

SENTIMENT ANALYSIS USING K-NEAREST NEIGHBOR BASED ON PARTICLE SWARM OPTIMIZATION ACCORDING TO SUNSCREEN'S REVIEWS

Anita Nur Syifa Rahayu¹, Teguh Iman Hermanto², Imam Ma'ruf Nugroho³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Sekolah Tinggi Teknologi Wastukencana, Indonesia
Email: ¹anitanur22@wastukencana.ac.id, ²teguhiman@wastukencana.ac.id, ³imam.ma@wastukencana.ac.id

(Naskah masuk: 28 Juni 2022, Revisi: 20 Juli 2022, diterbitkan: 26 Desember 2022)

Abstract

High UV exposure and tropical climate are affected by the equator that passing through Indonesia. In this case, early aging will haunted not only by skincare lovers but also the rest of Indonesians including male and female. By so, we need to protect our skin using sun protector like sunscreen. Sunscreen application awareness through reviews by many different user probably are the most effective way to get to know the suitable sunscreen. By scrolling through the reviews surely will be time wasting. As of, sentiment analysis is the solution to classifying between negative and positive sentiments from the reviews. This research uses K-Nearest Neighbor (K-NN) algorithm as classification method because this method way more easy and efficient to use by its self learning, thus K-NN can learned its own data through its neighbor. Particle Swarm Optimization is used to increasing the accuration. Evaluation method using Confusion Matrix and the results are accuracy, precision and recall. Classification result using only k-NN and optimized with PSO are getting increase. Brand 'Azarine' rank the first accuration value with 91,89% using k-NN and 92,80% using k-NN PSO.

Keywords: *k nearest neighbor, particle swarm optimization, review, sentiment analysis, sunscreen*

ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION PADA REVIEW SUNSCREEN

Abstrak

Paparan sinar UV yang tinggi karena Indonesia terletak di garis khatulistiwa dan beriklim tropis. Hal ini pun dapat menyebabkan penuaan dini pada kulit karena radiasi sinar ultraviolet (UV). Karenanya, kulit perlu perlindungan dengan memakai produk sun protector pada kulit seperti halnya penggunaan sunscreen. Sehingga meningkat pula pengetahuan yang dibutuhkan akan sunscreen melalui review dari berbagai orang agar mendapatkan sunscreen yang cocok. Dengan scrolling keseluruhan komentar maupun review yang ada tentu akan sangat memakan waktu. Dengan demikian, analisis sentimen dijadikan pemecahan atas masalah tersebut guna mengklasifikasi review menjadi positif dan negatif sehingga dapat memberikan informasi apakah sunscreen tersebut cocok atau tidak. Penelitian ini menggunakan metode klasifikasi K-Nearest Neighbor (K-NN) karena algoritma ini tergolong mudah dan efisien untuk pengklasifikasian teks karena bersifat self learning, sehingga dapat mempelajari struktur data yang ada dan mengkategorikan data nya sendiri melalui tetangga terdekatnya. Penggunaan algoritma optimasi Particle Swarm Optimization (PSO) bertujuan untuk meningkatkan akurasi hasil klasifikasi. Metode pengujian hasil klasifikasi menggunakan metode Confusion Matrix yang akan menghasilkan 3 nilai yaitu Accuracy, Precision dan Recall. Hasil klasifikasi menggunakan k-NN saja dan dioptimasi oleh PSO secara keseluruhan hasilnya meningkat. Brand 'Azarine' menjadi pemegang nilai akurasi terbesar sebesar 91,89% menggunakan K-NN dan 92,80% menggunakan K-NN & PSO.

Kata kunci: *analisis sentimen, k nearest neighbor, particle swarm optimization, review, sunscreen*

1. PENDAHULUAN

Paparan sinar UV yang tinggi karena Indonesia terletak di garis khatulistiwa dan beriklim tropis. Hal ini pun dapat menyebabkan penuaan dini pada kulit karena radiasi sinar ultraviolet (UV) [1]. Sinar UV yang terbagi kedalam tiga jenis yaitu: UV A, UV B, UV C. Sinar UV C ini mampu diserap seluruhnya

oleh lapisan stratosfer pada atmosfer bumi. Sedangkan sinar UV B hanya diserap sebagian. Lain halnya dengan UV A yang tidak dapat diserap sama sekali[2]. Selain itu, berbagai faktor yang dapat merusak lapisan ozon seperti meningkatnya penggunaan Bahan Perusak Ozon atau BPO yang sedianya bersifat melindungi radiasi sinar UV sehingga meningkatkan resiko paparan sinar UV.

Karenanya, kulit perlu perlindungan. Bahaya sinar UV dapat dilindungi dengan 2 macam cara, yaitu perlindungan fisik dan perlindungan kimiawi. Perlindungan fisik dapat dengan memakai sarung tangan, kacamata *panoramic*, patron, pakaian panjang dan lain-lain. Sedangkan perlindungan kimiawi dengan memakai produk sun protector pada kulit seperti halnya penggunaan *sunscreen*. [1]. Hal ini pun meningkatkan kesadaran pentingnya penggunaan *sunscreen* baik diluar maupun di dalam ruangan. Berbagai *brand* kosmetika baik dari luar negeri maupun dalam negeri terus berinovasi dalam membuat formula *sunscreen* yang ringan tetapi ampuh memberikan perlindungan kepada kulit agar terhindar dari efek negatif sinar UV. Sepanjang tahun 2018, produk perawatan kulit (*skin care*) termasuk *sunscreen* di Indonesia memberikan kontribusi dan pasar terbesar. [3]

Hal ini berdampak pada kebutuhan *sunscreen*, mulai dari kecocokan dengan jenis kulit, tekstur dan formula ringan maupun berat, hasil dari penggunaannya apakah lembab ataupun kering, efek negatif dari penggunaan *sunscreen* tersebut, kecocokan dalam penggunaan aktifitas sehari-hari yang berbeda, hingga kisaran harga yang sesuai. Sehingga meningkat pula pengetahuan yang dibutuhkan akan *sunscreen* melalui *review* dari berbagai orang agar mendapatkan *sunscreen* yang cocok. Akan tetapi, tidak semua produk *sunscreen* mempunyai kualitas yang diharapkan konsumen, sehingga pemilihan *sunscreen* yang tepat sangat krusial [4]. Saat ini ada berbagai forum khusus untuk berbagi *review* penggunaan kosmetik khususnya *sunscreen*, salah satunya ialah *website Female Daily* [5]. Dalam forum ini kita dapat melihat *review* dari berbagai macam orang dengan perbedaan jenis kulit, warna kulit, dengan berbagai aktifitas sehari-hari berdasarkan *filter* menu yang dipilih seperti "*Highest Rating*", "*Lower Rating*" sehingga dapat memberikan gambaran terhadap produk yang dicari. Dengan scrolling keseluruhan komentar maupun *review* yang ada tentu akan sangat memakan waktu. Sehingga, analisis sentimen dijadikan pemecahan atas masalah tersebut guna mengklasifikasi *review* menjadi positif dan negatif sehingga dapat memberikan informasi apakah *sunscreen* tersebut cocok atau tidak [6]. Sebagai tambahan, dari banyaknya *review* ada juga yang tidak dapat diklasifikasikan dengan jelas kelas sentimennya [7]. Analisis sentimen merupakan metode yang dilakukan untuk mengidentifikasi emosi serta sikap yang diketahui melalui teks yang diklasifikasikan sebagai sentimen negatif dan sentimen positif [8]. Analisis sentimen yang termasuk ke dalam salah satu metode text mining, ditujukan untuk menganalisis hasil perasaan emosional guna mengklasifikasikannya ke dalam kelas negatif dan positif [9]. Sentimen dapat diklasifikasikan menjadi sentimen negatif dan positif [10]. Dalam penelitian ini, analisis sentimen dilakukan untuk menentukan

respon pengguna *sunscreen* terhadap produk melalui opini untuk mengetahui tingkat kepuasan dari kualitas produk tersebut agar pengguna *sunscreen* dapat memilih produk yang sesuai [4]. Selain itu, analisis sentimen juga digunakan dalam dunia bisnis untuk menganalisis tren [11].

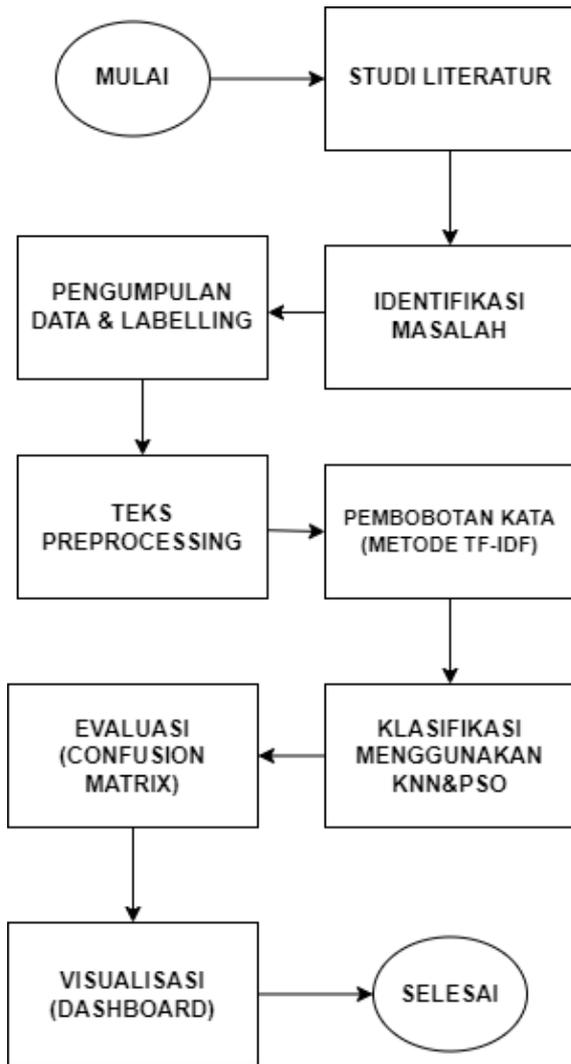
Pada analisis sentimen digunakan metode klasifikasi salah satu metode klasifikasi adalah *K-Nearest Neighbor* (K-NN) [12]. Penelitian ini menggunakan metode klasifikasi K-NN karena algoritma ini tergolong mudah dan efisien untuk pengklasifikasian text. Metode ini dianggap mudah karena bersifat *self learning*, sehingga dapat mempelajari struktur data yang ada dan mengkategorikan data nya sendiri melalui tetangga terdekatnya. Secara singkat, cara kerjanya ialah dengan cara menemukan jarak terdekat dari data *testing* dan data *training*, lalu dilihat data yang mana yang paling dominan dari antar data tersebut. [13]. Pada penelitian ini klasifikasi menggunakan rumus Cos Similarity karena dianggap mampu lebih optimal pada penelitian sebelumnya. [14]. Penggunaan algoritma optimasi *Particle Swarm Optimization* (PSO) bertujuan untuk meningkatkan akurasi hasil klasifikasi. PSO biasa digunakan untuk menangani masalah optimasi karena prosesnya yang tergolong mudah dan cepat. [15].

Untuk mengetahui adanya peningkatan akurasi atau tidak, dilakukan uji evaluasi [16]. Hasil klasifikasi akan di uji menggunakan salah metode evaluasi yaitu *confusion matrix*. Hasil nya dapat ditampilkan melalui nilai *Accuracy*, *Precision* dan *Recall*. *Accuracy* merupakan nilai benar (akurat) antara prediksi dan hasil. Sedangkan *Precision* merupakan nilai benar (akurat) antara prediksi dan data diinginkan. Lain halnya dengan *Recall* yang merupakan akurasi dari keberhasilan model klasifikasi mendapatkan kembali sebuah informasi. [16].

2. METODE PENELITIAN

Sebuah penelitian diharuskan mempunyai metodologi penelitian yang divisualisasikan kedalam kerangka penelitian. Hal ini memudahkan untuk melihat langkah-langkah sistematis guna memperoleh tujuan dengan permasalahan spesifik.

Pada penelitian ini terbagi menjadi 8 tahapan penelitian untuk memudahkan pengerjaan dan pembacaan laporan hasil penelitian. Kerangka penelitian ini runut berdasarkan sistematika penelitian, mulai dari tahapan studi literatur, identifikasi masalah, pengumpulan data, text preprocessing, pembobotan kata, klasifikasi, evaluasi, hingga visualisasi dengan pembuatan dashboard pada Microsoft Power BI. Sehingga disajikanlah kerangka penelitian yang dapat dilihat pada Gambar 1. :



Gambar 1. Kerangka Penelitian.

Dari Gambar 1. dapat diuraikan penelitian ini dilakukan dengan melakukan beberapa tahapan, terbagi menjadi 8 tahapan yaitu meliputi:

1. Studi Literatur, merupakan langkah awal dalam mencari dan membaca berbagai referensi berupa jurnal dan laporan penelitian yang telah ada sebelumnya dari situs Internet yang memiliki keterkaitan dengan topik mengenai analisis sentimen dan penelitian yang menggunakan algoritma K-NN serta penelitian analisis sentimen yang dilakukan berdasarkan *website Female Daily* maupun *website reviewer* lainnya.
2. Identifikasi Masalah, setelah mempelajari referensi, dilakukan identifikasi masalah terhadap data yang akan diteliti hingga pemilihan algoritma klasifikasi yang akan digunakan serta cara meningkatkan akurasi dari hasil klasifikasi menggunakan algoritma klasifikasi. Data yang berasal dari *review* sebuah produk cocok untuk dijadikan penelitian analisis sentimen untuk mencari hasil sentimen negatif dan positifnya. Pada tahap ini pula, ditentukan algoritma klasifikasi nya yaitu *K-Nearest*

Neighbor beserta algoritma optimasinya yaitu *Particle Swarm Optimization*.

3. Pengumpulan Data dan Pelabelan, selanjutnya merencanakan pengambilan data melalui metode data *scrapping* menggunakan *Google Extension* yaitu *web scrapper*. Hal ini menghasilkan data sebanyak 357 data dengan rincian 194 data kelas positif dan 163 kelas negatif. Data yang diambil berasal dari *review* produk *sunscreen brand* “Wardah”, “Azarine” dan “Skin Aqua”. Data *scrapping* dibantu oleh filter “*Highest Rating*” dan “*Lowest Rating*” dalam *website Female daily*. *Labelling* merupakan tahapan pelabelan sentimen terhadap data. Data akan diklasifikasikan ke dalam kelas negatif dan positif. Pelabelan dilakukan dengan bantuan aplikasi *Microsoft Excel*. Tabel 1 merupakan *data preview* yang digunakan pada penelitian ini:

Tabel 1. Data preview.

Komentar	Sentimen
Jujur aku suka banget sama sunsreen ini, gabikin muka aku kusam, ga lengket muka aku cerah gitu ga nyesel deh	Positif

4. *Text Preprocessing* merupakan tahapan yang paling krusial dalam *text mining*, karena dalam tahap inilah akurasi dalam analisis sentimen optimal atau tidak. Dalam penelitian ini dilakukan 3 tahapan dalam *text preprocessing* yaitu, (1) *tokenization*, (2) *case folding* dan (3) *stopword removal*. *Tokenization* ialah tahapan yang dilakukan guna memecah atau memenggal kalimat menjadi kata per kata. Tiap kata sudah dikembalikan kedalam bentuk dasarnya. Tahapan *case folding* akan mengubah semua huruf kapital menjadi *lowercase*, hal ini dilakukan untuk meminimalisir ambiguitas maupun *case sensitive* dalam tahapan selanjutnya. Langkah terakhir dalam *text preprocessing* yaitu *stopword removal*, dalam tahapan ini kata yang telah dipenggal di tahap *tokenization* selanjutnya akan dipilah kembali dan menghapuskan kata yang tidak dibutuhkan.

Tabel 2. Tahapan Tokenization

Sebelum	Sesudah
Jujur aku suka banget sama sunsreen ini, gabikin muka aku kusam, ga lengket muka aku cerah gitu ga nyesel deh belinya.	[’Jujur’,’aku’,’suka’,’banget’,’sama’,’sunsreen’,’ini’,’ga’,’bikin’,’muka’,’aku’,’cerah’,’gitu’,’ga’,’nyesel’,’deh’,’beli’]

Dari tabel 2, dapat dilihat bahwa tahapan *tokenization* ialah menghilangkan semua karakter noise baik emoticon hingga tanda baca. Pada tahap

ini pula kalimat dipecah (dipenggal) menjadi kata per kata.

Tabel 3. Tahapan Case Folding.

Sebelum	Sesudah
['Jujur','aku','suka', 'banget','sama', 'sunscreen','ini','ga', 'bikin','muka','aku', 'cerah','gitu','ga', 'nyesel','deh','beli']	['jujur','aku','suka', 'banget','sama', 'sunscreen','ini','ga', 'bikin','muka','aku', 'cerah','gitu','ga', 'nyesel','deh','beli']

Pada Tabel 3, dapat dikemukakan bahwa tahapan case folding merupakan tahapan untuk mengkonversi semua kata menjadi huruf kecil, sehingga hasil dari tahapan ini data tidak memiliki huruf kapital.

Tabel 4. Tahapan Stopword Removal.

Sebelum	Sesudah
['jujur','aku','suka', 'banget','sama', 'sunscreen','ini','ga', 'bikin','muka','aku', 'cerah','gitu','ga', 'nyesel','deh','beli']	['suka','ga','bikin', 'muka','kusam','ga', 'lengket','muka','cerah', 'ga','nyesel','beli']

Pada Tabel 4, data melalui tahapan stopwords removal setelah sebelumnya melalui 2 tahapan text preprocessing lainnya yaitu case folding dan tokenization. Pada tahap ini data akan dihilangkan kata yang tidak mengandung sentimen. Sehingga, pada tahap ini data hanya akan memiliki kata yang mengandung sentimen negatif dan sentimen positifnya.

5. Pembobotan kata menggunakan metode TF IDF, setelah melalui *text preprocessing* data yang dihasilkan pun sudah bersih dan siap untuk dilakukan pembobotan tiap kata. Hal ini bertujuan untuk mengurutkan peringkat data agar diketahui relevansi kata dalam dokumen. Langkah pertama, dilakukan penghitungan setiap kata pada dokumen hal ini merupakan proses *Term Frequency* (TF). Sedangkan perhitungan IDF dilakukan dengan persamaan (1):

$$\log\left(\frac{n}{df}\right) \tag{1}$$

Setelah itu dihitung bobot nya TF dikalikan hasil IDF.

Keterangan:

n = banyaknya dokumen pada dataset

df = data frequency

6. *Modelling* klasifikasi menggunakan Algoritma K-NN dan PSO, setelah

melewati tahapan pembobotan kata dilakukan proses *modelling* menggunakan algoritma K-NN dan PSO dengan cara dihitung menggunakan persamaan *Cos Similarity*. Dalam realisasinya tahap ini dilakukan menggunakan aplikasi *Rapid Miner*. Adapun persamaan *Cos Similarity* nya dilakukan menggunakan persamaan (2) adalah sebagai berikut:

$$\text{Cos Similarity} = \frac{\sum_k(Wte \times Wtr)}{\sqrt{\sum_k Wte} \sqrt{\sum_k Wtr}} \tag{2}$$

Keterangan:

Wte = Bobot data testing

Wtr = Bobot data training

Hal ini bertujuan untuk mengukur tingkat kesamaan antar data sesuai dengan hakikatnya algoritma ini ialah menghitung kesamaan dengan tetangga terdekat. Setelah mendapatkan hasil dari *Cos Similarity* maka akan dilakukan pemeringkatan kesamaan antara data *training* dan data *testing*. Peringkat tertinggi diasumsikan sebagai perkiraan kelas dari data *testing*. Pada penelitian ini digunakan $k=1$ dan $k=3$.

7. Evaluasi Model pada penelitian ini menggunakan metode *confusion matrix*. Hal ini bertujuan untuk membandingkan nilai yang sebenarnya dengan nilai hasil prediksi. Evaluasi menggunakan metode ini dapat menghasilkan *accuracy*, *precision* dan *recall*. *True Positive* merupakan data yang diprediksi positif dan benar positif. Sama halnya dengan nilai *True Negative* merupakan data yang diprediksi sebagai negatif dan benar negatif. *False Positive* merupakan data yang bernilai negatif tetapi diprediksi positif. Begitu juga dengan *False Negative* merupakan data yang bernilai positif dan diprediksi sebagai negatif. Adapun perhitungan akurasi menggunakan persamaan (3):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Dibawah ini merupakan perhitungan *Precision* dan *Recall* menggunakan persamaan (4) dan (5):

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

Keterangan:

TP = True Positive.

TN = True Negative.

FN = False Negative.

FP = False Positive.

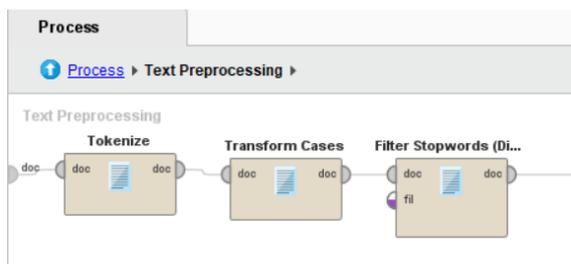
- dan Visualisasi, hasil penelitian ini divisualisasikan dengan menggunakan *dashboard*. Aplikasi yang digunakan untuk membuat *dashboard* ialah *Microsoft Power BI*.

3. HASIL DAN PEMBAHASAN

Data didapatkan dengan cara metode scrapping dengan menggunakan Google Extension Bernama Web Scrapper. Data diambil dari reviews 3 brand produk sunscreen yaitu ‘Azarine’, ‘Wardah’, dan ‘Skin Aqua’ pada website Female Daily sebagai website honest review untuk berbagai produk kecantikan mulai dari hair care, body care, skin care hingga make up. Setelah didapatkan ketiga brand tersebut akan dilakukan proses labelling di Microsoft Excel seperti pada Gambar 2.

text	sentiment
Sorryyyyyyy..... Tapi aku gak rekomen produk ini.... Why karena di aku gak cocok, Bo. negatif	negatif
Super suka sama produk ini. Dia itu selain melindungi juga melembabkan karena memiliki kac positif	positif
Awalnya cari-cari sunscreen yang cocok buat kulit kering kaya aku sih, terus nemu skin aqua positif	positif
simple, pemakaian botol bulan, nuka jadi kerasa lembab juga sih.. meresap ke kulit juga cepi positif	positif
Favorit sunscreenu Udah gakeitung berapa kali repurchase produk ini saking sukanya. Setiap positif	positif
teksturnya cair, jadi mudah banget diaply ke wajah, gak ada white cast, malah kasih efek glr positif	positif
Pernah dulu pake ini memang teksturnya enak buat kulit yg berminyak cepet meresapnya gal negatif	negatif
Sunscreen favorite akuuu, teksturnya ringan jadi pas dipake tu ga terasa berat dan ada efek i positif	positif
Suka bgt sama sunscreen ini. udah habis 3 botol. Diaku sih efek nya pas baru dipake bikin wa positif	positif
sunscreen favorit awalnya bingung mau pake ini atau produk lain.. tp pas d drugstore, nemu positif	positif
ss kedua setelah biore yg ini free alcohol harganya juga worth it dan alhamdulillah ga bikin negatif	negatif
Skin Aqua UV Moisture Milk ini formulanya light banget, ga bikin white cast kalo dipake, positif	positif
duuh gmn yaa, sebenarnya suka sama teksturnya doi karna encer gtt ga yg thick, cuman di w negatif	negatif
Awal-awal pake ini sih fine aja, kulit jadi lebih lembab juga ga kering, tapi pas lagi jerawatan (negatif	negatif
Udah pake ini kayanya lebih dari tahun. Teksturnya ringan bgt tapi sangat melembabkan. Fin positif	positif
Gaktau kenapa produknya greasy bgt di aku. ganyaman makenya dari awal dan hasilnya juga negatif	negatif
nggak ngerti lagi ini baguuss banget meskipun teksturnya agak lengket tapi cepat meresap d positif	positif
Ini sunscreen pertama yg bikin aku rajin pake sunscreen, gak bikin kusam juga gak bereaksi me positif	positif
Super luv banget sama dia teksturnya cair banget tapi moisturizer banget. Kebetulan kulit mu positif	positif
Good product, harganya ekonomis walau ternyata kecil ya, hanya se-telapak tangan dan ga i positif	positif
SUPEERRR CINTAAAKKK sama sunscreen ini, baru setahunan make kayaknya udh ngabisin s positif	positif
Tertarik beli karena sunscreen ini termasuk aman buat kulit yang prone to fungal acne. But t negatif	negatif
kulitku sensitif banget terutama kalo skincarenya oil based. Mana sunscreen kebanyakan pai negatif	negatif

Gambar 2. Hasil data scrapping dan labelling.



Gambar 3. Text Preprocessing pada Rapid Miner.

Tahapan *preprocessing* yang disajikan dalam Gambar 3, dan pembobotan kata menggunakan metode TF-IDF hingga modelling klasifikasi pada penelitian ini dibuat menggunakan aplikasi *Rapid Miner*. Operator *tokenize* akan mengeksekusi tahapan *tokenization*. Sedangkan operator *transform cases* akan mengerjakan tahapan *case folding*, untuk proses *text preprocessing* yang terakhir menggunakan operator *filter stopwords* untuk melaksanakan perintah *stopword removal*.

Data yang berformat *.xlsx*. diimport menjadi sebuah input. Tiap data (*brand*) akan dibuat 2 model klasifikasi, menggunakan algoritma K-NN dan PSO. Data akan terbagi menjadi 2 yaitu *data testing* dan *data training*. Operator *cross validation* yang akan membagi data tersebut. Adapun *number of folds* pada operator *cross validation* yang digunakan dalam penelitian ini sejumlah k = 1 dan k = 3.

Rumus *Cos similarity* digunakan sebagai perhitungan klasifikasi menggunakan K-NN. Nilai dari hasil pembobotan menggunakan metode TF-IDF akan diklasifikasikan dengan rumus *Cos Similarity*. Sehingga rumus ini akan mengukur jarak kedekatan tetangga *data testing* beserta *data training*. Disajikan dokumen berjumlah 5 yang sudah melalui tahapan *text preprocessing* untuk 4 *data training* dan 1 *data testing* pada Tabel 5.

Tabel 5. Data Training dan Testing.

Term	Sentimen
Fine lebih lembab ga kering jerawat bruntus parah makin radang jerawat bruntus makin parah fine fine	Negatif
Tekstur ringan sangat lembab finisih dewy matte nyaman	positif
Greasy ga nyaman tambah ga nyaman minyak sedih ga bisa	negatif

Data pada Tabel 5 akan dihitung menggunakan rumus *Cos Similarity*. Hasil terdekat akan diasumsikan sebagai nilai kelas *data testing* tersebut. Sebagai contoh dalam tabel dibawah akan disajikan hasil perbandingan jumlah *Cos Similarity* dari 4 *data training* dan 1 *data testing* yang merujuk pada nilai data.

Tabel 6. Tabel Perhitungan *Cos Similarity*.

Jarak Dts - Dtr	Cos similarity	Ranking
Cos (D5 ,D1)	0,01348911	3
Cos (D5 ,D2)	0,01664091	2
Cos (D5 ,D3)	0	4
Cos (D5 ,D4)	0,10002597	1

Keterangan:

Dtr = Data Training

Dts = Data Testing

Dari Tabel 6 dapat diketahui bahwa; (1) jika $k=1$, maka hasil peringkat 1 (D4) dengan hasil $Cos Similarity = 0,10002597$ sehingga akan menghasilkan prediksi D5 (Data Testing) akan bernilai sama dengan D4 (Data Training). (2) jika $k = 3$, maka hasil peringkat 1 s.d 3 (D4,D2,D1) terdapat 2 Positif dan 1 Negatif, maka bisa disimpulkan D5 termasuk ke dalam kelas positif.

Setelah dilakukan *modelling* akan dihasilkan hasil evaluasi berupa nilai *Accuracy*, *Precision* dan *Recall* dari masing-masing model *brand*. Nilai dari *Accuracy* yang akan dijadikan sebagai acuan untuk melihat *performance* dari algoritma tersebut untuk mengukur jarak kedekatan hasil pengukuran dan nilai sebenarnya. Sebagaimana tujuan dari penelitian ini bahwa algoritma PSO berguna untuk menaikkan akurasi hasil klasifikasi dibandingkan dengan hanya menggunakan algoritma K-NN saja. Hal ini dibuktikan pada Tabel 7.

Tabel 7. Perbandingan hasil evaluasi.

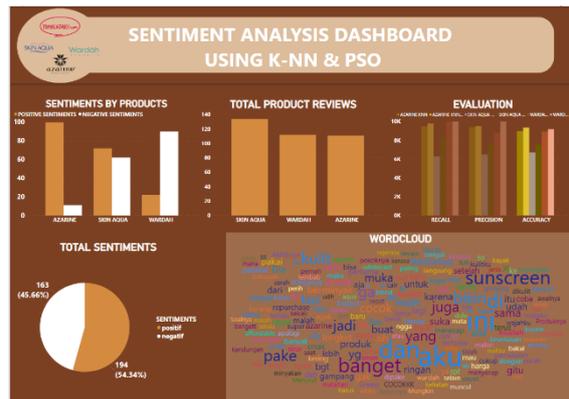
	k-NN	PSO
Azarine	<i>Accuracy</i> =91,89%	<i>Accuracy</i> =92,80%
	<i>Precision</i> = 100%	<i>Precision</i> = 100%
	<i>Recall</i> = 91,74%	<i>Recall</i> = 92,59%
Wardah	<i>Accuracy</i> =89,32%	<i>Accuracy</i> =90,30%
	<i>Precision</i> = 100%	<i>Precision</i> = 100%
	<i>Recall</i> = 88,24%	<i>Recall</i> = 89,11%
Skin Aqua	<i>Accuracy</i> =64,95%	<i>Accuracy</i> =71,76%
	<i>Precision</i> =62,71%	<i>Precision</i> =79,03%
	<i>Recall</i> = 59,68%	<i>Recall</i> = 66,22%

Berdasarkan Tabel 7, *performance* klasifikasi optimasi oleh PSO pada seluruh brand memiliki akurasi lebih tinggi dibandingkan dengan klasifikasi hanya dengan menggunakan K-NN. Hasil akurasi tertinggi yang didapatkan berasal dari produk *sunscreen* brand ‘Azarine’ dengan rincian *Accuracy* menggunakan K-NN 91,89% dan 92,80% menggunakan K-NN & PSO. Hal ini pun membuktikan bahwa hasil akurasi berbanding lurus dengan peringkat *sunscreen* pada artikel yang dirilis oleh Infografik.

Visualisasi dilakukan dengan cara pembuatan *dashboard* dinamis pada *tools Microsoft Power BI*. Dengan judul *dashboard “Sentiment Analysis Dashboard Using K-NN & PSO”* (disajikan pada Gambar 5). Tujuan dibuatnya dashboard ini ialah untuk mempermudah pembacaan hasil klasifikasi beserta nilai akurasi yang didapatkan. Selain itu, dalam dashboard ini ditampilkan pula jumlah dari data keseluruhan ketiga produk, baik sentimen negatif dan sentimen positifnya. Di dalam *dashboard*, terdapat beberapa grafik diantaranya;

1. Grafik *sentiments by products*, menunjukkan jumlah sentimen positif (kolom warna merah muda) dan sentimen negatif (kolom warna putih) dari masing-masing brand *sunscreen*.

2. Disamping nya terdapat grafik total product reviews yang menunjukkan jumlah reviews (data) dari masing-masing brand *sunscreen*.
3. Pada grafik ketiga dengan judul *evaluation*, dapat dilihat tingkat *accuracy*, *precision*, dan *recall* dari setiap brand *sunscreen*.
4. Total sentimen baik negatif maupun positif dari keseluruhan brand dapat dilihat melalui grafik jenis *pie chart*.
5. Selain berbentuk grafik, disajikan pula bentuk *wordcloud* dari data *reviews*.



Gambar 5. Dashboard Visualisasi.

4. DISKUSI

Pada penelitian ini, metode pengambilan data menggunakan data scrapping dengan bantuan Google Extension, metode ini sangat mempermudah pengambilan data komen melalui website khususnya website *reviewer* karena menggunakan tools yang sudah ada pada browser. Hal ini tentu lebih efisien dibandingkan dengan pengambilan data menggunakan metode data crawling khususnya dengan bahasa pemrograman Python yang cenderung menggunakan aplikasi tambahan lainnya. Sama halnya dengan pengambilan data menggunakan aplikasi Orange, dengan metode ini data yang dihasilkan formatnya lebih rapi dan teratur.

Pengolahan data dimulai dari text preprocessing hingga pengklasifikasian sentimen dilakukan dengan menggunakan aplikasi Rapid Miner. Rapid miner mengolah data dengan format .xls. dengan bantuan operator yang dibutuhkan, misalnya tahapan tokenization digunakan operator tokenize.

Pembuatan dashboard menggunakan aplikasi Microsoft Power BI. Tools wordcloud yang hanya dimiliki oleh Microsoft Power BI menjadi salah satu daya tariknya. Penggunaan wordcloud membantu visualisasi dashboard lebih maksimal. Selain itu, penggunaan aplikasi ini pun karena secara umum digunakan oleh peneliti lain karena tools dan extension yang mendukungnya sangat lengkap.

5. KESIMPULAN

Berdasarkan hasil penelitian ini dapat diambil kesimpulan bahwa analisis sentimen data *review 3 brand* produk *sunscreen* yang diambil menggunakan metode scrapping menggunakan *Google Extension*. Proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* dilakukan dengan menggunakan bantuan *tools Rapid Miner*. Pada *Rapid Miner* pula di optimasi menggunakan algoritma *Particle Swarm Optimization*. Diketahui perbedaan yang signifikan antara tingkat akurasi yang dihasilkan oleh algoritma K-NN saja terhadap hasil klasifikasi setiap brand yang di optimasi oleh algoritma *Particle Swarm Optimization*. Brand 'Azarine' menjadi pemegang nilai akurasi terbesar sebesar 91,89% menggunakan K-NN dan 92,80% menggunakan K-NN & PSO. Perhitungan K-NN dapat dilakukan dengan menggunakan rumus *Cos Similarity* yang berbasis jarak kedekatan tetangga.

DAFTAR PUSTAKA

- [1] E. Fadilah Mumtazah *et al.*, "Pengetahuan Mengenai Sunscreen Dan Bahaya Paparan Sinar Matahari Serta Perilaku Mahasiswa Teknik Sipil Terhadap Penggunaan Sunscreen," *J. Farm. Komunitas*, vol. 7, no. 2, pp. 63–68, 2020.
- [2] S. H. Isfardiyana and S. R. Safitri, "Pentingnya melindungi kulit dari sinar ultraviolet dan cara melindungi kulit dengan sunblock buatan sendiri," *J. Inov. dan Kewirausahaan*, vol. 3, no. 2, pp. 126–133, 2014.
- [3] N. F. Andriani and A. H. Setiawan, "Faktor Berhubungan Konsumen Terhadap Penggunaan Produk Skincare," *Diponegoro J. Kesehat.*, vol. 9, no. 4, pp. 1–8, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jme/article/view/29050>.
- [4] S. V. M. Classifier, N. F. Putri, S. Al Faraby, and M. Dwifebri, "Analisis Sentimen Pada Produk Kecantikan Dari Ulasan Female Daily Menggunakan Information Gain Dan Svm Classifier," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 10068–10079, 2021.
- [5] amaliyah gian f riszki wijayatun pratiwi, sharfina febbi h, dairoh, dwi intan af'idah, qirani rifa a, "analisis sentimen pada review skincare female daily menggunakan metode support vector machine," *J. informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. sentiment analysis, 2021.
- [6] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
- [7] R. Syafaat Amardita and M. Dwifebri Purbolaksono, "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *J. Ris. Komput.*, vol. 9, no. 1, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i1.3793.
- [8] H. Nurrun Muchammad Shiddieqy, S. Paulus Insap, and W. Wing Wahyu, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, no. February, pp. 57–64, 2016.
- [9] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [10] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "COMPARISON OF CLASSIFICATION ALGORITHM AND FEATURE SELECTION IN PERBANDINGAN ALGORITMA KLASIFIKASI DAN FEATURE SELECTION," *J. Tek. Inform.*, vol. 3, no. 3, pp. 739–744, 2022.
- [11] A. Muzaki and A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier," *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [12] A. Yoga Pratama *et al.*, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 897–910, 2021.
- [13] T. Setiyorini, ; Rizky, and T. Asmono, "Implementation of K-Nearest Neighbor and Gini Index Method in Classification of Student Performance," *Techno Nusa Mandiri J. Comput. Inf. Technol.*, vol. 16, no. 2, pp. 121–126, 2019, [Online]. Available: <https://ejournal.nusamandiri.ac.id/index.php/techno/article/view/747>.
- [14] R. Samuel, R. Natan, and U. Syafiqoh, "Penerapan Cosine Similarity dan K-Nearest Neighbor (K-NN) pada Klasifikasi dan Pencarian Buku," *J. Big Data Anal. Artif. Intell.*, vol. 1, no. 1, pp. 9–14, 2018.
- [15] A. Taufik, "Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes," *J. Tek. Komput.*, vol. III, no. 2, pp. 40–47, 2017.

- [16] S. Sudianto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study : Internet Selebgram Rachel Venny Escape From Quarantine) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *J. Tek. Inform.*, vol. 3, no. 1, pp. 141–145, 2022.