

**KOMPARASI PENGGUNAAN METODE PROBLEM
TRANSFORMATION BERBASIS XGBOOST UNTUK
KLASIFIKASI MULTI-LABEL EFEK KOSMETIK**

SKRIPSI

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



disusun oleh

LUTHFIYYAH NUR NAFI'AH

22.11.4987

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**KOMPARASI PENGGUNAAN METODE PROBLEM
TRANSFORMATION BERBASIS XGBOOST UNTUK
KLASIFIKASI MULTI-LABEL EFEK KOSMETIK**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

LUTHFIYYAH NUR NAFI'AH

22.11.4987

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI PENGGUNAAN METODE PROBLEM
TRANSFORMATION BERBASIS XGBOOST UNTUK
KLASIFIKASI MULTI-LABEL EFEK KOSMETIK**

yang disusun dan diajukan oleh

Luthfiyyah Nur Nafi'ah

22.11.4987

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 April 2026

Dosen Pembimbing,



Windha Mega Pradnya Duhita, S.Kom., M.Kom
NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

**KOMPARASI PENGGUNAAN METODE PROBLEM
TRANSFORMATION BERBASIS XGBOOST UNTUK
KLASIFIKASI MULTI-LABEL PADA EFEK KOSMETIK**

yang disusun dan diajukan oleh

Luthfiyyah Nur Nafi'ah

22.11.4987

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 April 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Uyock Anggoro Saputro, S.Kom., M.Kom.
NIK. 190302419

Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276

Arifivanto Hadinegoro, S.Kom., M.T.
NIK. 190302289



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 April 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Luthfiyyah Nur Nafi'ah
NIM : 22.11.4987

Menyatakan bahwa Skripsi dengan judul berikut:

KOMPARASI PENGGUNAAN METODE PROBLEM TRANSFORMATION BERBASIS XGBOOST UNTUK KLASIFIKASI MULTI-LABEL EFEK KOSMETIK

Dosen Pembimbing : Windha Mega Pradnya Dhuhiita, S.Kom, M.Kom

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, 20 April 2026

Yang Menyatakan,




Luthfiyyah Nur Nafi'ah

HALAMAN PERSEMBAHAN

Dengan mengucapkan *Alhamdulillahirabbil'alamiin*, puji syukur kehadiran Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Sholawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh ilmu pengetahuan. Karya sederhana ini penulis persembahkan kepada:

1. Ayahanda dan Ibunda tercinta, sebagai wujud bakti, hormat, dan terima kasih yang mendalam atas segala doa, kasih sayang, serta pengorbanan yang tidak terhingga.
2. Keluarga besar, yang telah memberikan dukungan dan motivasi selama penulis menempuh pendidikan.
3. Almamater tercinta, Universitas AMIKOM Yogyakarta, sebagai tempat penulis menimba ilmu dan mengembangkan diri dalam dunia akademik.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi untuk memenuhi salah satu syarat kelulusan dalam memperoleh gelar sarjana pada program studi informatika, Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa keberhasilan penyelesaian skripsi ini dapat terwujud berkat bantuan, bimbingan, dan kerja sama dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua penulis, Bapak Giarto dan Ibu Tri. Terima kasih atas cinta yang tak pernah menuntut balas, doa yang tak pernah terputus, serta perhatian yang selalu hadir dalam setiap langkah penulis, bersama pengorbanan yang sering kali tak terlihat namun begitu terasa. Semua itu menjadi kekuatan yang mengantarkan penulis hingga sampai pada titik ini.
2. Ibu Windha Mega P.D., S.Kom., M.Kom., selaku dosen pembimbing yang dengan kesabaran dan ketulusan telah meluangkan waktu, memberikan arahan serta masukan yang menjadi penuntun bagi penulis dalam menyelesaikan setiap bagian dari skripsi ini.
3. Dosen penguji yang telah memberikan saran serta kritik konstruktif demi kesempurnaan skripsi ini.
4. Farah dan Azril, kedua adik penulis yang selalu menemani selama penyusunan skripsi ini. Terima kasih atas keceriaan yang selalu dihadirkan, sehingga mampu menjadi penghibur bagi penulis
5. Nisa, Deka, dan Esti, sahabat penulis sejak awal masa perkuliahan hingga penyusunan skripsi ini, serta teman-teman kelas Informatika 07 yang tidak dapat disebutkan satu per satu. Terima kasih atas kebersamaan dan solidaritas yang terjalin selama masa perkuliahan.
6. Kepada Firdaus, terima kasih telah senantiasa hadir menemani setiap proses penyusunan skripsi, serta memberikan dukungan dan semangat yang berarti hingga penulis mampu menyelesaikan karya ini.

7. Kepada Rebellion Rose dan MCPR, terima kasih telah menjadi *soundtrack* perjuangan yang membakar semangat serta mengubah rasa lelah menjadi daya juang.
8. Terima kasih kepada diri penulis sendiri, atas segala kerja keras, kesabaran, dan keteguhan hati untuk tetap melangkah serta tidak menyerah dalam menghadapi setiap tantangan hingga skripsi ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Akhir kata, semoga karya ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan.

Yogyakarta, 10 April 2026

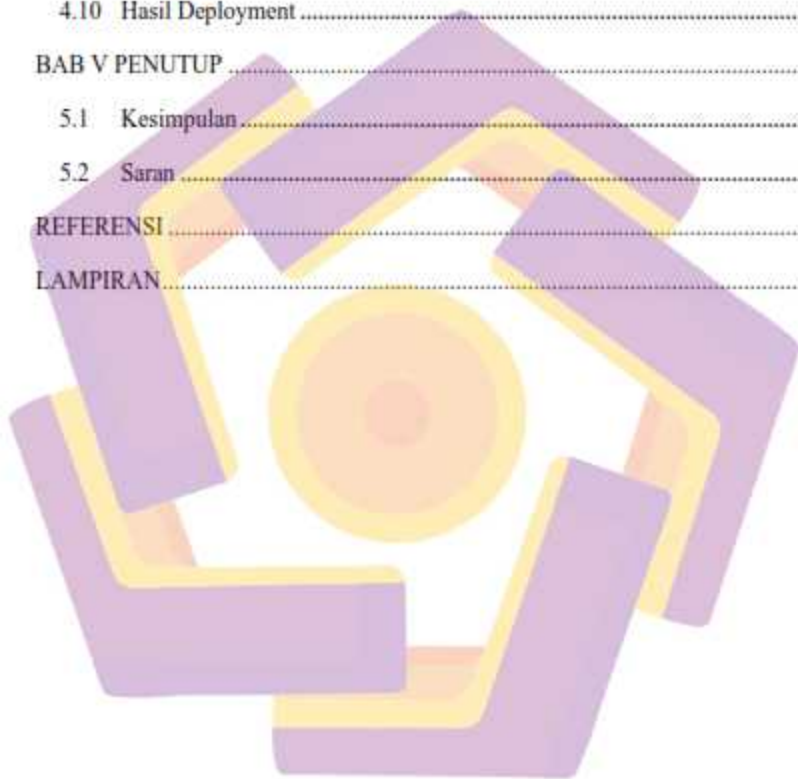
Penulis

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	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH.....	xvii
INTISARI	xxi
ABSTRACT	xxii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6

2.1	Studi Literatur	6
2.2	Dasar Teori	14
2.2.1	Kosmetik	14
2.2.2	Komposisi dan Efek Kosmetik	14
2.2.3	Klasifikasi Multilabel.....	15
2.2.4	Metode Problem Tranformation	15
2.2.5	XGBOOST.....	17
2.2.6	Feature Extraction.....	19
2.2.7	Seleksi fitur	21
2.2.8	Evaluasi performansi	23
BAB III METODE PENELITIAN		27
3.1	Objek Penelitian.....	27
3.2	Alur Penelitian	28
3.2.1	Identifikasi Masalah.....	28
3.2.2	Pengumpulan Data	28
3.2.3	Merumuskan Masalah.....	29
3.2.4	Implementasi.....	29
3.2.5	Analisis Hasil.....	36
3.2.6	Dokumentasi	36
3.3	Alat dan Bahan.....	36
3.3.1	Data Penelitian	36
3.3.2	Alat dan Instrumen.....	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Hasil Data Collection.....	38
4.2	Hasil Exploratory Data Analysis (EDA)	38
4.3	Hasil Data Preprocessing	41
4.4	Labeling	43
4.5	Feature Extraction.....	46

4.6	Feature Selection.....	48
4.7	Splitting Data	49
4.8	Modeling.....	49
4.9	Evaluation	49
4.10	Hasil Deployment	53
BAB V PENUTUP		57
5.1	Kesimpulan.....	57
5.2	Saran	58
REFERENSI		60
LAMPIRAN.....		64



DAFTAR TABEL

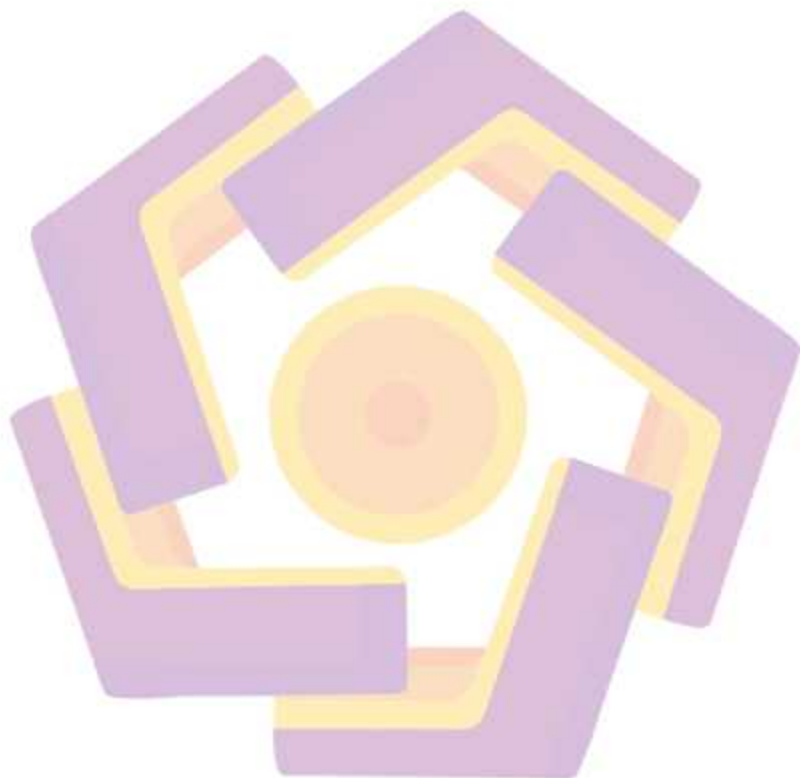
Tabel 2.1 Keaslian Penelitian	10
Tabel 3. 1 Fitur Dataset.....	30
Tabel 3. 2 Parameter TF-IDF.....	33
Tabel 3. 3 Parameter XGBoost	34
Tabel 4. 1 Jumlah Missing Value	41
Tabel 4. 2 Jumlah Sesudah dan Sebelum Penanganan Missing Value	42
Tabel 4. 6 Hasil Ekstraksi Fitur Kategorikal Menggunakan One-Hot Encoding ..	47
Tabel 4. 7 Top-10 Ingredients Paling Dominan Berdasarkan Nilai TF-IDF	47
Tabel 4. 8 Top-5 Name Paling Dominan Berdasarkan Nilai TF-IDF.....	48
Tabel 4. 9 Hasil Mean Mutual Information	48
Tabel 4. 10 Shape dataset.....	49
Tabel 4. 12 Perbandingan precision, recall, dan F1-score metode BR, CC, dan LP pada tiap label	51

DAFTAR GAMBAR

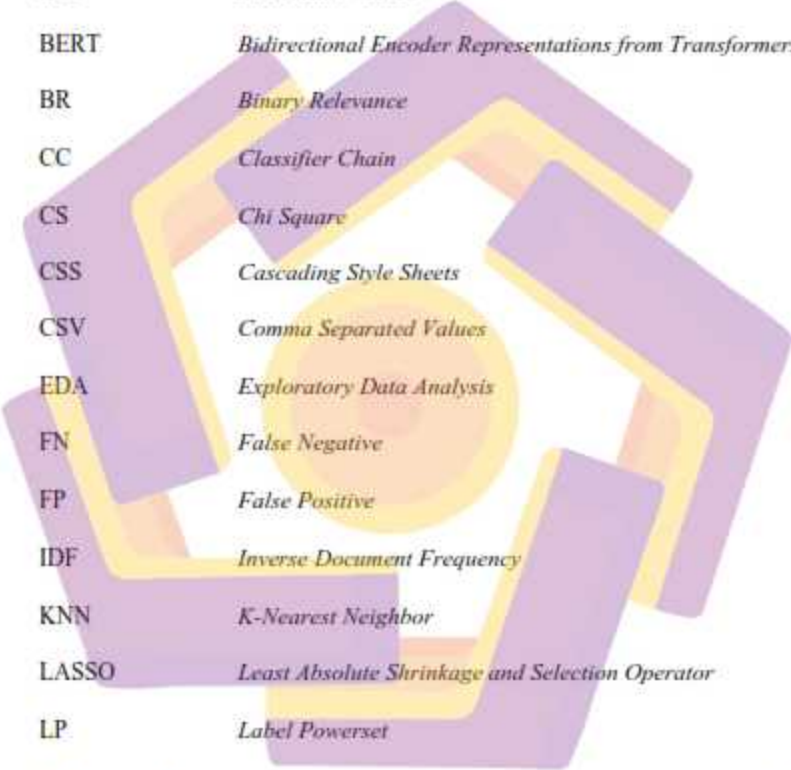
Gambar 3. 1 Diagram Alur Penelitian	28
Gambar 3. 2 Alur Implementasi Model	29
Gambar 4. 1 Isi Dataset	38
Gambar 4. 2 Informasi Dataframe	39
Gambar 4. 3 Distribusi Statistik Deskriptif	39
Gambar 4. 4 Distribusi fitur type	40
Gambar 4. 5 Representasi Data pada kolom type kategori Tool	40
Gambar 4. 6 Contoh data duplikat	41
Gambar 4. 7 Informasi Dataframe Setelah Menghapus Data Duplikat	42
Gambar 4. 8 Informasi Dataframe setelah preprocessing	43
Gambar 4. 10 Distribusi jumlah data pada setiap label efek penggunaan kosmetik	45
Gambar 4. 11 Distribusi Kelas Tiap Label (0 dan 1)	46
Gambar 4. 12 Perbandingan Performa Metode Problem Transformation	50
Gambar 4. 13 Tampilan Halaman Utama Aplikasi	54
Gambar 4. 14 Tampilan Tentang Sistem	54
.....	54
Gambar 4. 15 Tampilan Layanan sistem	54
Gambar 4. 16 Tampilan Keunggulan dan Visi Sistem	55
Gambar 4. 17 Tampilan Proses Analisis	55
Gambar 4. 18 Tampilan Hasil Analisis	56

DAFTAR LAMPIRAN

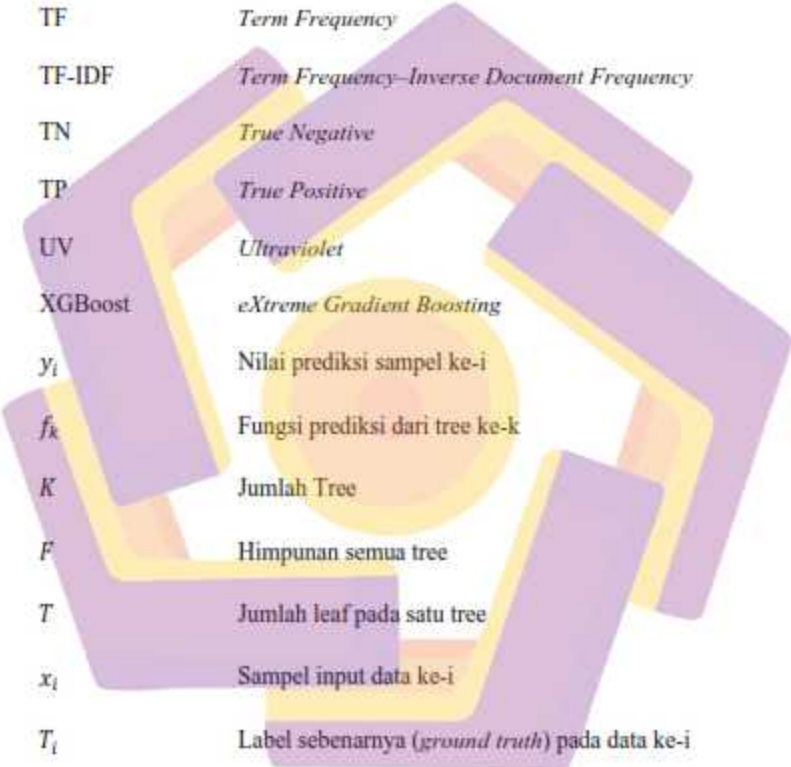
Lampiran 1. Link Dataset	64
Lampiran 2. Deploy	64



DAFTAR LAMBANG DAN SINGKATAN



AHA	<i>Asam Alpha Hidroksil</i>
API	<i>Application Programming Interface</i>
AUC	<i>Area Under Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BR	<i>Binary Relevance</i>
CC	<i>Classifier Chain</i>
CS	<i>Chi Square</i>
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma Separated Values</i>
EDA	<i>Exploratory Data Analysis</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
IDF	<i>Inverse Document Frequency</i>
KNN	<i>K-Nearest Neighbor</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LP	<i>Label Powerset</i>
MALDI-TOF MS	<i>Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry</i>
MI	<i>Mutual Information</i>
NaN	<i>Not a Number</i>
NLP	<i>Natural Language Processing</i>

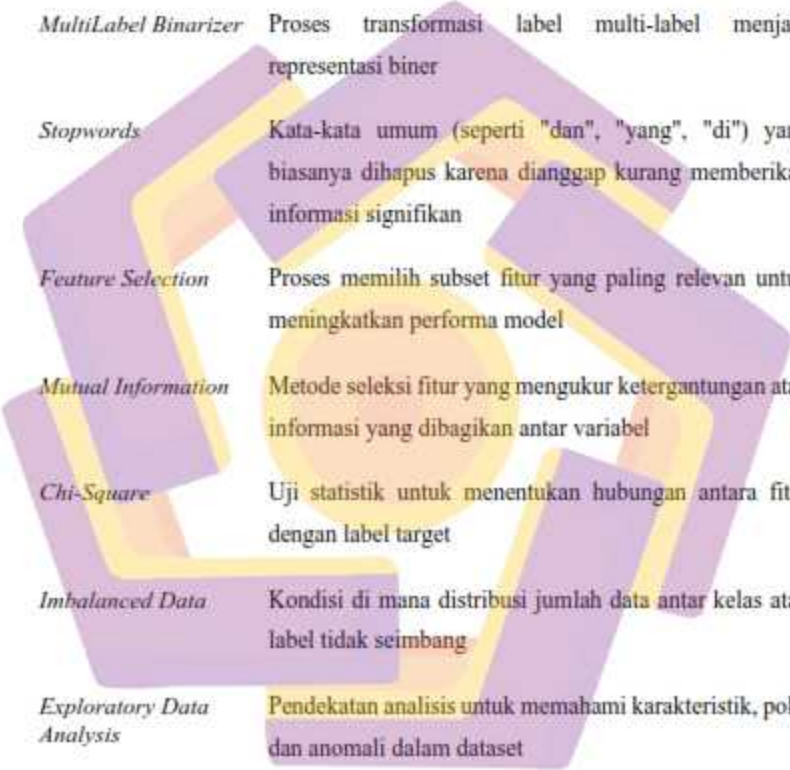


RAM	<i>Random Access Memory</i>
RF	<i>Random Forest</i>
SSD	<i>Solid State Drive</i>
SVM	<i>Support Vector Machine</i>
TF	<i>Term Frequency</i>
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
UV	<i>Ultraviolet</i>
XGBoost	<i>eXtreme Gradient Boosting</i>
y_i	Nilai prediksi sampel ke- i
f_k	Fungsi prediksi dari tree ke- k
K	Jumlah Tree
F	Himpunan semua tree
T	Jumlah leaf pada satu tree
x_i	Sampel input data ke- i
T_i	Label sebenarnya (<i>ground truth</i>) pada data ke- i
P_i	Label hasil prediksi pada data ke- i
w_i	Bobot proporsi jumlah <i>instance</i> pada kelas ke- i untuk perhitungan <i>Weighted F1-score</i>
$W_{t,d}$	Bobot TF-IDF akhir untuk kata tertentu dalam dokumen
$f_{t,d}$	Jumlah kemunculan term pada dokumen

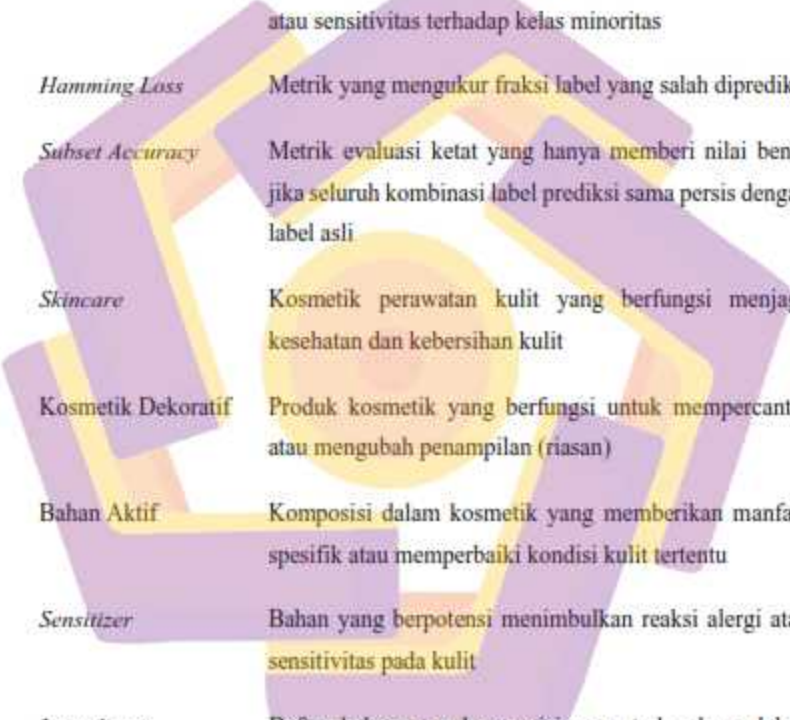
N	Total jumlah dokumen dalam korpus
df_t	Jumlah dokumen yang mengandung term
$kf_{k,d}$	Total frekuensi seluruh term dalam dokumen
$Obj(\theta)$	<i>Objective function</i> (fungsi tujuan) yang diminimalkan model
L	<i>Loss function</i> untuk mengukur kesalahan prediksi
λ	Parameter regularisasi ridge regression
Ω	<i>Regularization function</i> untuk mengontrol kompleksitas model
γ	Parameter batas minimum gain untuk pemecahan node
ω	Bobot prediksi pada <i>leaf</i>
Δ	Fungsi XOR (<i>exclusive disjunction</i>) yang menghitung perbedaan label antara prediksi dan label sebenarnya
Σ	Notasi penjumlahan total

DAFTAR ISTILAH

<i>Problem Transformation</i>	Pendekatan untuk menyelesaikan klasifikasi multi-label dengan mengubahnya menjadi satu atau beberapa masalah klasifikasi satu-label
<i>Binary Relevance</i>	Metode transformasi yang melatih satu classifier biner secara mandiri untuk setiap label
<i>Classifier Chain</i>	Metode transformasi yang melatih classifier dalam bentuk rantai, di mana hasil prediksi label sebelumnya digunakan sebagai fitur tambahan untuk label berikutnya
<i>Label Powerset</i>	Metode transformasi yang menganggap setiap kombinasi label unik sebagai satu kelas baru dalam klasifikasi tunggal
<i>K-Nearest Neighbor</i>	Algoritma klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas label dari tetangga terdekatnya
<i>Random Forest</i>	Algoritma pembelajaran mesin yang terdiri dari banyak pohon keputusan (decision trees)
<i>Support Vector Machine</i>	Algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan antar kelas
<i>Base Classifier</i>	Model dasar yang digunakan dalam metode problem transformation untuk melakukan prediksi
<i>Multi-label Classification</i>	Klasifikasi di mana satu contoh data (instance) dapat memiliki lebih dari satu label secara bersamaan
<i>Ensemble Method</i>	Teknik yang menggabungkan beberapa model prediktif untuk meningkatkan kinerja keseluruhan



<i>Feature Extraction</i>	Proses mengubah data teks atau mentah menjadi representasi numerik
<i>One Hot Encoding</i>	Teknik mengubah variabel kategorikal menjadi vektor biner (0 dan 1)
<i>MultiLabel Binarizer</i>	Proses transformasi label multi-label menjadi representasi biner
<i>Stopwords</i>	Kata-kata umum (seperti "dan", "yang", "di") yang biasanya dihapus karena dianggap kurang memberikan informasi signifikan
<i>Feature Selection</i>	Proses memilih subset fitur yang paling relevan untuk meningkatkan performa model
<i>Mutual Information</i>	Metode seleksi fitur yang mengukur ketergantungan atau informasi yang dibagikan antar variabel
<i>Chi-Square</i>	Uji statistik untuk menentukan hubungan antara fitur dengan label target
<i>Imbalanced Data</i>	Kondisi di mana distribusi jumlah data antar kelas atau label tidak seimbang
<i>Exploratory Data Analysis</i>	Pendekatan analisis untuk memahami karakteristik, pola, dan anomali dalam dataset
<i>Confusion Matrix</i>	Tabel yang membandingkan label aktual dengan label prediksi
<i>Precision</i>	Rasio prediksi positif yang benar dibandingkan dengan total prediksi positif



<i>Recall</i>	Rasio prediksi positif yang benar dibandingkan dengan total data yang sebenarnya positif
<i>F1-Score</i>	Nilai rata-rata harmonik antara precision dan recall
<i>Micro, Macro, & Weighted F1-Score</i>	Berbagai cara menghitung rata-rata F1-score pada data multi-label/multi-kelas untuk melihat performa global atau sensitivitas terhadap kelas minoritas
<i>Hamming Loss</i>	Metrik yang mengukur fraksi label yang salah diprediksi
<i>Subset Accuracy</i>	Metrik evaluasi ketat yang hanya memberi nilai benar jika seluruh kombinasi label prediksi sama persis dengan label asli
<i>Skincare</i>	Kosmetik perawatan kulit yang berfungsi menjaga kesehatan dan kebersihan kulit
Kosmetik Dekoratif	Produk kosmetik yang berfungsi untuk mempercantik atau mengubah penampilan (riasan)
Bahan Aktif	Komposisi dalam kosmetik yang memberikan manfaat spesifik atau memperbaiki kondisi kulit tertentu
<i>Sensitizer</i>	Bahan yang berpotensi menimbulkan reaksi alergi atau sensitivitas pada kulit
<i>Ingredients</i>	Daftar bahan atau komposisi yang terkandung dalam produk kosmetik
Data Sekunder	Data yang diperoleh dari sumber pihak ketiga yang sudah ada (misalnya dari Kaggle)
<i>Overfitting</i>	Kondisi di mana model terlalu menyesuaikan dengan data latih sehingga performanya buruk pada data baru



<i>Underfitting</i>	Kondisi di mana model terlalu sederhana sehingga tidak mampu menangkap pola data dengan baik
<i>Hyperparameter Tuning</i>	Proses mencari konfigurasi parameter terbaik untuk mengoptimalkan performa model
<i>Scikit-learn</i>	Pustaka Python yang digunakan untuk machine learning
<i>Skmultilearn</i>	Pustaka Python khusus yang digunakan untuk mengimplementasikan metode problem transformation dalam klasifikasi multi-label
<i>N-gram Range</i>	Parameter TF-IDF yang menentukan penggunaan unigram (satu kata tunggal) dan bigram (dua kata berurutan)
<i>Max Features</i>	Batas jumlah fitur kata yang diambil dalam proses TF-IDF
<i>Learning Rate</i>	Parameter XGBoost yang mengontrol kecepatan model dalam mempelajari pola
<i>N-estimators</i>	Jumlah pohon keputusan (trees) yang dibangun dalam model XGBoost
<i>Max Depth</i>	Kedalaman maksimum setiap pohon keputusan dalam model
<i>Random State</i>	Parameter untuk memastikan hasil pembagian data atau pelatihan model tetap konsisten (reproducible)

INTISARI

Ketertarikan masyarakat terhadap penggunaan kosmetik sangat tinggi, namun konsumen masih kesulitan memahami kandungan produk dan efeknya. Di Indonesia, sekitar 30% pengguna tidak mengetahui komposisi produk yang digunakan, sehingga berpotensi menimbulkan efek samping seperti jerawat, iritasi atau kemerahan. Permasalahan ini semakin kompleks karena satu produk kosmetik dapat memiliki berbagai kombinasi manfaat dan efek samping secara bersamaan, sehingga diperlukan pendekatan klasifikasi multi-label untuk mengidentifikasi berbagai kategori efek secara lebih akurat.

Penelitian ini membandingkan tiga metode *problem transformation*, yaitu *Binary Relevance*, *Classifier Chain*, dan *Label Powerset* menggunakan algoritma *XGBoost* sebagai *base classifier*. Tahapan penelitian meliputi pengumpulan data dari Kaggle, prapemrosesan data, ekstraksi fitur menggunakan *TF-IDF* dan *One-Hot Encoding*, serta seleksi fitur dengan *Mutual Information* dan *Chi-Square*. Evaluasi performa model dilakukan menggunakan metrik *Micro F1-Score*, *Macro F1-Score*, *Weighted F1-Score*, *Hamming Loss*, dan *Subset Accuracy*.

Hasil penelitian menunjukkan bahwa metode *Binary Relevance* dan *Classifier Chain* memberikan performa terbaik dan stabil dengan *F1-Score Micro* dan *F1-Score Weighted* sebesar 94%. Pada *F1-Score Macro*, *Binary Relevance* (94%) sedikit lebih unggul dibandingkan *Classifier Chain* (93%). Kedua metode juga memiliki *Hamming Loss* rendah (5%) dan *Subset Accuracy* sebesar 64%, sedangkan *Label Powerset* menunjukkan performa lebih rendah. Penelitian ini dapat dimanfaatkan oleh konsumen, industri kosmetik, dan pengembang sistem rekomendasi dalam memahami efek penggunaan produk secara lebih akurat. Penelitian selanjutnya dapat menggunakan hyperparameter tuning dan deep learning untuk meningkatkan performa pada data yang lebih kompleks.

Kata kunci: Problem Transformation, XGBoost, Klasifikasi Multi-Label, Kosmetik.

ABSTRACT

The public's interest in using cosmetics is very high, but consumers still struggle to understand the product's ingredients and functions. In Indonesia, around 30% of users do not know the composition of the products they use, which can potentially cause side effects such as acne, irritation, or redness. This issue becomes even more complex because a single cosmetic product can have various combinations of benefits and side effects simultaneously, necessitating a multi-label classification approach to more accurately identify the different categories of effects.

This study compares three problem transformation methods, namely Binary Relevance, Classifier Chain, and Label Powerset, using the XGBoost algorithm as the base classifier. The stages of the research include data collection from Kaggle, data preprocessing, feature extraction using TF-IDF and One-Hot Encoding, and feature selection with Mutual Information and Chi-Square. Model performance evaluation was conducted using the metrics Micro F1-Score, Macro F1-Score, Weighted F1-Score, Hamming Loss, and Subset Accuracy.

The research results indicate that the Binary Relevance and Classifier Chain methods provide the best and most stable performance with a Micro F1-Score and Weighted F1-Score of 94%. In terms of Macro F1-Score, Binary Relevance (94%) slightly outperformed Classifier Chain (93%). Both methods also have a low Hamming Loss (5%) and a Subset Accuracy of 64%, while the Label Powerset method showed lower performance. This research can be utilized by consumers, the cosmetics industry, and recommendation system developers to understand the effects of product usage more accurately. Future research can utilize hyperparameter tuning and deep learning to improve performance on more complex data.

Keyword: Problem Transformation, XGBoost, Multi-Label Classification, Cosmetics.