

**Analisis Sentimen Menggunakan Metode Random forest Dan
TF-IDF Terhadap Ulasan Produk Kecantikan Pada Website
Female Daily**

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

dari Program Studi Informatika

Fakultas Informatika

Universitas Telkom

1301194289

Agung Prayogo



Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2024

LEMBAR PENGESAHAN

**Analisis Sentimen Menggunakan Metode Random Forest Dan TF-IDF Terhadap Ulasan Produk
Kecantikan Pada Website Female Daily**

**Sentiment Analysis Using Random Forest dan TF-IDF Method Of Beauty Product Reviews On
Female Daily Website**

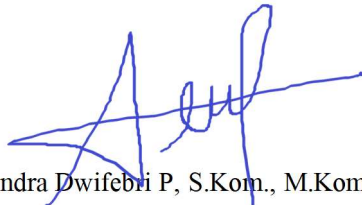
NIM : 1301194289

Agung Prayogo

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada
Program Studi Sarjana Informatika
Fakultas Informatika Universitas
Telkom

Bandung,
Menyetujui

Pembimbing I,



Mahendra Dwifebri P, S.Kom., M.Kom.

NIP. 20940009-3

Pembimbing II,



Widi Astuti, S.T.,M.Kom.

NIP.19910018

Ketua Program Studi
Sarjana Informatika,



Dr. Erwin Budi Setiawan, S.Si., M.T.

NIP: 00760045

LEMBAR PERNYATAAN

Dengan ini saya, Agung Prayogo, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan Analisis Sentimen Terhadap Ulasan Produk Kecantikan Pada Website *Female Daily* Menggunakan Metode *Random Forest* dan *TF-IDF* beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 11 Januari 2024

Yang Menyatakan



Agung Prayogo

Analisis Sentimen Menggunakan Random Forest Dan TF-IDF Terhadap Ulasan Produk Kecantikan Pada Website Female Daily

Abstrak

Pengguna internet Indonesia menggunakan layanan e-commerce untuk membeli berbagai barang. Female Daily adalah salah satu website yang menyediakan berbagai ulasan produk kecantikan. Salah satu fitur yang sangat membantu pelanggan saat membeli barang yang akan mereka beli adalah ulasan yang tersedia di website tersebut. Banyak ulasan akan membutuhkan waktu lama untuk dibaca, dan pengguna hampir tidak mungkin membaca keseluruhan informasi. Oleh karena itu, penelitian diperlukan untuk membantu konsumen mempertimbangkan produk seperti analisis sentimen. Analisis sentiment ini secara otomatis membagi komentar pengguna menjadi kategori positif, netral, dan negatif. Dalam penelitian ini, fitur ekstraksi digunakan untuk sentimen analisis. Pada penelitian ini, fitur Ekstraksi yang digunakan untuk sentimen analisis menggunakan metode Random Forest dan TF-IDF. Penelitian ini menghasilkan nilai terbaik dengan rata-rata accuracy 81,75% dan f1-score 71,90%

Kata kunci : *review* produk, analisis sentimen, Random Forest, TF-IDF

Abstract

Indonesian internet users use e-commerce services to buy various goods. Female Daily is a website that provides various beauty product reviews. One feature that really helps customers when purchasing the items they want to buy is the reviews available on the website. Many reviews will take a long time to read, and it is almost impossible for users to read the entire information. Therefore, research is needed to help consumers consider products such as sentiment analysis. This sentiment analysis automatically divides user comments into positive, neutral and negative categories. In this research, extraction features are used for sentiment analysis. In this research, extraction features are used for sentiment analysis using the Random Forest and TF-IDF methods. This research produced the best scores with an average accuracy of 81.75% and f1-score of 71.90%

Keywords: product reviews, sentiment analysis, Random Forest, TF-IDF

1. Pendahuluan

Latar Belakang

Tahun 2021, OJK mencatat 88,1 persen pengguna internet di Indonesia telah menggunakan layanan *e-commerce* dalam membeli berbagai produk. Angka tersebut didapatkan dari survei We Are Social. Survei tersebut menunjukkan bahwa negara Indonesia berada di peringkat pertama dari segi layanan *e-commerce* [1]. Website Female Daily adalah Salah satu situs web yang menawarkan informasi tentang produk kecantikan adalah. Pada website <https://reviews.femaledaily.com/>, lebih dari 30.000 produk kecantikan dan tercatat terdapat 300.000 *reviews*. Fitur *review* merupakan salah satu fitur yang sangat membantu pelanggan dalam memilih produk kecantikan untuk digunakan dalam *daily beauty care* pengguna[2]. Karena terlalu banyak Ulasan akan membutuhkan banyak waktu untuk dibaca, dan pengguna hampir pasti tidak akan dapat membaca keseluruhan informasi. [3]. Sehingga diperlukan proses analisis sentiment.

Analisis sentimen atau dikenal dengan *opinion mining* merupakan bagian ilmu yang menganalisis pendapat orang, sentimen, penilaian, sikap, dan emosi[4]. Klasifikasi terhadap sentimen memiliki tujuan untuk menemukan solusi dari masalah dengan cara mengelompokkan pendapat atau ulasan pengguna menjadi opini positif, netral, atau negatif.

Sebelumnya telah dilakukan penelitian terdahulu yang berkaitan dengan topik serupa. Analisis sentimen pada ulasan produk konsumen bahasa Indonesia menggunakan metode *Naïve Bayes* dan TF-IDF. Hasil dari penelitian tersebut mampu menghasilkan nilai akurasi 82% [5]. Penelitian lainnya adalah analisis sentimen menggunakan *Random Forest Ensemble* untuk review mobile product di Kanada. Dari penelitian tersebut menghasilkan bahwa *Random Forest* merupakan classifier yang bagus untuk mengklasifikasikan multiclass dengan menghasilkan nilai accuracy 72% [6]. Dalam penelitian [7] metode *Chi-square* berfungsi untuk memberikan nilai pada kata atau fitur. Fitur yang terpilih selanjutnya digunakan untuk tahap klasifikasi. Hasil dari penelitian ini mampu meningkatkan performansi dari model yang dibangun. Pada penelitian sebelumnya terdapat pendekatan baru yaitu *Random Forest*. Metode tersebut tidak hanya meningkatkan akurasi tetapi juga kompleksitas waktu [8]. Klasifikasi *Random Forest* menghasilkan performa sistem yang sangat bagus untuk dataset ulasan *movie* berbahasa Inggris. RF dan *Mutual Information* sebagai seleksi fitur dikombinasikan sehingga menghasilkan nilai akurasi dan nilai *f1-scores* dengan nilai masing-masing sebesar 79% dan 75% [9].

Pada penelitian ini penulis menggunakan metode *Random Forest* dan *TF-IDF* sebagai Classification dan Ekstraksi fitur dalam menganalisis sentimen terhadap ulasan produk kecantikan pada website *Female Daily*.

Sebelum melakukan klasifikasi dengan banyak fitur, menyeleksi fitur sangat diperlukan untuk meningkatkan kinerja sistem. *TF-IDF* dapat meningkatkan kinerja sistem dalam studi sebelumnya[7]. Klasifikasi *Random Forest* untuk mengklasifikasikan data yang imbalanced dan metode tersebut memiliki kinerja yang sangat baik pada penelitian sebelumnya [8] [9]. *TF-IDF* dipilih menggantikan *Word2vec*, karena *Word2Vec* tidak terlalu cocok dengan penelitian yang dibuat. Oleh sebab itu *TF-IDF* digunakan agar bisa meningkatkan akurasi dan relevansi hasil pencarian[17].

Topik dan Batasannya

Topik dan batasan masalah pada penelitian ini adalah pengujian untuk mengetahui performansi dari model berdasarkan skenario pengujian. Batasan masalah untuk penelitian adalah dilakukan analisis sentimen menggunakan metode *Random Forest* dan *TF-IDF* sebagai seleksi fitur. Data yang digunakan adalah data review produk kecantikan berbahasa Indonesia yang berasal dari website *Female Daily*, dengan jumlah 3959 data dengan 4 level aspek, yaitu harga, kemasan, produk dan aroma. Pada setiap aspek memiliki kelas yaitu positif, netral dan negatif.

Tujuan

Tujuan dari penelitian tugas akhir ini berdasarkan latar belakang dan perumusan masalah yaitu melakukan analisis sistem pada ulasan produk kecantikan dari website *female daily* menggunakan metode *Random Forest* dan *TF-IDF* sebagai seleksi fitur, kemudian mengukur hasil performansi dari sistem yang dibangun.

Organisasi Tulisan

Pada bab 2 membahas studi terkait dengan penelitian ini. Pada bab 3 membahas teori dan rancangan sistem penelitian. Hasil dan analisis dibahas pada bab 4. Kemudian pada bab 5 membahas kesimpulan dari penelitian ini.

2. Studi Terkait

Penelitian ini menggunakan beberapa referensi mengenai penelitian sebelumnya yang berkaitan dengan metode klasifikasi, seleksi fitur, preprocessing, ekstraksi fitur dan evaluasi. Banyak penelitian tentang analisis sentimen serupa seperti review produk, rumah makan, film dan data lainnya.

Pada penelitian[8] oleh Trian Basofi Rohman, Devi Dwi Purwanto, Joan Santoso dengan judul “Sentiment Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest” mengusulkan sistem yang mampu mengelompokkan review rumah makan ke sebuah kelas sentiment Positif, Negatif, ataupun Netral. Penelitian ini menggunakan sebuah metode preprocessing pemecahan kalimat, replacement, POS Tagging, Content Words, Case Folding, Remove Stopword, Stemming, serta Convert Negation. Lalu tahap selanjutnya yaitu pembobotan term menggunakan *TF-IDF*, terakhir menganalisis sentimen dengan metode *Random Forest*. Hasil yang diperoleh cukup baik dengan nilai akurasi 74%, precision 62%, dan recall 55%.

Pada penelitian berikutnya [9] Hana Chyntia Morama, Dian Eka Ratnawati, Issa Arwani dengan judul “Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier”. Untuk melakukan klasifikasi, algoritma *Random Forest Classifier* dan pembobotan kata Term Frequency-Inverse Document Frequency (*TF-IDF*) digunakan. Setelah analisis pengujian selesai, hanya elemen kamar yang digunakan karena memiliki proporsi sentimen yang seimbang dibandingkan dengan elemen lainnya. Proporsi sentimen dianggap penting untuk klasifikasi sentimen. Jumlah pohon dalam studi ini adalah 300, dengan 10 kedalaman pohon. Situasi ini digunakan untuk menguji. Hasil pengujian menunjukkan bahwa hasil prediksi lebih baik dengan jumlah pohon dan kedalaman pohon yang lebih besar. Hasil klasifikasi terbaik kedua parameter terhadap aspek kamar yaitu 90% untuk nilai akurasi dan skor *f1*-nya..

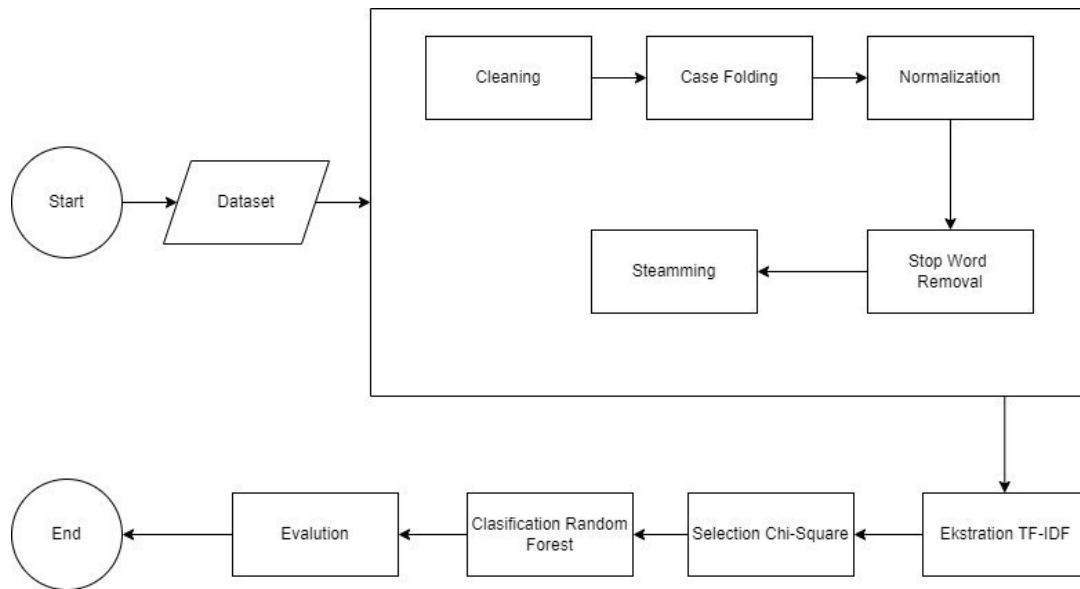
Penelitian [10] melakukan analisis sentimen terhadap Bahasa Indonesia. Digunakan metode *Random Forest* dengan 3 tahap yang dilakukan yaitu *preprocessing*, *features extraction* dan *classification*. Terdapat 2 kategori dari hasil yang akan diuji diantaranya adalah positif dan negatif. Terbukti dari hasil pengujian bahwa *Random Forest* berhasil membawa performa yang baik dengan *average OBB* 0.829.

Pada penelitian [11] oleh Okta Ihza Gifari, Muh. Adha, Ivan Rifky Hendrawan, Fernandito Freddy Setlight Durrand, dengan judul “Analisis Sentimen Review Film Menggunakan *TF-IDF* dan *Support Vector Machine*” Algoritma *TF-IDF* dan *svm* berhasil dalam penelitian ini. Hasil pengujian dan analisis menunjukkan bahwa algoritma *SVM* dapat digunakan untuk mengklasifikasikan data opini film berbahasa Indonesia berdasarkan sentimen. Semakin banyak data pelatihan yang digunakan, semakin baik kinerja sistem. Hal ini dibuktikan dengan hasil nilai *accuracy* 0.85, nilai *precision* 1.0, nilai *recall* 0.7, dan nilai *F1-Score*

0.82. Hal ini sudah bisa dikatakan baik. Untuk penelitian kedepan disarankan adanya proses pengoptimalan sebuah tweet, agar kata tidak baku atau singkatan-singkatan bisa melewati proses preprocessing secara baik.

3. Sistem yang Dibangun

Gambar 1. Desain Sistem



Sistem yang dibangun pada penelitian ini adalah sistem yang menganalisa sentimen terhadap ulasan produk kecantikan pada website *Female Daily*.

Fokus penelitian ini adalah mengembangkan model analisis sentimen untuk menganalisis ulasan pengguna terhadap aplikasi Practo. Pendekatan yang dipilih melibatkan pemanfaatan algoritma Random Forest bersama dengan fitur TF-IDF ekstraksi. Diagram yang menguraikan proses model ini digambarkan pada “Gambar 1”

3.1. Dataset

Data didapatkan dari situs web <https://femaledaily.com/>. Pada dataset ini terdapat 3959 data dan terdapat 4 level aspek (*price*, *packaging*, *product*, dan *aroma*) yang sudah dilabeli secara manual. Data diklasifikasi menjadi positif, negatif, maupun netral berdasarkan kriteria.

Tabel 1. Pelabelan Data

<i>review_text</i>	<i>price</i>	<i>packaging</i>	<i>product</i>	<i>aroma</i>
Pertama pake, di kulit enak banget . cocok dipake sebelum bedak sehari2.. harganya emg mahal	-1	0	1	0

Berdasarkan contoh ulasan produk kecantikan pada tabel diatas, data sudah diberi label 1 untuk ulasan yang bersifat positif, 0 untuk ulasan yang bersifat netral, sedangkan untuk ulasan yang bersifat negatif diberikan label -1. Pada ulasan tersebut tidak terdapat kata yang menunjukkan aspek kemasan dan aroma sehingga aspek kemasan dan aroma diberi label 0. Selanjutnya pada ulasan tersebut terdapat kata “mahal” yang termasuk aspek harga, dimana kata mahal merupakan kata negatif sehingga ulasan tersebut untuk aspek harga diberi label -1. Pada ulasan tersebut terdapat juga kata “enak banget” yang termasuk aspek produk, dimana kata tersebut menunjukkan kata positif sehingga ulasan tersebut untuk aspek produk diberi label 1.

3.2. Preprocessing

Preprocessing adalah tahap yang dilakukan text mining, dimana data yang tidak penting dan tidak diperlukan akan dihilangkan. Tahapan *preprocessing* yang dilakukan adalah data *cleaning*, *case folding*, *normalization*, *stopword removal*, dan *stemming*.

Cleaning merupakan tahapan membersihkan data yang mengandung tanda baca, angka, simbol dan lainnya [13].

Tabel 2. Cleaning Data

<i>review_text</i>	Hasil
Pertama pake, di kulit enak banget. cocok dipake sebelum bedak sehari2..harganya emg mahal	Pertama pake di kulit enak banget cocok dipake sebelum bedak sehari harganya emg mahal

Case Folding merupakan merupakan mengubah huruf kapital menjadi huruf kecil yang terdapat pada data ulasan agar tidak ada duplikasi term yang hanya dibedakan oleh bentuk hurufnya [10].

Tabel 3. Case Folding

<i>review_text</i>	Hasil
Pertama pake di kulit enak banget cocok dipake sebelum bedak sehari harganya emg mahal	pertama pake di kulit enak banget cocok dipake sebelum bedak sehari harganya emg mahal

Normalization merupakan tahapan menggantikan kata yang tidak normal seperti kesalahan penulisan, tidak baku dan sebagainya.

Tabel 4. Normalization Data

<i>review_text</i>	Hasil
pertama pake di kulit enak banget cocok dipake sebelum bedak sehari harganya emg mahal	pertama pakai di kulit enak banget cocok dipakai sebelum bedak sehari harganya memang mahal

Dengan menggunakan file kamus yang bernama *meaningless.csv.*, *stopword* digunakan untuk memilah (filter) kata-kata umum atau sering digunakan yang sebenarnya tidak bermakna dan kemudian menghapusnya. Ini terjadi dalam bahasa Indonesia dengan kata-kata seperti dan, atau, yang, dari, ke kemudian, untuk, dan sebagainya

Tabel 5. Stopword Removal

<i>review_text</i>	Hasil
pertama pakai di kulit enak banget cocok dipakai sebelum bedak sehari harganya memang mahal	pakai kulit enak banget cocok dipakai bedak sehari harganya mahal

Stemming merupakan merupakan proses yang dilakukan untuk memotong imbuhan kata(awalan, akhiran, kombinasi)[10].

Tabel 6. Stemming Data

<i>review_text</i>	Hasil
pakai kulit enak banget cocok dipakai bedak sehari harganya mahal	pakai kulit enak banget cocok pakai bedak sehari harga mahal

Dari tabel diatas, dapat dilihat bawah kata yang memiliki imbuhan baik awalan, akhiran ataupun kombinasi akan dipotong. Contoh dari ulasan diatas misalkan kata “dipakai” menjadi “pakai”, kata “harganya” menjadi “harga”.

3.3. Feature Extraction TF-IDF

Setelah dilakukan tahap preprocessing dilanjutkan dengan proses pembobotan TF-IDF (Term Frequency – Inverse Document Frequency). Pembobotan ini digunakan untuk merubah kata menjadi data numerik [14]. Langkah yang perlu dilakukan yaitu menghitung TF, kemudian menghitung jumlah inverse frekuensi dapat dilihat dari Persamaan berikut :

$$w(t, d) = tf(t, d) \times idf(t) = tf(t, d) \times \log \frac{N}{df(t)} \quad (1)$$

Keterangan :

menghitung TF-IDF menggunakan Proses pembobotan TF-IDF (Term Frequency—Inverse Document Frequency) adalah suatu proses untuk mengubah data teks menjadi data numerik dengan memotong setiap kata atau fitur.

$w(t,d)$ = bobot TF-IDF atau bobot term(t) dalam dokumen(d)
 $tf(t,d)$ = jumlah kemunculan term(t) dalam dokumen(d)
 $idf(t)$ = jumlah inverse frekuensi dokumen tiap kata
 $df(t)$ = jumlah frekuensi dokumen tiap kata
 N = jumlah total dokumen

IDF berfungsi mengurangi bobot suatu *term* jika kemunculannya banyak tersebar diseluruh dokumen

3.4. Feature Selection Chi-Square

Data yang telah melalui proses ekstraksi dengan TF-IDF akan dilakukan seleksi fitur. Metode seleksi fitur *Chi-square* dapat meningkatkan kinerja sistem pada beberapa penelitian sebelumnya [7] [11] [12][15]. Fungsi metode ini yaitu untuk menguji independensi sebuah kata atau fitur terhadap kategorinya serta untuk menghilangkan fitur pengganggu sebelum dilakukan klasifikasi. *Chi-square* memiliki 3 uji, yaitu frekuensi pengamatan, frekuensi harapan, dan *test* [7]. Langkah pertama yaitu dilakukan perhitungan dengan rumus pada Persamaan berikut :

$$e_{ij} = \frac{o_i \cdot o_j}{N} \quad (2)$$

Keterangan :

e_{ij} = *expected frequency*
 o_i = frekuensi yang diharapkan kolom marginal
 o_j = frekuensi yang diharapkan baris marginal
 N = jumlah sample

$$x^2 = \sum_{i=1}^n \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \quad (3)$$

Ketereangan :

x^2 = *Chi-square*
 o_{ij} = *Observed Frequency*
 e_{ij} = *Expected Frequency*

Menggunakan `sklearn.feature_selection import chi2` menghitung nilai chi-square Menggunakan `sklearn.feature_selection import SelectKBest` untuk memilih k feature $x^2 = \sum (\text{Hasil yang diamati} - \text{Hasil yang diharapkan})^2 / \text{Hasil yang diharapkan}$ Sehingga Semakin tinggi nilai chi-square maka semakin banyak informasi dari

label yang dibawa oleh fitur tersebut dan semakin relevan fitur tersebut dengan label

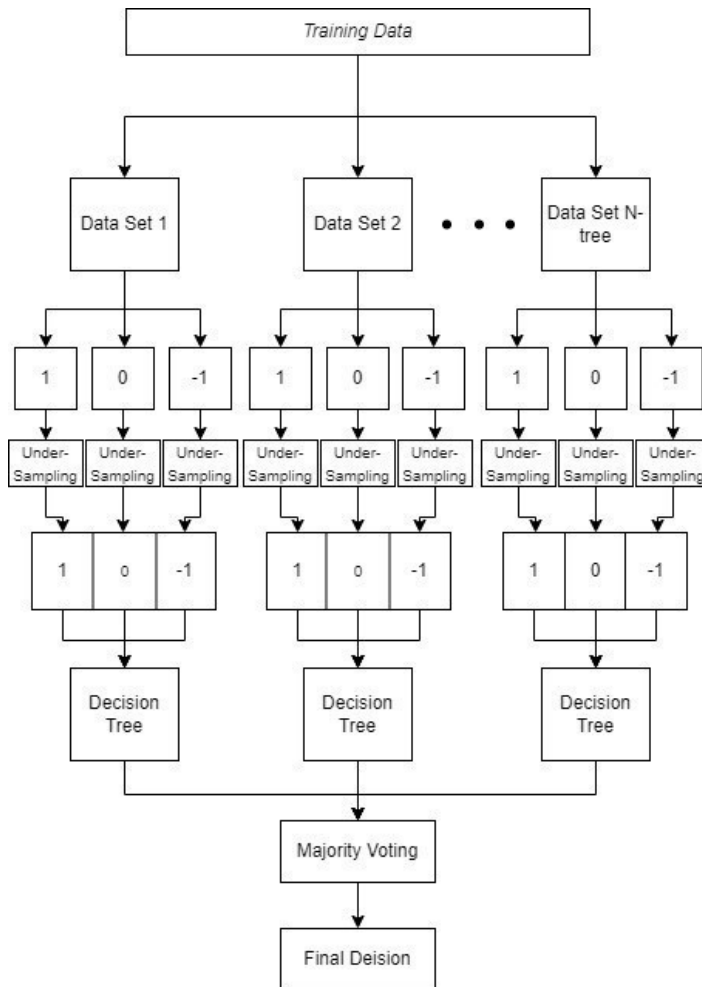
Pada penelitian ini, data yang digunakan memiliki 8294 fitur. Dan fitur yang diambil yaitu 5000 fitur terbaik dari hasil nilai *chi-square*. Kemudian data tersebut dilakukan klasifikasi.

3.5. Data Splitting

Data pada penelitian ini dibagi menjadi *data train* dan *data test*. Presentase pembagian datanya yaitu *data train* 80% dan *data test* 20%. Pembagian data juga dilakukan dengan metode *k-fold*. *K-Fold cross-validation* merupakan proses pengujian dengan tujuan untuk menilai kinerja dari model yang dibangun. Tahap ini bekerja dengan membagi data secara acak dan membagi data sebanyak *k*-partisi. Untuk jumlah *k* yang digunakan pada penelitian ini yaitu $k=5$.

3.6. Classification Random Forest

Penelitian ini menggunakan metode Random Forest.. BRF menangani data *imbalanced* dengan menerapkan *undersampling* untuk setiap formasi decision tree dari *Random Forest*(RF). Metode Random Forest meningkatkan akurasi dan mengurangi kompleksitas waktu. Pada metode ini diterapkan *undersampling* dikarenakan dataset yang digunakan merupakan data *imbalanced*.



Gambar 2. Alur Kerja Random Forest[8]

Pada gambar diatas menunjukkan bahwa langkah pertama yaitu membaca data latih. Selanjutnya, sampling untuk penentuan data yang menjadi tree sejumlah N data. Pada setiap pembentukan pohon, undersampling dilakukan dengan memilih kelas mayoritas dan dihapus secara acak sehingga jumlah data seimbang dengan kelas minoritas. Dalam penelitian ini ada empat aspek. Setiap aspek memiliki distribusi kelas yang berbeda. Jadi, undersampling di setiap aspek berbeda. Misalnya, "1" dan "0" adalah kelas mayoritas dalam aspek harga. Sedangkan "-1" dan "0" merupakan kelas mayoritas pada aspek produk. . Dan kemudian, data diberi label ulang untuk melalui proses decision tree. Setelah itu pemroses dengan *Decision Tree*. Pemrosesan akan berlangsung sampai sebanyak N-tree. Terakhir setelah hasil *decision tree* keluar, selanjutnya dilakukan *majority voting* untuk mendapatkan hasilnya. Kemudian dilakukan evaluasi untuk pengukuran performance dari metode ini.

3.7. Evaluation

Evaluasi dilakukan untuk mengukur performansi dari model sistem dengan *confusion matrix*. *Confusion matrix* digunakan untuk memberikan informasi dari hasil klasifikasi berupa kelas aktual dan kelas prediksi. Hasil performansi dari sistem yang dibangun dievaluasi dengan metrik data[16]. Berikut ilustrasi table dari *confusion matrix*.

Tabel 7. Confusion Matrix [16]

<i>confusion matrix</i>		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	TP	FP
	Negatif	FN	TN

Keterangan:

TP = data prediksi positif dan data sebenarnya positif

FN = data prediksi negatif dan data sebenarnya positif

FP = data prediksi positif dan data sebenarnya negatif

TN = data prediksi negatif dan data sebenarnya negatif

Untuk mengukur performa dari model sistem dalam menghitung hasil prediksi dari data uji dilakukan perhitungan *accuracy*, *recall*, *precision*, dan *f1-score*. Berikut rumus-rumus untuk menghitung evaluasi:

Accuracy adalah rasio prediksi benar (positif dan negatif) dengan keseluruhan data [17].

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

Recall adalah rasio prediksi benar positif dengan keseluruhan prediksi positif[17].

$$recall = \frac{TP}{TP + FN} \quad (2)$$

Precision adalah proporsi dari memprediksi kasus positif yang benar[17].

$$precision = \frac{TP}{TP + FP} \quad (6)$$

F1-Score adalah perbandingan rata-rata antara presisi dan recall[17].

$$f1 - score = 2 x \frac{Recall \times Precision}{Recall + Precision} \quad (7)$$

4. Evaluasi

Pada tahap evaluasi ini dilakukan pengujian sistem untuk melihat hasil performansi dari model yang dibangun dengan melakukan beberapa skenario pengujian. Perbandingan performansi dari setiap skenario dilihat dari rata-rata akurasi dari seluruh level aspek. Skenario pertama yaitu pada tahap *preprocessing* dilakukan pengujian dengan tahap stemming dan tidak dengan tahap stemming. Selanjutnya, skenario kedua yaitu pada tahap *feature selection* dilakukan pengujian dengan tahap seleksi fitur dan tidak dengan tahap seleksi fitur. Pada skenario ketiga yaitu tahap pengujian dilakukan dengan menggunakan *k-fold cross-validation* dan tanpa *k-fold cross-validation*.

Selanjutnya peneliti mengambil model terbaik dari nilai terbaik rata-rata akurasi dan *f1-score*.

4.1 Pengujian Skenario 1

Pengujian skenario 1 dengan melakukan pengujian pada tahap *preprocessing*, dimana pengujian ini bertujuan melihat pengaruh dari dua dataset yang digunakan yaitu data yang melalui tahap *stemming* dan tanpa *stemming*. Pada pengujian ini menggunakan dataset yang melalui tahap *feature selection*. Hasil dari pengujian skenario 1 dapat dilihat pada tabel berikut :

Tabel 8. Pengujian Skenario 1

Dataset		Accuracy	Precision	Recall	F1-Score
Stemming	Harga	93,18%	92,50%	92,08%	92,28%
	Kemasan	88,64%	67,72%	71,37%	69,17%
	Produk	58,08%	54,64%	63,40%	54,43%
	Aroma	86,49%	68,09%	77,36%	70,20%
	Rata-rata	81,60%	70,74%	76,05%	71,52%
Tanpa Stemming	Harga	92,05%	91,68%	91,08%	91,37%
	Kemasan	87,75%	66,06%	68,42%	67,11%
	Produk	57,32%	53,75%	62,24	53,55%
	Aroma	82,95%	62,10%	71,48%	64,70%
	Rata-rata	80,01%	68,40%	73,31%	64,70%

Berdasarkan hasil pengujian menggunakan skenario 1, data yang melalui tahap *stemming* pada tahap *preprocessing* mendapatkan hasil performansi lebih baik dibandingkan dengan hasil performansi tanpa tahap *stemming*. Perbedaan rata-rata nilai performansi *accuracy*, *precision* dan *recall* lebih sedikit jaraknya daripada rata-rata performansi nilai *f1-score*. Nilai *f1-score* memiliki perbedaan yang paling jauh yaitu 6,82%. Tahap *stemming* merupakan proses mengolah data dengan memotong imbuhan kata seperti awalan, akhiran maupun kombinasi. Contoh pemotongan kata pada dataset seperti kata “dipakai” atau “memakai” menjadi pakai, dengan pemotongan kata tersebut sistem lebih mudah dalam memproses kata. Sehingga hal tersebut menyebabkan peningkatan hasil rata-rata performansi dari model yang dibangun.

4.2 Pengujian Skenario 2

Pada pengujian skenario 2 dilakukan pengujian pada tahap *feature selection*, dimana pengujian ini bertujuan untuk melihat pengaruh dari dua dataset yang digunakan yaitu data yang melalui tahap *feature selection* dan tanpa *feature selection*. Pada pengujian 1 sudah didapatkan hasil terbaik yaitu dengan menggunakan dataset *stemming*, sehingga pada skenario kedua digunakan dataset yang melalui tahap *stemming*. Hasil dari pengujian skenario 2 dapat dilihat pada tabel berikut :

Tabel 9. Pengujian Skenario 2

Dataset		Accuracy	Precision	Recall	F1-Score
Feature Selection	Harga	93,18%	92,50%	92,08%	92,28%
	Kemasan	88,64%	67,72%	71,37%	69,17%
	Produk	58,08%	54,64%	63,40%	54,43%
	Aroma	86,49%	68,09%	77,36%	70,20%
	Rata-rata	81,60%	70,74%	76,05%	71,52%
Tanpa Feature	Harga	91,54%	90,40%	91,03%	90,69%
	Kemasan	86,36%	64,27%	71,56%	67,32%
	Produk	57,45%	54,93%	64,66%	54,56%

Selection	Aroma	83,59%	63,58%	72,52%	65,82%
	Rata-rata	79,74%	68,30%	74,94%	69,60%

Berdasarkan hasil pengujian menggunakan skenario 2, data yang melalui tahap *preprocessing* menggunakan *stemming* dan melalui tahap *feature selection* mendapatkan nilai performansi lebih baik dibandingkan dengan hasil performansi dari data yang melalui tahap *preprocessing* menggunakan *stemming* dan tanpa melalui tahap *feature selection*. Dapat dilihat perbedaan rata-rata nilai *f1-score* mencapai 1,92%. Hal ini disebabkan nilai recall dan precision memiliki hasil yang lebih baik juga, dimana recall memprediksi data benar positif dengan keseluruhan prediksi positif dan *precision* yang memprediksi data positif yang benar. Pada tahap *feature selection*, peneliti mengambil sebanyak 5000 fitur terbaik (tanpa alasan tertentu) dari perhitungan nilai *chi-square*. Pemilihan fitur ini bertujuan untuk mengurangi fitur yang tidak terlalu penting dan klasifikasi hanya menggunakan fitur yang diperlukan saja. Hal tersebut menyebabkan kerja sistem semakin meningkat dengan hasil rata-rata nilai akurasi 81,60% dan *f1-score* 71,52% lebih tinggi dibandingkan model tanpa melalui tahap *feature selection*.

4.3 Pengujian Skenario 3

Pada pengujian skenario 3 dilakukan pengujian menggunakan dataset terbaik dari hasil skenario sebelumnya. Dataset yang digunakan yaitu data yang melalui tahap *preprocessing* menggunakan *stemming* dan melalui tahap *feature selection*.

Tabel 11. Pengujian Skenario 3 Menggunakan K-fold

Aspek	k-fold	Performansi			
		Accuracy	Precision	Recall	F1-Score
Harga	1	92%	91%	92%	91%
	2	93%	92%	92%	92%
	3	95%	94%	95%	94%
	4	95%	94%	94%	94%
	5	94%	93%	93%	93%
	Rata-rata	93,8%	92,8%	93,2%	92,8%
Kemasan	1	89%	69%	71%	70%
	2	87%	64%	70%	66%
	3	89%	69%	74%	70%
	4	89%	71%	74%	72%
	5	88%	61%	71%	65%
	Rata-rata	88,4%	66,8%	72%	68,6%
	1	59%	56%	66%	56%
	2	60%	57%	66%	57%
	3	61%	59%	66%	58%

Produk	4	59%	56%	65%	56%
	5	60%	56%	65%	57%
	Rata-rata	59,8%	56,8%	65,6%	56,8%
Aroma	1	86%	66%	75%	69%
	2	85%	68%	81%	72%
	3	84%	67%	76%	69%
	4	85%	66%	73%	68%
	5	85%	65%	77%	69%
	Rata-rata	85%	66,4%	76,4%	69,4%
Rata-rata		81,75%	70,7%	76,8%	71,90%

Berdasarkan hasil menggunakan skenario 3, dapat dilihat bahwa sistem dengan menggunakan *k-fold* dengan $k=5$ mendapatkan hasil performansi yang lebih baik dibandingkan dengan pengujian tanpa menggunakan *k-fold*. *K-fold* memiliki cara kerja dengan membagi data sebanyak k -partisi secara acak. Pada penelitian ini penggunaan *k-fold* menghasilkan performansi yang lebih baik. Walaupun, hasil performansi dengan *k-fold* tidak meningkatkan secara signifikan.

Tabel 12. Model Terbaik

Dataset	Accuracy	Precision	Recall	F1-Score
Harga	93,8%	92,80%	93,2%	92,80%
Kemasan	88,4%	66,8%	72%	68,6%
Produk	59,8%	56,8%	65,6%	56,8%
Aroma	85%	66,4%	76,4%	69,4%
Rata-Rata	81,75%	70,7%	76,8%	71,90%

Berdasarkan tabel diatas, hasil model terbaik memiliki rata-rata nilai *accuracy* sebesar 81,75%, nilai *precision* sebesar 70,7%, *recall* sebesar 76,8%, dan *f1-score* sebesar 71,90%. Pada dataset memiliki empat level aspek yaitu harga, kemasan, produk dan aroma. Dapat dilihat nilai *accuracy*, *precision*, *recall* dan *f1-score* pada aspek harga lebih tinggi dibandingkan level aspek yang lainnya. Hal ini disebabkan oleh keseimbangan data dari persebaran kelas-kelas pada aspek harga lebih seimbang dibandingkan persebaran kelas-kelas aspek yang lainnya. Ketidakseimbangan persebaran kelas dapat disebabkan beberapa faktor salah satunya proses pelabelan yang tidak konsisten dan kurang sesuai. Pada aspek harga memiliki persebaran kelas positif, negatif dan netral yang cukup seimbang. Sedangkan aspek produk memiliki persebaran kelas tidak seimbang dengan kelas positif lebih banyak dibandingkan kelas negatif dan netral. Pada aspek produk terdapat fitur yang kurang spesifik sehingga pelabelan tidak konsisten. Selanjutnya untuk persebaran kelas pada aspek kemasan dan aroma memiliki persebaran kelas netral lebih banyak dibandingkan kelas positif dan negatif.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, peneliti telah membangun sistem untuk menganalisa sentimen dari ulasan produk kecantikan pada website female daily menggunakan metode *Random Forest* dan *TF-IDF* sebagai fitur seleksi. Pada tahap pengujian dilakukan tiga skenario pengujian. Pada skenario pertama melihat tahap *stemming* dan tanpa *stemming*. Pada skenario kedua melihat pengaruh tahap *feature selection* dan tanpa *feature*

selection. Selanjutnya pada tahap ketiga melihat pengaruh metode evaluasi yang digunakan.

Peneliti menarik kesimpulan dari ketiga pengujian, pemilihan tahap *preprocessing*, penggunaan *feature selection*, dan penggunaan tahap evaluasi dengan *k-fold* mempengaruhi hasil performansi pada analisis sentimen pada ulasan produk kecantikan. Hasil terbaik yaitu menggunakan dataset melalui tahap *preprocessing* menggunakan *stemming*, melalui tahap *feature selection* dan uji performansi dengan *k-fold* dengan nilai akurasi terbaik sebesar 81,75% dan nilai *f1-score* sebesar 71,95%.

Saran untuk penelitian selanjutnya adalah melakukan labeling yang lebih spesifik dan lebih akurat, melakukan pengecekan terhadap daftar kamus kata pada stopword dan normalization untuk menghindari kata yang mempengaruhi performansi, menggunakan metode fitur ekstraksi dan seleksi yang lain, mencoba memodifikasi metode untuk mendapatkan sistem yang lebih efisien dan performansi yang lebih baik.

Daftar Pustaka

- [1] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.
- [2] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, 2011, doi: 10.1016/j.eswa.2010.12.147.
- [3] H. Ardian and S. Kosasi, "Analisis Sentimen Pada Review Produk Kosmetik Bahasa Indonesia Dengan Metode Naive Bayes," *J. ENTER*, vol. 2, no. 1, pp. 306–320, 2019.
- [4] A. Sharma and S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis," *Int. J. Comput. Appl.*, no. June, pp. 15–20, 2012, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Performance+Investigation+of+Feature+Selection+Methods+and+Sentiment+Lexicons+for+Sentiment+Analysis#0>
- [5] S. Ballı and O. Karasoy, "Development of content-based SMS classification application by using Word2Vec-based feature extraction," *IET Softw.*, vol. 13, no. 4, pp. 295–304, 2019, doi: 10.1049/iet-sen.2018.5046.
- [6] A. Ramadhan, B. Susetyo, and Indahwati, "Penerapan Metode Klasifikasi Random Forest Dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan," *J. Pendidik. dan Kebud.*, vol. 4, no. 2, pp. 169–182, 2019, doi: 10.24832/jpnk.v4i2.1327.
- [7] H. Pouransari, "Deep learning for sentiment analysis of movie reviews," *CS224N Proj.*, pp. 1–8, 2014, [Online]. Available: <http://web.stanford.edu/class/cs224d/reports/PouransariHadi.pdf>
- [8] T. B. Rohman, D. D. Purwanto, and J. Santoso, "Sentiment Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest," *Pros. Semin. Nas. Teknol. Inf. Apl.*, pp. 7–11, 2018.
- [9] Hana Chyntia Moramal¹, Dian Eka Ratnawati², Issa Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *Vol. 6, No. 4, April 2022*, hlm. 1702-1708.
- [10] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [11] Okta Ihza Gifari¹, Muh. Adha², Ivan Rifky Hendrawan³, Fernandito Freddy Setlight Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH*, Vol. 2, No. 1, Maret 2022.
- [12] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.
- [13] X. Rong, "word2vec Parameter Learning Explained," pp. 1–21, 2014, [Online]. Available: <http://arxiv.org/abs/1411.2738>
- [14] T. S. Kasus, P. Pada, and S. Tripadvisor, "Analisis Sentimen Pelanggan Hotel di," vol. 8106, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.
- [15] M. A. A. Jihad, Adiwijaya, and W. Astuti, "Analisis sentimen terhadap ulasan film menggunakan algoritma random forest," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10153–10165, 2021.
- [16] I. Hemalatha and A. Govardhan, "Preprocessing the Informal Text for efficient ALGORITHM FOR," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 1, no. 2, p. 4, 2012.
- [17] Dwi Septiani, Ica Isabela, "ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS", *Vol. 01, No. 2, Maret 2022*, e-ISSN 2807-9108