PENGEMBANGAN SISTEM REKOMENDASI PRODUK KECANTIKAN LOKAL TERDAFTAR BPOM MENGGUNAKAN PENDEKATAN *NATURAL LANGUAGE PROCESSING* (NLP)

Tugas Akhir

Diajukan untuk memenuhi persyaratan mencapai derajat

Sarjana Sistem Informasi

Cinta Chantika Lestari

NPM: 221711729

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2026

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

PENGEMBANGAN SISTEM REKOMENDASI PRODUK KECANTIKAN LOKAL TERDAFTAR BPOM MENGGUNAKAN PENDEKATAN NATURAL LANGUAGE PROCESSING (NLP)

yang disusun oleh

Cinta Chantika Lestari

221711729

dinyatakan telah memenuhi syarat pada tanggal 22 Juni 2026

		Keterangan
Dosen Pembimbing 1	: Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Dosen Pembimbing 2	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Penguji 2	: Paulus Mudjihartono, S.T., M.T., Ph.D.	Telah Menyetujui
Penguji 3	: Emanuel Ristian Handoyo, S.T., M.Eng.	Telah Menyetujui

Yogyakarta, 22 Juni 2026

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Prof. Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU, ASEAN Eng.

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

LEMBAR PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Cinta Chantika Lestari

NPM : 221711729

Program Studi : Sistem Informasi

Fakultas : Fakultas Teknologi Industri

Judul Penelitian : Pengembangan Sistem Rekomendasi Produk Kecantikan Lokal Terdaftar BPOM menggunakan Pendekatan *Natural Language Processing* (NLP)

Menyatakan dengan ini:

1. Skripsi ini adalah benar merupakan hasil karya sendiri dan tidak merupakan salinan sebagian atau keseluruhan dari karya orang lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta, berupa Hak Bebas Royalti non eksklusif (*Non-Exclusive-Royalty-Free Right*) atas penelitian ini, dan berhak menyimpan, mengelola dalam pangkalan data, mendistribusikan, serta menampilkan untuk kepentingan akademis, tanpa perlu meminta izin selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum yang mengikuti atas pelanggaran Hak Cipta dalam pembuatan skripsi ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 3 Juni 2026

Yang menyatakan,



Cinta Chantika Lestari

221711729

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas penyertaan dan kasih karunia-Nya, sehingga penulis mampu menyelesaikan laporan tugas akhir yang berjudul "Pengembangan Sistem Rekomendasi Produk Kecantikan Lokal Terdaftar BPOM menggunakan Pendekatan *Natural Language Processing* (NLP)" sesuai dengan target penulis. Laporan ini disusun sebagai bagian dari pemenuhan persyaratan akademik dalam meraih gelar Sarjana Komputer di Program Studi Sistem Informasi, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.

Penulis memahami bahwa penyusunan laporan ini bukan hal yang mudah. Meskipun terdapat berbagai keterbatasan dan tantangan selama pelaksanaan, penyusunan laporan ini dapat berjalan dengan baik berkat doa, bimbingan, semangat, dan dukungan dari berbagai pihak. Untuk itu, saya menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Orangtua dan keluarga penulis, yang selalu dengan tulus mendoakan, memberikan dukungan, semangat, dan nasihat sehingga penulis dapat menyelesaikan tugas akhir studi ini.
2. Ibu Citra Yayu' Palangan, S.T., M.Sc. dan Bapak A. Gonzaga Pradnya Sidhawara, S.T., M.Eng. selaku dosen pembimbing tugas akhir penulis, atas waktu yang telah diberikan, bimbingan, arahan, dukungan, serta kesediaan untuk menjawab berbagai pertanyaan penulis selama penyusunan laporan.
3. Seluruh dosen Sistem Informasi Universitas Atma Jaya Yogyakarta, yang telah membagikan ilmu, mendukung, serta membimbing penulis selama masa perkuliahan hingga penulis bisa berada di titik ini.
4. Sahabat, teman-teman, dan *partner* penulis, yang selalu ada untuk penulis baik saat suka maupun duka, menjadi tempat *sharing* dan berkeluh kesah, memberikan doa, semangat, serta motivasi bagi penulis di tengah-tengah kesibukan.

5. Kepada diri penulis sendiri, atas segala ketangguhan, kesabaran, dan kerja keras yang telah dikerahkan hingga titik ini. Terima kasih karena telah memilih untuk tidak menyerah di saat-saat tersulit, tetap tegak berdiri meski harus membagi waktu antara kegiatan akademik dan tanggung jawab lainnya, serta atas keberanian untuk terus melangkah menyelesaikan apa yang telah dimulai.

Penulis berharap dengan adanya laporan tugas akhir ini dapat memberikan manfaat dan kontribusi bagi pihak-pihak yang berkepentingan, serta menjadi referensi mahasiswa tahun ajaran mendatang.

"*We've always defined ourselves by the ability to overcome the impossible. And we count these moments… when we dare to aim higher, to break barriers, to reach for the stars, to make the unknown known.*"

—Interstellar (2014)

ABSTRAK

Banyaknya pilihan produk kecantikan yang tersedia di *platform e-commerce* tidak jarang menyulitkan konsumen dalam menentukan produk yang sesuai dengan kebutuhan dan kondisi kulit mereka, terlebih dengan minimnya informasi mengenai kandungan produk serta belum terintegrasinya informasi legalitas BPOM sebagai jaminan keamanan. Penelitian ini bertujuan mengembangkan sistem rekomendasi produk kecantikan lokal berbasis *content-based filtering* yang memanfaatkan skor kemiripan deskripsi produk, sekaligus mengintegrasikan informasi nomor registrasi BPOM sebagai aspek keamanan dan legalitas. Dataset yang digunakan mencakup 327 produk *skincare* dan *makeup* dari lima perusahaan kecantikan lokal terbesar di Indonesia yang diperoleh melalui *web scraping* dari Shopee pada periode Februari hingga Maret 2026. Evaluasi dilakukan berdasar data penilaian oleh responden dan dinilai menggunakan metrik Precision@5, F1@5, MAP@5, dan NDCG@5. Sistem terpilih, yakni penggabungan IndoBERT dan TF-IDF dengan bobot seimbang 50:50, berhasil meraih performa terbaik dengan Precision@5 sebesar 0,9833, F1@5 sebesar 0,9907, MAP@5 sebesar 0,9837, dan NDCG@5 sebesar 0,9473. Hasil ini menunjukkan bahwa pendekatan *hybrid* mampu menangkap kecocokan tekstual sekaligus makna kontekstual secara lebih komprehensif dibandingkan metode tunggal. Sistem diimplementasikan sebagai *prototype* aplikasi web menggunakan Streamlit, di mana pengguna dapat memperoleh rekomendasi produk lokal yang relevan beserta informasi legalitas BPOM-nya.

Kata kunci: sistem rekomendasi, *content-based filtering*, produk kecantikan lokal, IndoBERT, TF-IDF

ABSTRACT

The large number of beauty products available on e-commerce platforms often makes it difficult for consumers to identify products that match their skin conditions and needs, particularly due to limited ingredient information and the lack of integrated BPOM legality verification. This study aims to develop a content-based filtering recommendation system for local beauty products by leveraging product description similarity and integrating BPOM registration information as an indicator of product safety and legality. The dataset consists of 327 skincare and makeup products from five major Indonesian local beauty companies, collected through web scraping from Shopee between February and March 2026. The system was evaluated using respondent ratings and assessed with Precision@5, F1@5, MAP@5, and NDCG@5. The best-performing model, a hybrid combination of IndoBERT and TF-IDF with equal weighting (50:50), achieved a Precision@5 of 0.9833, F1@5 of 0.9907, MAP@5 of 0.9837, and NDCG@5 of 0.9473. These results indicate that the hybrid approach captures both textual similarity and contextual meaning more effectively than individual methods. The system was implemented as a Streamlit-based web application that provides personalized local beauty product recommendations along with BPOM legality information.

***Keywords**: recommendation system, content-based filtering, local beauty products, IndoBERT, TF-IDF*

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN PENELITIAN ..... **ii**

KATA PENGANTAR ..... **iv**

ABSTRAK ..... **vii**

ABSTRACT ..... **viii**

DAFTAR ISI ..... **ix**

DAFTAR GAMBAR ..... **xiii**

DAFTAR TABEL ..... **xv**

BAB I PENDAHULUAN **1**

1.1. Latar Belakang 1

1.2. Rumusan Masalah 3

1.3. Pertanyaan Penelitian 4

1.4. Batasan Masalah 4

1.5. Tujuan Penelitian 5

1.6. Manfaat Penelitian 5

1.7. Bagan Keterkaitan 6

BAB II TINJAUAN PUSTAKA **7**

2.1. Studi Sebelumnya 7

2.2. Dasar Teori 13

2.2.1. Produk Kecantikan 13

2.2.2. Badan Pengawas Obat dan Makanan (BPOM) 14

2.2.3. *E-Commerce* 14

2.2.4. Sistem Rekomendasi 15

2.2.5. Natural Language Processing (NLP) 16

2.2.6. *Term Frequency-Inverse Document Frequency* (TF-IDF) 17

2.2.7. *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT) 18
2.2.8. *Cosine Similarity* 19
2.2.9. *Weighted Composition* (Pembobotan Komposisi) 20
2.2.10. Evaluasi Model 20
2.2.11. Streamlit 22
BAB III METODE PENELITIAN 23
3.1. Tahapan Penelitian 23
3.2. Metode Penelitian 25
3.2.1. Identifikasi Masalah 25
3.2.2. Studi Literatur 25
3.2.3. Pengumpulan Data 25
3.2.4. *Preprocessing* dan Normalisasi Data 26
3.2.5. Representasi Fitur Teks 27
3.2.6. Pemodelan Sistem Rekomendasi 27
3.2.7. Evaluasi Model 28
3.2.8. Pembuatan *Prototype Web* 29
3.2.9. Interpretasi Hasil 29
3.2.10. Pembahasan dan Kesimpulan 29
3.3. Alat yang Digunakan 29
3.3.1. Google Colab 29
3.3.2. Visual Studio Code 30
3.3.3. BigSeller Scraper Extension 30
3.3.4. Pandas & NumPy 30
3.3.5. Scikit-learn 30
3.3.6. Sastrawi 30

3.3.7. Transformers (IndoBERT) & PyTorch 31
3.3.8. Matplotlib & Seaborn 31
3.3.9. Streamlit 31
BAB IV HASIL DAN PEMBAHASAN 32
4.1. Pengumpulan Data 32
4.2. *Preprocessing* dan Normalisasi Data 33
4.2.1 *Data Cleaning* 33
4.2.2 *Data Partitioning / Column Selection* 35
4.2.3 *Feature Engineering* 35
4.2.4 *Text Cleansing* 37
4.2.5 *Case Folding* 38
4.2.6 *Stemming* 39
4.2.7 *Stopwords Removal* 41
4.3. Representasi Fitur Teks 42
4.3.1. TF-IDF 42
4.3.2. IndoBERT 44
4.4. Pemodelan Sistem Rekomendasi 46
4.4.1. *Cosine Similarity* 47
4.4.2. *Weighted Composition* (Pembobotan Komposisi) 48
4.4.3. Peringkat dan Hasil Rekomendasi (*Top-K Recommendation*) 49
4.5. Evaluasi Model 50
4.5.1. Data Preferensi Pengguna (*Ground Truth*) 50
4.5.2. Penghitungan Metrik Evaluasi 53
4.6. Pembuatan *Prototype Web* 57
4.6.1. Implementasi Sistem 57
4.6.2. Konfigurasi Awal dan Pemuatan Model 58

4.6.3. Fitur Input Pengguna dan Pemilihan Sistem 59
4.6.4. Komputasi dan Pengembalian Hasil Rekomendasi 60
4.6.5. Tampilan Hasil Rekomendasi 61
4.7. Interpretasi Hasil 63
4.7.1. Skor Metrik Evaluasi 63
4.7.2. Analisis Kecocokan dengan *Ground Truth* Pengguna 64
4.8. Pembahasan dan Kesimpulan 64
4.8.1. Temuan Penelitian 64
4.8.2. Penerapan Sistem Rekomendasi 66
4.8.3. Kelebihan Sistem 67
4.8.4. Keterbatasan Sistem 67
4.8.5. Kesimpulan Pembahasan 68
BAB V KESIMPULAN DAN SARAN 70
5.1. Kesimpulan 70
5.2. Saran 70
DAFTAR PUSTAKA 72
LAMPIRAN 78

DAFTAR GAMBAR

Gambar 1. 1 Bagan Keterkaitan 6
Gambar 2. 1 Ilustrasi Content-Based Filtering 16
Gambar 2. 2 Arsitektur Representasi IndoBERT 18
Gambar 3. 1. Tahapan Penelitian 24
Gambar 4. 1 Sampel Data Hasil Web Scraping 32
Gambar 4. 2 Kode dan Jumlah Nilai Kosong pada Dataset 33
Gambar 4. 3 Kode untuk Menghapus Data Duplikat 34
Gambar 4. 4 Kode untuk Data Partitioning/Column Selection 35
Gambar 4. 5 Kode untuk Feature Engineering 35
Gambar 4. 6 Kode untuk Text Cleansing 37
Gambar 4. 7 Kode untuk Case Folding 38
Gambar 4. 8 Kode untuk Stemming 40
Gambar 4. 9 Kode untuk Stopwords Removal 41
Gambar 4. 10 Kode untuk Vektorisasi Teks dengan TF-IDF 43
Gambar 4. 11 Nilai Hasil Vektorisasi TF-IDF 43
Gambar 4. 12 Kode untuk Memuat dan Inisialisasi IndoBERT 44
Gambar 4. 13 Kode untuk Eksplorasi dan Ekstraksi Nilai Embedding 45
Gambar 4. 14 Kode Penerapan Fungsi Ekstraksi IndoBERT 45
Gambar 4. 15 Hasil Ekstraksi Fitur Teks dengan IndoBERT 46
Gambar 4. 16 Kode untuk Menghitung Cosine Similarity TF-IDF dan IndoBERT 47
Gambar 4. 18 Tabel Nilai Cosine Similarity IndoBERT 47
Gambar 4. 17 Tabel Nilai Cosine Similarity TF-IDF 47
Gambar 4. 19 Kode untuk Normalisasi Skala dan Weighted Composition 48
Gambar 4. 20 Tabel Nilai Matriks Weighted Composition TF-IDF dan IndoBERT 50:50 48
Gambar 4. 21 Kode untuk Proses Peringkat dan Rekomendasi 49
Gambar 4. 22 Contoh Hasil Rekomendasi Top-5 49
Gambar 4. 23 Kuesioner Penilaian Relevansi Produk Rekomendasi 51
Gambar 4. 24 Grafik Hasil Pemilihan dari Responden 52
Gambar 4. 25 Kode untuk Menghitung Metrik Evaluasi 53

Gambar 4. 26 Kode untuk Memuat Kuesioner dan Menyusun Ground Truth 54
Gambar 4. 27 Data Ground Truth Weighted IndoBERT+TF-IDF 54
Gambar 4. 28 Grafik Perbandingan Hasil Evaluasi Model 56
Gambar 4. 29 Kode Konfigurasi Halaman 58
Gambar 4. 30 Kode Pembuatan Komponen Dropdown dan Input Query 59
Gambar 4. 31 Kode Pembuatan Komponen Tombol "Rekomendasi Produk" 60
Gambar 4. 32 Kode Komputasi Skor Similarity 61
Gambar 4. 33 Kode Komponen Hasil Produk Rekomendasi 61
Gambar 4. 34 Contoh Tampilan Hasil Produk Rekomendasi 62
Gambar 4. 35 Kode Komponen Modal Informasi Lengkap Produk 62
Gambar 4. 36 Contoh Tampilan Modal Informasi Lengkap Produk 62

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu 9
Tabel 2. 2 Perbandingan Metode 10
Tabel 4. 1 Contoh Hasil Feature Engineering 36
Tabel 4. 2 Contoh Hasil Text Cleansing 37
Tabel 4. 3 Contoh Hasil Case Folding 39
Tabel 4. 4 Contoh Hasil Stemming 40
Tabel 4. 5 Contoh Hasil Stopwords Removal 41
Tabel 4. 6 Hasil Metrik Evaluasi Seluruh Sistem 55
Tabel 4. 7 Tabel Sistem yang Diimplementasikan 57