION**



Michelle Novanty Lestuny

235312716

**MAGISTER INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2025**

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

EKSTRAKSI ASPEK-ASPEK PERAWATAN KULIT PADA KOMENTAR TIKTOK DENGAN PENDEKATAN
NAMED-ENTITY RECOGNITION

yang disusun oleh

Michelle Novanty Lestuny

235312716

dinyatakan telah memenuhi syarat pada tanggal 22 Juli 2025

		Keterangan
Dosen Pembimbing 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Dosen Pembimbing 2	: Dra. Ernawati, M.T., Ph.D	Telah Menyetujui
Tim Penguji		
Penguji 1	: Y. Sigit Purnomo WP., S.T.,M.Kom., Ph.D.	Telah Menyetujui
Penguji 2	: Yonathan Dri Handarkho, S.T., M.Eng., Ph.D	Telah Menyetujui

Yogyakarta, 22 Juli 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

HALAMAN PERNYATAAN TESIS

Saya yang bertandatangan dibawah ini:

Nama : Michelle Novanty Lestuny

Nomor Mahasiswa : 235312716

Judul Tesis : EKSTRAKSI ASPEK-ASPEK PERAWATAN KULIT
PADA KOMENTAR TIKTOK DENGAN
PENDEKATAN NAMED-ENTITY RECOGNITION

Menyatakan dengan sesungguhnya bahwa Tesis ini merupakan hasil penelitian, pemikiran dan pemaparan asli saya sendiri. Saya tidak mencantumkan tanpa pengakuan bahan - bahan yang telah dipublikasikan sebelumnya atau ditulis oleh orang lain, atau sebagai bahan yang pernah diajukan untuk gelar atau ijasah pada Universitas Atma Jaya Yogyakarta atau perguruan tinggi lainnya.

Apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan peraturan yang berlaku di Universitas Atma Jaya Yogyakarta.

Demikian pernyataan ini saya buat.

Yogyakarta, 7 Juli 2025

Yang membuat pernyataan,



(Michelle Novanty Lestuny)

INTISARI

Di era digital sekarang sudah banyak perempuan yang sadar tentang pentingnya perawatan kulit. Hal ini didukung oleh platform media sosial yang menyediakan berbagai konten edukasi dan rekomendasi *skincare*. Aplikasi *TikTok* menjadi salah satu *platform* populer tempat pengguna saling berbagi pengalaman, ulasan, dan tips melalui kolom komentar. Namun, komentar-komentar tersebut umumnya tidak terstruktur, sehingga sulit dimanfaatkan secara langsung untuk memahami kebutuhan dan kebingungan pengguna terkait *skincare*. Penelitian ini bertujuan untuk membangun korpus dan mengembangkan model *Named Entity Recognition* (NER) menggunakan pendekatan *Natural Language Processing* (NLP) agar mampu mengenali delapan entitas penting dalam komentar, yaitu kandungan *skincare*, jenis kulit, permasalahan kulit, tips, *impact*, jenis produk, waktu perawatan, dan perawatan. Proses dilakukan melalui *scraping* data komentar *TikTok*, pembuatan korpus yang didalamnya terdapat proses pembuatan panduan *keyword*, *pre-processing*, *multilabel classification*, anotasi menggunakan *Prodigy*, dan verifikasi *expert* dan pengembangan model dengan *spaCy*. Validasi dilakukan bersama *expert* dari industri *skincare* untuk memastikan akurasi dan keamanan informasi. Hasil evaluasi menunjukkan bahwa model yang dikembangkan memiliki performa tinggi dengan *precision* sebesar 94.44, *recall* sebesar 92.27, dan *F1-score* sebesar 93.34. Temuan ini menunjukkan bahwa pendekatan NLP dan NER efektif dalam mengekstraksi informasi dari komentar *TikTok*, serta berpotensi dikembangkan lebih lanjut sebagai sistem rekomendasi *skincare* yang berbasis data pengguna.

Kata-kata kunci: *Skincare*, *TikTok*, *Natural Language Processing* (NLP), *Named Entity Recognition* (NER), *spaCy*, *Prodigy*, ekstraksi informasi

ABSTRAK

In this digital era, many women are now aware of the importance of skincare. This is supported by social media platforms that provide various educational content and skincare recommendations. TikTok app is one of the popular platforms where users share experiences, reviews and tips through the comment section. However, these comments are generally unstructured, making it difficult to directly utilize to understand users' needs and confusion related to skincare. This study aims to build a corpus and develop a Named Entity Recognition (NER) model using Natural Language Processing (NLP) approach to be able to recognize eight important entities in comments, namely skincare content, skin type, skin problems, tips, *impact*, product type, time, and treatment. The process is carried out through scraping TikTok comment data, corpus creation which includes the process of making *keyword* guides, pre-processing, multilabel classification, annotation using Prodigy, and expert verification and model development with spaCy. Validation is done with experts from the skincare industry to ensure accuracy and information security. The evaluation results show that the developed model has high performance with precision of 94.44, recall of 92.27, and F1-score of 93.34. The findings show that NLP and NER approaches are effective in extracting information from TikTok comments, and have the potential to be further developed as a skincare recommendation system based on user data.

Keyword s: Skincare, TikTok, Natural Language Processing (NLP), Named Entity Recognition (NER), spaCy, Prodigy, information extraction.

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Tesis dengan judul EKSTRAKSI ASPEK-ASPEK PERAWATAN KULIT PADA KOMENTAR TIKTOK DENGAN PENDEKATAN NAMED-ENTITY RECOGNITION ini telah selesai disusun. Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Tesis ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan yang maha kasih,
2. Kedua Orang Tua penulis yang selama ini memberikan doa yang tidak putus-putus dan semangat untuk penulis yang rela menahan semua beban asam garam kehidupan dan rasa sakit untuk bertahan demi melihat penulis berhasil mendapatkan gelar dengan memakai toga,
3. Bapak Yonathan Dri Handarkho, S.T., M.Eng., Ph.D selaku Kaprodi Magister Informatika Universitas Atma Jaya Yogyakarta, sekaligus dosen penguji yang telah banyak memberikan motivasi dan masukan dalam menyelesaikan tesis,
4. Bapak Yohanes Sigit Purnomo W.P, S.T., M.Kom., Ph.D dan Dra. Ernawati, M.T., Ph.D selaku Dosen Pembimbing yang telah bersedia memberi bimbingan, gagasan ide, koreksi dan pengarahan kepada penulis sehingga tesis ini dapat diselesaikan dengan baik,
5. Bapak Stanilaus Rahmadi selaku General Manager pada PT L'Essential telah menerima, membantu, dan memberi semangat dalam proses verifikasi data pada tesis ini
6. Seluruh Dosen, Staf Pengajar, dan Karyawan Departemen Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama masa kuliah di Program Studi Magister Informatika, Departemen Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta

7. Teman-teman tercinta, Dhea Salamena, Febby CA, Aulya Sujud, Theoda Boboy, Anggres Monica, Alody Entertainment, Puja Praditya, Niko, dan semua teman-teman Magister Informatika UAJY angkatan 2023 yang sudah memberikan semangat dan dukungan kepada penulis,
8. Semua pihak yang telah membantu menyelesaikan tesis ini, terima kasih atas semua dukungan dan doanya selama ini

Dengan keterbatasan pengalaman dan pengetahuan, penulis menyadari bahwa tesis ini masih jauh dari sempurna dan memiliki kelemahan serta perlu pengembangan lebih lanjut. Oleh karena itu segala bentuk kritik maupun saran yang sifatnya membangun akan sangat diharapkan kedepannya. Akhir kata, semoga tesis ini dapat bermanfaat dan berguna bagi semua pihak.

Yogyakarta, 7 Juli 2025



Michelle Novanty Lestuny

DAFTAR ISI

HALAMAN PENGESAHAN.....	2
HALAMAN PERNYATAAN TESIS.....	3
INTISARI.....	4
ABSTRAK	5
KATA PENGANTAR	6
DAFTAR ISI.....	8
DAFTAR GAMBAR	10
DAFTAR TABEL	11
BAB I	12
PENDAHULUAN	12
1.1. Latar Belakang	12
1.2. Rumusan Masalah	14
1.3. Pertanyaan Penelitian	14
1.4. Tujuan Penelitian	14
1.5. Batasan Masalah.....	15
1.6. Manfaat Penelitian	15
1.7. Sistematika Penulisan	16
BAB II.....	18
TINJAUAN PUSTAKA	18
BAB III	36
LANDASAN TEORI.....	36
3.1. <i>Natural Language Processing</i> (NLP)	36
3.2. <i>Named Entity Recognition</i> (NER).....	38
BAB IV	43
METODOLOGI PENELITIAN.....	43
4.1. Tahapan Penelitian	43
4.2. Identifikasi Masalah	44
4.3. Studi Literatur	44
4.4. Pengambilan Data	44

4.5.	Membuat Korpus.....	45
4.5.1	Pre-Processing.....	45
4.5.2	Verifikasi <i>Expert</i>	48
4.5.3	Anotasi	48
4.6.	Mengembangkan Model NER.....	49
BAB V		51
5.1.	Pengambilan Data	51
5.2.	Membuat Korpus.....	53
5.2.1	Pre-processing	53
5.3.	Develop Model NER.....	65
5.3.1	Training Model.....	68
5.3.2	Validasi Model	71
5.3.3	Testing Model	72
5.4.	Analisis Pembahasan.....	74
BAB VI		77
6.1	Kesimpulan	77
6.2	Saran.....	78
DAFTAR PUSTAKA		79
Lampiran Bukti Verifikasi Data		83

DAFTAR GAMBAR

Gambar 3. 1 Alur Pre Processing NLP	37
Gambar 3. 2 Pelabelan pada proses pengembangan model NER (Franchino et al., 2024)	39
Gambar 3. 3 Arsitektur Transition Based Parsing	40
Gambar 3. 4 Gold Standard Labels (Purnomo et al., 2024).....	41
Gambar 4. 1 Alur Penelitian.....	43
Gambar 4. 2 Model NER dengan Token2Vec mengadopsi Multi Hash Embedding dan Maxout Window Encoder pada penelitian ini	49
Gambar 5. 1 Komentar tentang skincare pada aplikasi TikTok.....	51
Gambar 5. 2 Hasil scraping data	52
Gambar 5. 3 Panduan Keyword	54
Gambar 5. 4 Code cleaning dan lowercasing.....	56
Gambar 5. 5 Code tokenization dan lemmatization	58
Gambar 5. 6 Code custom_stem pada lemmatization.....	59
Gambar 5. 7 Code dictionary	60
Gambar 5. 8 Code multilabel classification	60
Gambar 5. 9 Hasil pre-processing.....	61
Gambar 5. 10 Profil PT L'Essential	62
Gambar 5. 11 Hasil verifikasi dengan expert.....	63
Gambar 5. 12 Proses anotasi dengan ner manual Prodigy	64
Gambar 5. 13 Code pada file project.yml	67
Gambar 5. 14 Proses membuat file konfigurasi config.cfg.....	67
Gambar 5. 15 Proses training model NER	68
Gambar 5. 16 Struktur konfigurasi pipeline model NER spaCy.....	69
Gambar 5. 17 Konfigurasi optimizer pada proses training model NER spaCy	69
Gambar 5. 18 Konfigurasi training model NER spaCy	70
Gambar 5. 19 Isi folder model-best.....	70
Gambar 5. 20 Pengaturan bobot skor evaluasi pada konfigurasi training.score_weights spaCy	71
Gambar 5. 21 Konfigurasi komponen NER pada pipeline spaCy	71
Gambar 5. 22 Proses validasi model NER.....	72
Gambar 5. 23 Proses testing model NER.....	73

DAFTAR TABEL

Tabel 2. 1 Daftar review penelitian terdahulu.....	26
Tabel 5. 1 Deskripsi data set	55
Tabel 5. 2 Contoh hasil proses cleaning.....	56
Tabel 5. 3 Contoh hasil proses lowercasing.....	57
Tabel 5. 4 Contoh hasil proses tokenization	58
Tabel 5. 5 Contoh hasil proses lemmatization	59