

PENERAPAN ALGORITMA *DECISION TREE CART* DALAM ANALISIS SENTIMEN PRODUK KECANTIKAN LOKAL STUDI KASUS PADA SOMETHINC DAN WARDAH

Sri Sutjiningtyas¹⁾, Samsul Budiarto²⁾, Iswanto³⁾, Rena Imanina Ilmiah⁴⁾

^{1,2,3,4}Program Studi Teknik Informatika, Universitas Nurtanio Bandung

Corresponding Author: srisutjiningtyas70@gmail.com

Article Info

Article history:

Received: Jan 14, 2026

Revised: Jan 29, 2026

Accepted: Jan 31, 2026

Published: Feb 01, 2026

Keywords:

Produk Kecantikan
Somethinc
Wardah
Analisis Sentimen
Decision Tree CART

ABSTRACT

Produk kecantikan terdiri atas dua kategori yaitu skincare dan make up. Penelitian ini bertujuan menganalisis sentimen dengan membangun pemodelan untuk memahami respons masyarakat terhadap produk kecantikan. Sentimen ulasan konsumen sebelumnya hanya berdasarkan perkiraan tanpa pemodelan ilmiah yang akurat, menyebabkan prediksi perusahaan sering meleset. Analisis sentimen menjadi solusi utama untuk mengkategorikan ulasan sebagai positif atau negatif, serta membantu perusahaan dalam pengambilan keputusan yang lebih baik. Dataset yang digunakan pada produk Somethinc berjumlah 2215 sedangkan pada produk Wardah 2098. Dalam penelitian ini, dilakukan up sampling karena ada ketidakseimbangan data dengan menyeimbangkan jumlah sampel model memiliki kesempatan lebih baik untuk belajar dari masing-masing kelas secara efektif. Penelitian ini menggunakan metode Decision Tree CART. Kedua produk Somethinc dan Wardah menunjukkan lebih banyak sentimen positif dibandingkan sentimen negatif. Ini menunjukkan bahwa kedua brand ini memiliki penerimaan yang baik di kalangan konsumen. Hasil penelitian dari kinerja model Decision Tree CART dalam mengklasifikasikan sentimen ulasan menghasilkan akurasi sebesar 80,13% untuk produk Somethinc sedangkan untuk produk Wardah menghasilkan akurasi sebesar 86,19%, yang berarti tingkat akurasi model yang dibangun sudah baik dalam mengklasifikasikan sentimen negatif dan positif terhadap produk Somethinc dan Wardah. Dengan menggunakan model ini sebagai referensi untuk mengevaluasi ulasan konsumen terhadap produk, perusahaan dapat menggunakannya sebagai masukan dalam pengambilan keputusan berikutnya.



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. PENDAHULUAN

Dari sekian banyak media sosial, Twitter merupakan salah satu media sosial yang paling pesat perkembangannya. Twitter dengan cepat mengambil perhatian masyarakat Indonesia, remaja khususnya. Pada penelitian We Are Social, ada sekitar 27,5 juta pengguna Twitter sehingga jumlah itu menempatkan Indonesia di peringkat keempat global [1]. Kebanyakan isi dalam Twitter adalah hal-hal pribadi dimana seorang berbagai cerita, opini, aktivitasnya, kepada orang-orang terpilih [2]. Sehingga bukan hal yang aneh jika Twitter menjadi platform untuk berbagi pemikiran, dan opini. Pendapat pada platform media sosial Twitter dapat mencakup semua topik, termasuk produk kecantikan.

Produk kecantikan wajah terdiri atas dua kategori yaitu skincare dan make up. Skincare adalah produk perawatan kulit yang digunakan untuk memelihara kesehatan dan penampilan kulit wajah, sedangkan make up adalah produk-produk yang digunakan untuk mengubah penampilan wajah dan menambah keindahan. Menurut dr. Kardiana Dewi, awareness orang untuk perawatan tubuh, termasuk wajah meningkat pesat [3]. Semakin banyak brand lokal kecantikan di pasaran maka ulasan yang muncul terhadap aspek pada produk, berpengaruh pada penilaian konsumen baik itu terhadap kualitas, kemasan, kecocokan dan harga produk. Brand lokal kecantikan menjadi kunci untuk konsumen menilai kualitas produk. Setiap orang memiliki pengalaman yang unik terkait dengan pemakaian produk skincare.

Analisis sentimen adalah proses untuk mengenali dan mengkategorikan opini yang diekspresikan dalam teks, seperti ulasan produk, komentar, atau posting di media sosial. Tujuan utama analisis ini adalah untuk mengklasifikasikan opini menjadi kategori seperti positif atau negatif. Dalam penerapan machine learning, analisis sentimen melibatkan penggunaan model untuk mengenali pola dan hubungan dalam data teks, sehingga model tersebut dapat memprediksi sentimen secara otomatis. Machine learning menggunakan metode seperti Decision Tree CART untuk membuat model berdasarkan data, yang kemudian dapat digunakan untuk mengklasifikasikan data baru.

Hampir 50% dari pengguna internet bergantung pada rekomendasi word-of-mouth (opini) sebelum membeli atau menggunakan suatu produk, termasuk opini-opini yang ada dalam sebuah forum [4]. Dari beberapa asumsi pendapat dari konsumen menunjukkan sentimen terkait harga kecenderungan orang menganggap bahwa jika dibandingkan dengan produk Wardah, produk Somethinc masih lebih mahal [5], tetapi jika dibandingkan secara kualitas Wardah masih ada jauh diatas Somethinc [6]. Sentimen sementara ini menggambarkan sebuah kondisi kecenderungan terhadap produk. Namun, penentuan sentimen ini masih berdasarkan pada perkiraan saja tidak berdasarkan sebuah pemodelan yang ilmiah sehingga prediksi yang dilakukan oleh perusahaan itu terkadang meleset. Hal ini menyebabkan bahwa faktanya di lapangan pihak perusahaan terkadang mengalami kondisi yang berbeda dengan apa yang di perkirakan. Hasil prediksi yang dilakukan oleh perusahaan terkait dari membaca ulasan tidak menunjukkan akurasi yang tepat maka sentimen tidak berjalan dengan baik dan tidak akurat.

Oleh karena itu, analisis sentimen dengan pemodelan untuk menganalisis klasifikasi jenis opini dalam ulasan dan mengkategorikan ulasan sebagai positif dan negatif menjadi solusi utama dalam mengatasi kendala tersebut. Maka dalam penelitian ini akan dibangun sebuah pemodelan sentimen untuk memahami respons masyarakat terhadap produk. Dalam menerapkan pemodelan sentimen dapat memahami respons masyarakat terhadap produk kecantikan yang dapat berguna untuk pengambilan keputusan dalam menilai produk agar menghasilkan prediksi yang lebih sistematis dan akurat. Selain itu, dengan mengetahui sentimen konsumen secara lebih baik, perusahaan juga dapat merespons lebih efektif terhadap umpan balik pasar, meningkatkan kualitas produk, dan memperkuat hubungan dengan pelanggan.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah (1) bagaimana implementasi dalam klasifikasi analisis sentimen

dengan pemodelan berbasis aspek terhadap brand lokal kecantikan pada Twitter menggunakan metode Decision Tree CART; (2) bagaimana hasil akurasi yang dihasilkan dari klasifikasi analisis sentimen berbasis aspek menggunakan metode Decision Tree CART.

Adapun tujuan dari penelitian ini adalah untuk mengimplementasikan klasifikasi analisis sentimen dengan pemodelan berbasis aspek terhadap brand lokal kecantikan pada Twitter menggunakan metode Decision Tree CART. Selain itu penelitian ini bertujuan untuk mengetahui hasil akurasi dari penerapan klasifikasi analisis sentimen berbasis aspek menggunakan metode Decision Tree CART.

2. METODE PENELITIAN

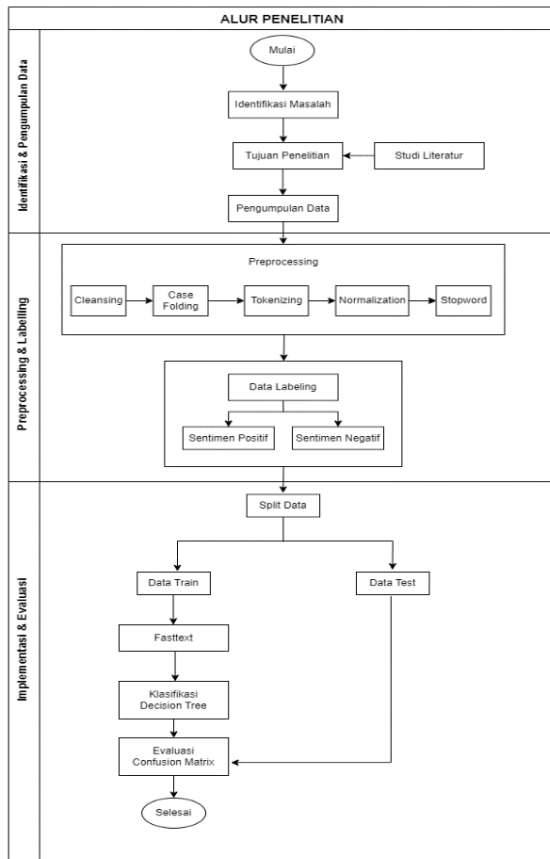
Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis *machine learning* untuk melakukan analisis sentimen berbasis aspek terhadap ulasan produk brand lokal kecantikan pada platform Twitter (X). Metode klasifikasi yang digunakan adalah *Decision Tree CART (Classification and Regression Tree)*.

Objek penelitian difokuskan pada dua brand lokal kecantikan, yaitu Somethinc dan Wardah, dengan aspek analisis meliputi: harga, kualitas, kemasan, dan kecocokan produk.

Dengan menggunakan kombinasi *word embedding Fasttext* dan algoritma *Decision Tree CART*, serta evaluasi menggunakan *confusion matrix*, penelitian ini diharapkan dapat menghasilkan klasifikasi opini yang lebih andal. Hasil dari analisis ini diharapkan membantu perusahaan dalam memahami sentimen konsumen secara lebih baik

2.1. Alur Penelitian

Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data hingga evaluasi model.



Gambar 1. Alur Penelitian

Alur penelitian pada gambar 1 dimulai dengan tahap "Identifikasi & Pengumpulan Data". Proses dimulai dari "Identifikasi Masalah" untuk menentukan isu yang akan diteliti. Setelah itu, tujuan penelitian ditetapkan dan dilengkapi dengan "Studi Literatur" untuk memahami latar belakang dan referensi yang relevan. Langkah berikutnya adalah "Pengumpulan Data" sebagai bahan utama untuk analisis.

Setelah data terkumpul, alur penelitian masuk ke tahap "Preprocessing & Labeling". Pada tahap ini, data terlebih dahulu di bersihkan melalui proses "Cleansing" untuk membersihkan data dari informasi yang tidak diperlukan. Kemudian di lakukan "Case Folding" untuk mengubah huruf besar ke huruf kecil, selanjutnya dengan "Tokenizing" untuk memecah teks menjadi token-token yang lebih kecil. Proses selanjutnya adalah "Normalization" untuk mengubah data ke dalam bentuk yang lebih standar, dan terakhir "Stopword" untuk menghapus kata-kata umum yang tidak memberikan informasi penting. Setelah *preprocessing*, data diberi label dengan kategori "Sentimen Positif" atau "Sentimen Negatif". Data yang telah dilabeli kemudian dipecah menjadi dua set

"Data Train" dan "Data Test" untuk keperluan pelatihan dan pengujian model.

Pada tahap "Implementasi & Evaluasi", "Data Train" digunakan untuk melatih model menggunakan "Fasttext". Hasil dari *Fasttext* kemudian diklasifikasikan menggunakan algoritma "Klasifikasi Decision Tree". Model yang telah dilatih kemudian dievaluasi menggunakan "Confusion Matrix" untuk menilai performa model berdasarkan data uji.

2.2. Pengumpulan Data

Data diperoleh dari media sosial X (Twitter) dengan menggunakan teknik *crawling* berdasarkan kata kunci yang berkaitan dengan kecantikan *make up* dan *skincare* something dan *wardah*. Proses pengumpulan data dilakukan dengan menggunakan *library pandas* untuk *python* dan *install node.js*, serta menggunakan *library tweet harvest* dengan token akses yaitu *auth token* untuk menyediakan otentikasi dan otorisasi untuk mengakses data dari API Twitter.

2.3. Preprocessing Data

Tahap *preprocessing* dilakukan untuk mengelola data mentah yang sudah diambil dengan memperbaiki dan membersihkan data yang belum sesuai. Tahapan *preprocessing* meliputi:

1. *Cleansing*: Menghapus URL, simbol, emotikon, hashtag, mention, dan karakter non-alfabet.
2. *Case folding*: Mengubah seluruh huruf menjadi huruf kecil.
3. *Tokenizing*: Memecah kalimat menjadi token kata.
4. *Normalization*: Mengubah kata tidak baku menjadi bentuk baku.
5. *Stopword*: Menghapus kata-kata umum yang tidak bermakna penting.

2.4. Pelabelan Data

Pelabelan data dilakukan secara manual untuk mengklasifikasikan tweet ke dalam dua kelas sentimen yaitu positif dan negatif.

data_berih	LABEL 1	Label 2	Label 3	Hasil
rahasia make up tahan flawless skin prep setting spray kali kenalin moisturizer	Positif	Positif	Negatif	Positif
ter the best something make up seharian flawless dempul				

Gambar 2. Proses Pelabelan Data

2.5. Pembagian Data

Pada tahap ini data akan dibagi menjadi dua subset yaitu data train dan data set yang digunakan untuk melatih dan mengevaluasi model yang telah dilatih.

2.6. Up Sampling

Data sentimen mengalami ketidakseimbangan kelas sehingga dilakukan teknik *up sampling* untuk menyeimbangkan jumlah data sentimen positif dan negatif. Pada tahap ini data akan melakukan penambahan dan pengurangan data.

2.7. Word embedding Fasttext

Pada tahap ini akan dilakukan *word embedding* menggunakan *library fasttext*, setiap kata akan berubah menjadi vektor.

2.8. Klasifikasi Model Decision Tree

Metode klasifikasi yang digunakan adalah *Decision Tree* CART, yang membangun struktur pohon keputusan berdasarkan nilai Gini Index untuk mendapatkan hasil akurasi *Decision Tree* yang optimal.

2.9. Evaluasi Model Confusion Matrix

Tahap terakhir dilakukan evaluasi setelah mendapatkan hasil dari model *Decision Tree* dengan menggunakan *confusion matrix* yang akan menghasilkan akurasi, presisi, recall dan F1-Score.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap brand lokal kecantikan Somethinc dan Wardah pada *platform* X (Twitter) menggunakan metode *Decision Tree* CART. Data yang digunakan diperoleh melalui proses *crawling* dengan total 4.313 data tweet, terdiri dari 2.215 tweet Somethinc dan 2.098 tweet Wardah.

3.1. Preprocessing Data

1. Cleansing

Tabel 1. Cleansing

Brand	Sebelum Cleansing	Setelah Cleansing
Somethinc	Rahasia make up tahan lama dan flawless adalah skin prep + setting spray Nah kali ini aku mau kenalin ke kalian moisturizer ter the best dari Somethinc yang buat make up aku seharian flawless no dempul https://t.co/B03DisGi2c	Rahasia make up tahan lama dan flawless adalah skin prep setting spray Nah kali ini aku mau kenalin ke kalian moisturizer ter the best dari Somethinc yang buat make up aku seharian flawless no dempul.
Wardah	kalau biasanya Wardah punya pelembab cream ini ada yang bentuknya gel & endul banget di kulit. tipe gel yang langsung lumer begitu nyentuh kulit. ini Hydra Rose Dewy Aqua Gel Day line paling cantik punya Wardah. #racuninskinicare https://t.co/AkqK7oKfON https://t.co/szpM5k2Auf	kalau biasanya Wardah punya pelembab cream ini ada yang bentuknya gel endul banget di kulit tipe gel yang langsung lumer begitu nyentuh kulit ini Hydra Rose Dewy Aqua Gel Day line paling cantik punya Wardah

2. Case Folding

Tabel 2. Case Folding

Brand	Sebelum Case Folding	Setelah Case Folding
Somethinc	Rahasia make up tahan lama dan flawless adalah skin prep setting spray Nah kali ini aku mau kenalin ke kalian moisturizer ter the best dari Somethinc yang buat make up aku seharian flawless no dempul	rahasia make up tahan lama dan flawless adalah skin prep setting spray nah kali ini aku mau kenalin ke kalian moisturizer ter the best dari Somethinc yang buat make up aku seharian flawless no dempul
Wardah	kalau biasanya Wardah punya pelembab cream ini ada yang bentuknya gel endul banget di kulit tipe gel yang langsung lumer begitu nyentuh kulit ini Hydra Rose Dewy Aqua Gel Day line paling cantik punya Wardah	kalau biasanya Wardah punya pelembab cream ini ada yang bentuknya gel endul banget di kulit tipe gel yang langsung lumer begitu nyentuh kulit ini hydra rose dewy aqua gel day line paling cantik punya Wardah

3. Tokenizing

Tabel 3. Tokenizing

Brand	Sebelum Tokenizing	Setelah Tokenizing
Somethinc	rahasia make up tahan lama dan flawless adalah skin prep setting spray nah kali ini aku mau kenalin ke kalian moisturizer ter the best dari Somethinc yang buat make up aku seharian flawless no dempul	['rahasia', 'make', 'up', 'tahan', 'lama', 'dan', 'flawless', 'adalah', 'skin', 'prep', 'setting', 'spray', 'nah', 'kali', 'ini', 'aku', 'mau', 'kenalin', 'ke', 'kalian', 'moisturizer', 'ter', 'the', 'best', 'dari', 'Somethinc', 'yang', 'buat', 'make', 'up', 'aku', 'sehari', 'an', 'flawless', 'no', 'dempul']
Wardah	kalau biasanya Wardah punya pelembab cream ini ada yang bentuknya gel endul banget di kulit tipe gel yang langsung lumer begitu nyentuh kulit ini hydra rose dewy aqua gel day line paling cantik punya Wardah	['kalau', 'biasanya', 'Wardah', 'punya', 'pelembab', 'cream', 'ini', 'ada', 'yang', 'bentuknya', 'gel', 'endul', 'banget', 'di', 'kulit', 'tipe', 'gel', 'yang', 'langsung', 'lumer', 'begitu', 'nyentuh', 'kulit', 'ini', 'hydra', 'rose', 'dewy', 'aqua', 'gel', 'day', 'line', 'paling', 'cantik', 'punya', 'Wardah']

4. Normalization

Tabel 4. Normalization

Brand	Sebelum Normalization	Setelah Normalization
Somethinc	['rahasia', 'make', 'up', 'tahan', 'lama', 'dan', 'flawless', 'adalah', 'skin', 'prep', 'setting', 'spray', 'nah', 'kali', 'ini', 'aku', 'mau', 'kenalin', 'ke', 'kalian', 'moisturizer', 'ter', 'the', 'best', 'dari', 'Somethinc', 'yang', 'buat', 'make', 'up', 'aku', 'seharian', 'flawless', 'no', 'dempul']	['rahasia', 'make', 'up', 'tahan', 'lama', 'dan', 'flawless', 'adalah', 'skin', 'prep', 'setting', 'spray', 'nah', 'kali', 'ini', 'saya', 'mau', 'kenalin', 'ke', 'kalian', 'moisturizer', 'ter', 'the', 'best', 'dari', 'Somethinc', 'yang', 'buat', 'make', 'up', 'saya', 'seharian', 'flawless', 'tidak', 'dempul']
	['kalau', 'biasanya', 'Wardah', 'punya', 'pelembab', 'cream', 'ini', 'ada', 'yang', 'bentuknya', 'gel', 'endul', 'banget', 'di', 'kulit', 'tipe', 'gel', 'yang', 'langsung', 'lumer', 'begitu', 'nyentuh', 'kulit', 'ini', 'hydra', 'rose', 'dewy', 'aqua', 'gel', 'day', 'line', 'paling', 'cantik', 'punya', 'Wardah']	['kalau', 'biasanya', 'Wardah', 'punya', 'pelembab', 'cream', 'ini', 'ada', 'yang', 'bentuknya', 'gel', 'endul', 'banget', 'di', 'kulit', 'tipe', 'gel', 'yang', 'langsung', 'lumer', 'begitu', 'nyentuh', 'kulit', 'ini', 'hydra', 'rose', 'dewy', 'aqua', 'gel', 'day', 'line', 'paling', 'cantik', 'punya', 'Wardah']

5. Stopword

Tabel 5. Stopword

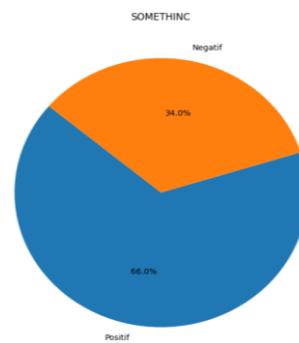
Brand	Sebelum Stopword	Setelah Stopword
Somethinc	['rahasia', 'make', 'up', 'tahan', 'lama', 'dan', 'flawless', 'adalah', 'skin', 'prep', 'setting', 'spray', 'nah', 'kali', 'ini', 'saya', 'mau', 'kenalin', 'ke', 'kalian', 'moisturizer', 'ter', 'the', 'best', 'dari', 'Somethinc', 'yang', 'buat', 'make', 'up', 'saya', 'seharian', 'flawless', 'tidak', 'dempul']	['rahasia', 'make', 'up', 'tahan', 'flawless', 'skin', 'prep', 'setting', 'spray', 'kali', 'kenalin', 'moisturizer', 'ter', 'the', 'best', 'Somethinc', 'make', 'up', 'saya', 'flawless', 'dempul']
	['kalau', 'biasanya', 'Wardah', 'punya', 'pelembab', 'cream', 'ini', 'ada', 'yang', 'bentuknya', 'gel', 'endul', 'banget', 'di', 'kulit', 'tipe', 'gel', 'yang', 'langsung', 'lumer', 'begitu', 'nyentuh', 'kulit', 'ini', 'hydra', 'rose', 'dewy', 'aqua', 'gel', 'day', 'line', 'paling', 'cantik', 'punya', 'Wardah']	['Wardah', 'pelembab', 'cream', 'bentuknya', 'gel', 'endul', 'banget', 'kulit', 'tipe', 'gel', 'langsung', 'lumer', 'nyentuh', 'kulit', 'hydra', 'rose', 'dewy', 'aqua', 'gel', 'day', 'line', 'cantik', 'Wardah']

3.2. Pelabelan Sentimen

Proses pelabelan menghasilkan dua kelas sentimen yaitu sentimen positif dan negatif. Untuk data Somethinc yang berjumlah 2215 data terdapat sentimen positif berjumlah 1461 data dan sentimen negatif berjumlah 754 data. Sedangkan untuk data Wardah yang berjumlah 2098 data terdapat sentimen positif berjumlah 1638 data dan sentimen negatif berjumlah 460 data.

	data_bersih	Hasil	Label
0	rahasia make up tahan flawless skin prep setti...	Positif	1
1	beli bedak hooman ekspektasi datang dapat pouc...	Positif	1
2	somethinc pcs best seller twinpack harga sehar...	Positif	1
3	somethinc pcs niacinamide bisa kulit cerah be...	Positif	1
4	somethinc product nya suka beli	Positif	1
...	...	...	...
2210	pakai somethinc oksidasi bikin keliatan kusam	Negatif	0
2211	mascara somethinc hangover nii kaak gumpal banget	Negatif	0
2212	somethinc harganya menguras kantong banget	Negatif	0
2213	maaf banget somethinc mahal kualitas nya	Negatif	0
2214	halo live Sedang malam skincare makeup underpr...	Positif	1

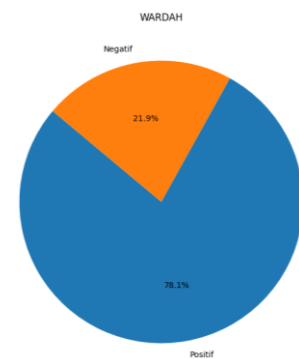
Gambar 3. Hasil Pelabelan Data Somethinc



Gambar 4. Pie Chart Somethinc

	data_bersih	Hasil	Label
0	my favourite toner ever dahla harga murah jima...	Positif	1
1	orang cakap kulit cerah pakai produk wardah aj...	Positif	1
2	haritu pergi watson lepastu nampak pwp wardah ...	Positif	1
3	i m using wardah skincare now range berry biru...	Positif	1
4	all this wardah lightening range is the best p...	Positif	1
...	...	...	...
2093	sungkem sunscreen wardah warna kuning worth it...	Positif	1
2094	cari skincare worth it coba pakek wardah cryst...	Positif	1
2095	racunin best ombre lips dari everyday matte...	Positif	1
2096	kesimpulanku produk wardah varian eyexpert wor...	Positif	1
2097	hii guys suka banget pakai wardah everyday ma...	Positif	1

Gambar 5. Hasil Pelabelan Data Wardah



Gambar 6. Pie Chart Wardah

3.3. Pembagian Data *Training* dan *Testing*

Pada tahap pembagian data menurut prinsip Pareto principle paling umum dikenal sebagai aturan 80:20. Maka untuk pembagian data ini menggunakan 80% sebagai data pelatihan, dan sisanya 20% dari data akan digunakan sebagai data uji. Adapun hasil dari tahap pembagi data latih dan data uji dapat dilihat pada tabel 6 untuk data Somethinc dan Wardah up sampling.

Tabel 6. Pembagian Data

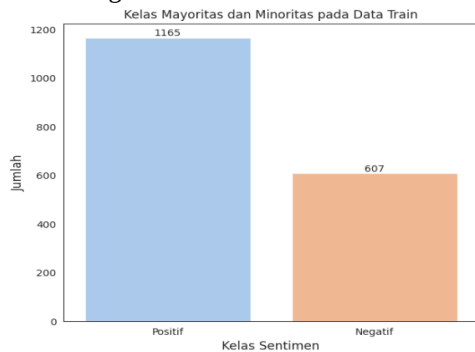
Brand	Jenis Data	Jumlah
Somethinc	Data Train	1772
	Data Test	443
Wardah	Data Train	1678
	Data Test	420

3.4. *Up Sampling* untuk menangani ketidakseimbangan data

Terdapat ketidakseimbangan data antara Somethinc dan Wardah, maka dilakukan *up sampling* pada data *train*. *Up sampling* ini bertujuan untuk menambah data negatif secara acak agar jumlahnya setara dengan data positif, sehingga model dapat dilatih dengan lebih baik.

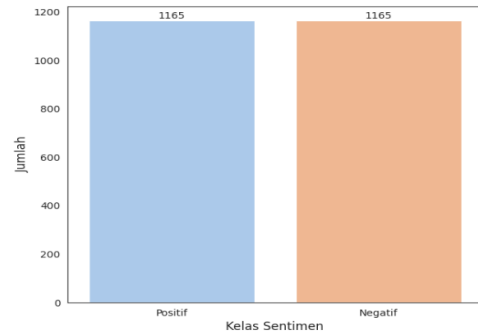
1. Somethinc

Data Somethinc 2215 dilakukan *up sampling* data bertambah menjadi 2773. Sebelum *up sampling* di lakukan pada data train berjumlah 1772 data yang terdapat 1165 data positif dan 607 data negatif.



Gambar 7. Data Somethinc Sebelum Up Sampling

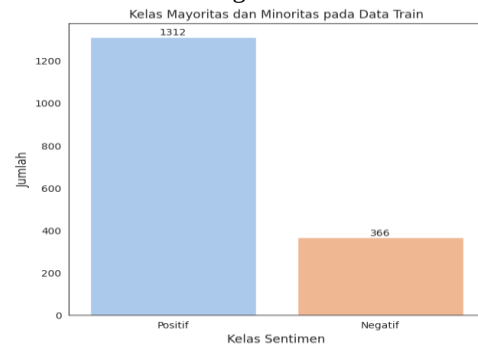
Setelah up sampling pada data train di lakukan, sentimen positif dan negatif menjadi sama rata berjumlah 1165 data.



Gambar 8. Data Somethinc Sesudah Up Sampling

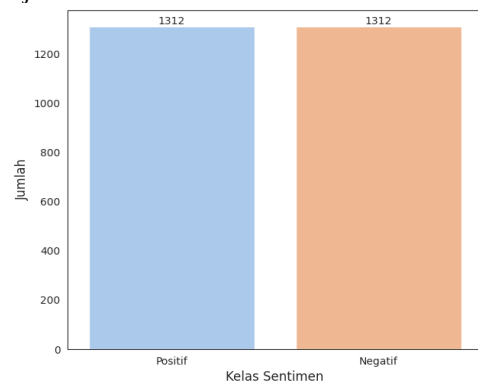
2. Wardah

Data Wardah 2098 dilakukan up sampling data bertambah menjadi 3044. Dapat di lihat pada gambar 4.6 data Wardah sebelum up sampling. Sebelum up sampling di lakukan pada data train berjumlah 1678 data yang terdapat 1312 data positif dan 366 data negatif.



Gambar 9. Data Wardah Sebelum Up Sampling

Setelah up sampling pada data train di lakukan, sentimen positif dan negatif menjadi sama rata berjumlah 1312.



Gambar 10. Data Wardah Sesudah Up Sampling

Maka hasil dari up sampling dari kedua produk Somethinc dan Wardah sebagai berikut :

Tabel 7. Pembagian Data Up Sampling

Brand	Jenis Data	Jml	Sebelum Up Sampling	Sesudah	Hasil
Somethinc	Data Train	1772	1165 (+)	1165	2330
	Data Telst	443	607 (-)	1165	
Wardah	Data Train	1678	1312 (+)	1312	2624
	Data Telst	420	366 (-)	1312	

3.5. Word Embedding Fasttext

	data_bersih	Hasil	Label	fasttext_embedding
479	lipstint warnanya cantik bikin bibir keliatan s...	Positif	1	[0.34819, -0.17964153, 0.1277826, 0.3127356, ...
1984	reponya bling somethinc skak emang somethinc c...	Positif	1	[0.012972163, -0.37701736, -0.22667605, 0.1024...
647	tb bagas tinted es somethinc skinlnt you rev...	Positif	1	[0.38436475, -0.45353624, -0.6143554, 0.37878...
787	lyaa nih alhamdulillah somethinc affordable so...	Positif	1	[0.019550036, -0.14628385, -0.04470942, 0.222...
1596	halley salmine nivea somethinc ortunya piang ...	Positif	1	[0.30088876, -0.3337287, -0.55634075, 0.20283...
...	...	...	...	...
880	somethinc gak nampol banget cushionnya coverag...	Negatif	0	[0.25046575, -0.30659258, -0.25230846, -0.050...
1243	pakai somethinc bikin kusam kulit	Negatif	0	[0.08382947, -0.06101203, 0.031103587, 0.2022...
586	somethinc ceramic skin saviour moisturizer gel...	Negatif	0	[0.2482621, 0.1307463, -0.43141195, 0.09064589...
429	kasih somethinc gak sebagai itu gaes mahal	Negatif	0	[0.12553224, -0.09539877, -0.24002672, -0.2266...
455	nct cerita sender beli somethinc nct dream set...	Negatif	0	[0.015484673, -0.21589941, -0.21299066, 0.502...

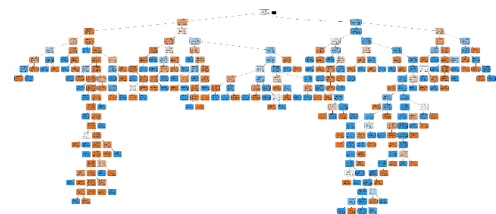
Gambar 11. Fasttext Embedding Somethinc

	data_bersih	Hasil	Label	fasttext_embedding
1591	my sunscreen a thread wardah spt apa mengand...	Positif	1	[0.5087487, -0.4092376, 0.18751384, 0.1304576...
1379	wardah nature daily mineral clarifying clay ma...	Positif	1	[0.45153764, -0.6936586, -0.14915557, 0.52616...
414	wardah murah kualitas lumayan sih kataku ngan...	Positif	1	[0.28500918, -0.2709616, 0.42223594, 0.237654...
704	suka produk muka merek wardah solal produk bad...	Positif	1	[0.34908798, -0.75570893, -0.15306152, 0.2011...
755	pingin nyoba pakai wardah micellar water kemar...	Positif	1	[0.6304204, -0.10197991, -0.33023423, -0.0051...
...	...	...	...	...
1648	wardah sih versi even dipakai makan geser bki...	Negatif	0	[0.9858707, -0.51553625, 0.5117541, 0.2148390...
2083	rek sunscreen garnier worth it make sunscreen ...	Negatif	0	[0.29932955, -0.23135662, -0.4123906, 0.18050...
637	gak recomd banget nih eyeshadow wardah gak b...	Negatif	0	[0.55769265, -0.42323178, -0.23600158, 0.1474...
1633	somethinc harga ketahanan pas omn somethinc wa...	Negatif	0	[0.84446585, -0.34464028, -0.21800473, 0.4068...
487	kualitasnya wardah gak banget	Negatif	0	[0.335171, -0.35303548, 0.13781168, -0.176754...

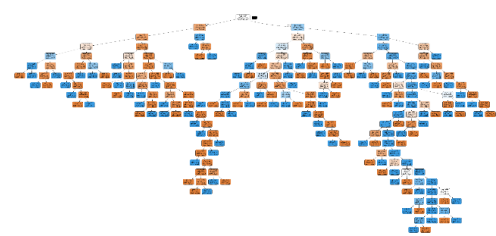
Gambar 12. Fasttext Embedding Wardah

FastText menghasilkan representasi vektor kalimat dengan memecah kata menjadi subkata (n-gram karakter), lalu menjumlahkan vektor tiap subkata untuk membentuk vektor kata. Untuk representasi kalimat, vektor kata dijumlahkan sesuai dimensinya. Misalnya, kalimat "aku suka makeup" direpresentasikan sebagai penjumlahan vektor kata "aku," "suka," dan "makeup," yang masing-masing diperoleh dari vektor n-gram. Hasil vektor ini menangkap makna kalimat secara keseluruhan, dengan nilai positif dan negatif yang membantu model memahami sentimen. Dalam klasifikasi decision tree, fitur vektor FastText digunakan sebagai dasar pengambilan keputusan klasifikasi.

3.6. Klasifikasi Model Decision Tree



Gambar 13. Grafik Decision Tree Somethinc



Gambar 14. Grafik Decision Tree Wardah

Pohon Decision Tree menjelaskan node paling atas merupakan node akar/root yang dimana adalah titik awal dari keputusan yang memiliki nilai gini, jumlah sampel, jumlah sampel per kelas, dan kelas mayoritas.

Node Internal atau cabang dari akar menunjukkan untuk arah cabang kiri dan kanan ditentukan oleh keputusan kurang atau sama dengan maka kondisi dianggap benar (true) cabang akan ada sebelah kiri, jika lebih besar maka kondisi dianggap salah (false) cadang ada di sebelah kanan. Pada bagian warna untuk biru menandakan kelas positif dan warna jingga menandakan kelas negatif, jika warnah lebih cerah menunjukkan node lebih murni yaitu mayoritas besar sampel dalam node tersebut berasal dari satu kelas, namun jika warna lebih pudar menunjukkan node kurang murni yaitu sampel lebih merata di antara dua kelas.

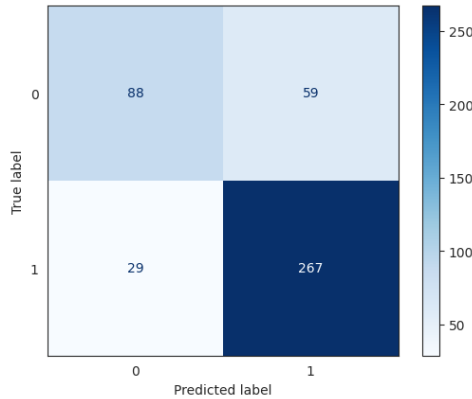
Node daun/leaf adalah titik akhir dari pohon keputusan yang memberikan keputusan akhir tentang kelas yang di prediksi. Sampel dari setiap kelas dan kelas mayoritas, yang merupakan prediksi akhir untuk sampel yang mencapai node tersebut. Yang menjadi node daun pada titik akhir ada beberapa kondisi penghentian seperti jumlah sampel minimum tercapai. Meskipun node ini tidak sepenuhnya murni, aturan penghentian yang berlaku dalam algoritma pembentukan pohon keputusan menyebabkan berhenti sebagai node daun.

3.7. Evaluasi Model menggunakan Confusion Matrix

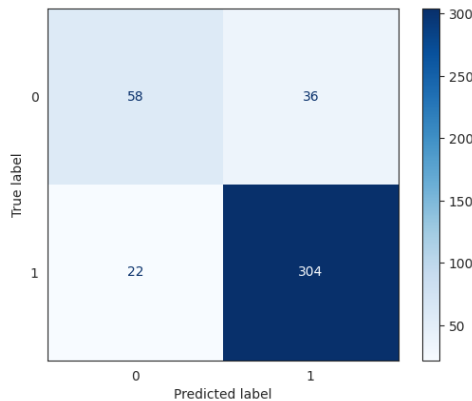
Evaluasi performa model dilakukan menggunakan confusion matrix untuk menghitung nilai akurasi, presisi, recall dan F1-Score. Uji coba data menggunakan up sampling dengan data pembagi 80:20.

Tabel 8. Hasil Evaluasi Model Decision Tree CART

Produk	Akurasi	Presisi	Recall	F1-Score
Somethinc	80,13%	81,90%	90,20%	85,84%
Wardah	86,19%	89,41%	93,25%	91,28%



Gambar 15. Hasil Evaluasi Confusion Matrix Somethinc Up Sampling



Gambar 16. Hasil Evaluasi Confusion Matrix Wardah Up Sampling

Maka hasil dari setiap produk yang telah diuji coba *up sampling* menghasilkan kinerja sebagai berikut :

Tabel 9. Hasil Evaluasi Model

Produk	Persentase
Somethinc	80,13%
Wardah	86,19%

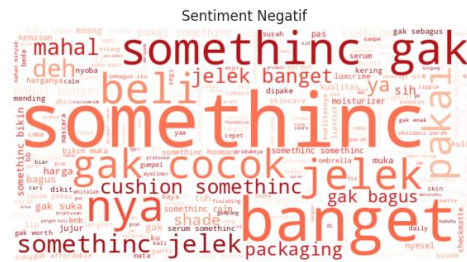
3.8. WordCloud

Wordcloud adalah visualisasi data teks yang menampilkan kata-kata dari kumpulan data teks di suatu dataset. Visualisasi tersebut menjelaskan bahwa semakin besar ukuran suatu kata maka kata-kata tersebut yang frekuensi kemunculannya lebih banyak atau sering muncul dari suatu data teks.

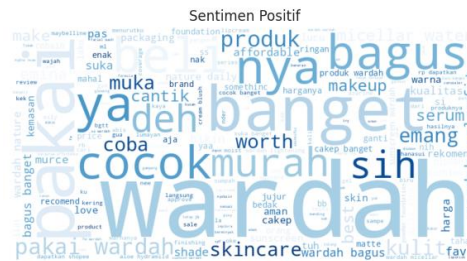
Visualisasi wordcloud kata-kata dalam suatu teks yang sering muncul pada sentimen positif dan sentimen negatif pada produk somethinc dan wardah dapat dilihat pada gambar 17 s.d gambar 20.



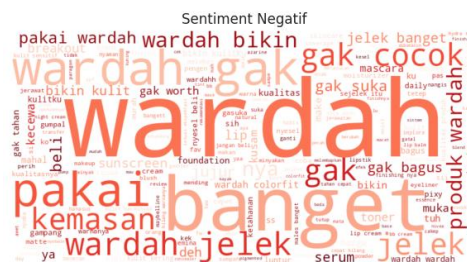
Gambar 17. Wordcloud Sentimen Positif Somethinc



Gambar 18. Wordcloud Sentimen Negatif Somethinc



Gambar 19. Wordcloud Sentimen Positif Wardah



Gambar 20. Wordcloud Sentimen Negatif Wardah

3. KESIMPULAN

Berdasarkan hasil keseluruhan maka dapat diambil kesimpulan sebagai berikut :

1. Dari kedua produk Somethinc dan Wardah menunjukkan lebih banyak sentimen positif dibandingkan sentimen negatif. Untuk produk Somethinc dari total data sebanyak 2215 data dengan sentimen positif berjumlah 1461 dan sentimen negatif berjumlah 754, sedangkan untuk produk Wardah sebanyak 2098 data terdapat sentimen positif berjumlah 1638 data dan sentimen negatif berjumlah 460 data.

Sehingga menunjukkan bahwa kedua brand ini memiliki penerimaan yang baik di kalangan konsumen.

2. Untuk mengatasi ketidakseimbangan kelas sentimen, upsampling dilakukan pada data latih, memungkinkan model belajar lebih efektif. Metode Decision Tree dengan upsampling menghasilkan akurasi 80,13%, presisi 81,90%, recall 90,20%, dan F1-Score 85,84% untuk produk Somethinc, sedangkan untuk produk Wardah menghasilkan akurasi 86,19%, presisi 89,41%, recall 93,25%, dan F1-Score 91,28%, hasil tersebut menunjukkan kinerja model yang di bangun sudah baik dalam mengklasifikasikan sentimen pada kedua produk.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan penghargaan setinggi-tingginya kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LP2M) Universitas Nurtanio Bandung atas dukungan fasilitasi serta pendanaan yang diberikan, sehingga kegiatan penelitian ini dapat terlaksana secara optimal. Penelitian ini merupakan bagian dari skema hibah internal LP2M Universitas Nurtanio Bandung yang ditujukan untuk mendorong peningkatan mutu serta produktivitas penelitian dosen. Selain itu, penulis juga mengucapkan terima kasih kepada seluruh pihak yang telah memberikan kontribusi, baik secara langsung maupun tidak langsung, selama proses pelaksanaan hingga penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] Annur, C. M. (2023, November 28). Ada 27 juta pengguna Twitter di Indonesia, terbanyak ke-4 global. Databoks. <https://databoks.katadata.co.id/datapublish/2023/11/28/ada-27-juta-pengguna-twitter-di-indonesia-terbanyak-ke-4-global>
- [2] Adriyanti, A., & Abubakar, A. H. (2023). Pengaruh perilaku konsumen terhadap keputusan pembelian produk kecantikan melalui e-commerce Shopee. *Jurnal Manajemen*, 17, 1–10.
- [3] Fanani, F., & Bustoni, I. A. (2017). Klasifikasi review software pada Google Play menggunakan pendekatan analisis sentimen (Skripsi). Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Gadjah Mada, Yogyakarta.
- [4] Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM model menggunakan metode decision tree dengan algoritma CART untuk prediksi curah hujan berpotensi banjir. *Journal of Applied Informatics and Computing* (JAIC). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] Harriz, M. A., & Setiyowati, H. (2023). Komparasi algoritma decision tree dan KNN dalam mengklasifikasi daerah berdasarkan produksi listrik. *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 167–175.
- [6] Ilmiyah, K., & Krishernawan, I. (2020). Pengaruh ulasan produk, kemudahan, kepercayaan, dan harga terhadap keputusan pembelian pada marketplace Shopee di Mojokerto. *Maker: Jurnal Manajemen*, 6(1), 31–42.
- [7] Maharani, S., & Nasution, M. A. (2023). Pengaruh harga, selebgram, dan kualitas produk terhadap keputusan pembelian skincare Somethinc pada mahasiswa FEB UMSU. *Jurnal Ekonomi dan Bisnis*, 12(2), 45–56.
- [8] Monarizqa, N., Lestari, N. E., & Hadi, B. S. (2014). Penerapan analisis sentimen pada Twitter berbahasa Indonesia sebagai pemberi rating. Universitas Gadjah Mada. <https://etd.repository.ugm.ac.id/penelitian/detail/71215>
- [9] Nurhadi, Z. F. (2017). Model komunikasi sosial remaja melalui media Twitter. *Jurnal ASPIKOM*, 3(3), 539–550.
- [10] Pamela, D. A. (2022). Produk skincare lokal mendominasi, tren kecantikan 2023 makin glowing tapi simpel. *Liputan6*. <https://www.liputan6.com/lifestyle/read/5142170/produk-skincare-lokal-mendominasi-tren-kecantikan-2023-makin-glowing-tapi-simpel>
- [11] Pratiwi, R. W., Fitriani, H. S., Dairoh, D., Af'idah, D. I., Qomariah, A., & Gunawan, F. A. (2021). Analisis sentimen pada review skincare Female Daily menggunakan metode support vector machine (SVM). *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 4(1), 40–46.
- [12] belbeautydiary. (2022). Somethinc vs Wardah. Lemon8. <https://www.lemon8-app.com/belbeautydiary/7124272505463898626?region=id>
- [13] Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. *TeknoIndonesia*. <https://www.teknoindonesia.com>
- [14] Isnaini, A. N. (2023). Pengertian data, fungsi, jenis-jenis, manfaat, dan contohnya. *Telkom University*. <https://telkomuniversity.ac.id/pengertian-data-fungsi-jenis-jenis-manfaat-dan-contohnya/>
- [15] Putri, S. K., Seltiawan, R., & Pambudi, A. (2023). Analisis sentimen terhadap brand skincare lokal menggunakan Naïve Bayes classifier
- [16] Nuryawan, O., Hasbullah, M., Rizal, M., Rajab, M. F., & Agustina, N. (2023). Algoritma decision tree untuk analisis sentimen publik terhadap marketplace di Indonesia. *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 5(1), 18–25.