

ANALISIS SENTIMEN PRODUK KECANTIKAN JENIS *MOISTURIZER* DI TWITTER MENGGUNAKAN ALGORITMA SUPER VECTOR MACHINE

Ria Cantika Larasati¹⁾, Christine Dewi²⁾, Henoch Juli Christanto³⁾

¹⁻³ Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana

email: 672018413@student.uksw.edu, christine.dewi@uksw.edu, henoch.christanto@atmajaya.ac.id

Abstract



This study employs the Support Vector Machine (SVM) algorithm for sentiment analysis of facial moisturizer products based on user opinions extracted from Application X, a prominent social media platform. A dataset of 50,868 tweets in Indonesian was collected from January 1, 2022, to October 31, 2023, with sentiment labels assigned using the TextBlob library. Of these, 35,565 tweets were used for training and 15,303 for testing. SVM classification achieved a high accuracy rate of 98%, with notable confusion matrix and recall values both peaking at 98% and 100%, respectively. The study aimed to assess whether sentiments expressed on Application X regarding moisturizer products leaned predominantly positive or negative. This research contributes to understanding public sentiment towards skincare products, providing insights valuable for product providers to gauge consumer perceptions effectively.

Keywords: *Sentiment Analysis, Facial Moisturizer, Support Vector Machine, Textblob*

1. PENDAHULUAN

Keberadaan produk perawatan wajah yang kian merebak di pasaran menjadi sorotan bagi masyarakat, khususnya kaum hawa [1]. Perkembangan produk perawatan wajah dengan berbagai macam varian harga dan komposisi serta cepatnya pertukaran arus informasi di media sosial membuat masyarakat memiliki kesadaran akan pentingnya merawat wajah, tubuh, hingga badan [2]. Dengan beragamnya pilihan produk perawatan kulit, setiap individu memiliki kesempatan untuk menemukan produk yang cocok dengan jenis kulit dan kebutuhan perawatan kulitnya sendiri [3]. Pelembab (*moisturizer*) wajah menjadi salah satu produk dasar dalam rangkaian perawatan wajah. Setiap produk *moisturizer* memiliki kekurangan dan kelebihan tergantung dengan komposisi yang terkandung di dalamnya [4]. Dalam menentukan produk *moisturizer* yang akan dibeli dapat dilakukan dengan memeriksa ulasan-ulasan pada internet. Ulasan yang diketik sesuai dengan pengalaman yang dirasakan oleh orang lain akan memiliki pengaruh terhadap keputusan pembelian suatu produk [5]. Salah satu aplikasi yang banyak digunakan dalam berbagi ulasan adalah Twitter.

Twitter yang kemudian berganti nama menjadi X merupakan sebuah platform media sosial bertipe *micro-blogging* yang memiliki jumlah maksimal 280 karakter untuk setiap kicauan (*tweet*) [6]. Dalam hal ini, X menjadi salah satu media sosial yang digemari masyarakat sebagai media pertukaran informasi baik itu dari segi pengalaman pengguna itu sendiri maupun ilmu pengetahuan. [7]. Pengguna X dapat melontarkan pesan yang bersifat positif, negatif, maupun netral terhadap suatu produk, layanan, ataupun personal [7]. Kebebasan berpendapat melalui *tweet* inilah yang menjadikan X sebagai salah satu platform yang dipercaya untuk melihat ulasan-ulasan produk *moisturizer*. Untuk menganalisa produk *moisturizer* lebih banyak memiliki ulasan positif atau negatif diperlukan sebuah analisis sentimen.

Analisis sentimen, atau yang sering disebut juga dengan penambangan opini (*opinion mining*), merupakan teknik yang digunakan untuk menganalisis dan mengklasifikasikan ulasan dari kata, kalimat, maupun dokumen dengan tujuan untuk memahami sentimen atau pendapat yang terkandung di dalamnya [8]. Tujuannya adalah untuk memperoleh opini pengguna di suatu platform. Sentimen dapat dibagi menjadi emosi



positif, negatif, dan emosi netral [9]. Tujuan dan manfaat penelitian ini adalah mengetahui sentimen dari ulasan mengenai *moisturizer* wajah. Penelitian ini juga dapat memberikan informasi kepada penyedia produk tentang pandangan masyarakat terkait produk *moisturizer* wajah.

Sentimen analisis dapat dilakukan dengan beberapa metode klasifikasi seperti model Naïve Bayes (NB), k-Nearest Neighbor (k-NN), dan Support Vector Machine (SVM) [10].

Penelitian yang dilakukan oleh Christine Dewi dan Rung-Ching Chen dalam jurnal yang berjudul “*Complement Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database*” mendapatkan akurasi tertinggi sebesar 75,13% untuk klasifikasi database sentimen film dengan model Naïve Bayes. [8].

Penelitian yang dilakukan oleh Irbah dan Sibaroni yang berjudul “Multi Aspect Sentiment of Beauty Product Reviews using SVM and Semantic Similarity” menggunakan metode klasifikasi SVM. Penelitian tersebut menggunakan kategori produk serum & essence, toner, scrub & exfoliating, serta sunscreen. Tiga aspek yang diteliti adalah harga, kemasan, dan aroma. Akurasi optimal yang dicapai adalah 93% untuk aspek harga, 92% untuk aspek kemasan, dan 86% untuk aspek aroma. [11]

Selanjutnya untuk penelitian yang dilakukan oleh Ristyani Slamet, et.al yang berjudul “Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal” membandingkan nilai akurasi pada model klasifikasi Naïve Bayes dan Support Vector Machine (SVM). Penelitian tersebut menghasilkan nilai akurasi model Naïve Bayes sebesar 82.65% dengan Precision 87.81%, dan nilai Recall sebesar 92.11%. Untuk model SVM sendiri memperoleh nilai *accuracy* sebesar 83.60% dengan nilai precision 83.86% dan nilai recall sebesar 99.62%. Dari penelitian ini dapat disimpulkan bahwa model SVM memiliki akurasi lebih tinggi dibandingkan dengan model Naïve Bayes. [7]

Penelitian yang berjudul “*Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database*” yang dilakukan oleh Christine Dewi, et.al menyebutkan bahwa tahap *preprocessing* teks adalah fase paling penting

dalam sistem *Natural Language Processing*. Penelitian ini melakukan kombinasi multinomial *Naïve Bayes* dan *TF-IDF* mencapai akurasi tertinggi sebesar 87,63% [8].

Selanjutnya pada penelitian Angelina Puput Giovani, et.al dari penelitian berjudul “Analisis Sentimen Aplikasi Ruang Guru di Twitter menggunakan Algoritma Klasifikasi” membandingkan model Naïve Bayes, SVM, K-Nearest Neighbor (k-NN) dengan *feature selection* dengan algoritma Particle Swarm Optimization (PSO), dengan aplikasi sentiment analisis pada aplikasi Ruang Guru di Twitter. Hasilnya menunjukkan bahwa model terbaik adalah SVM dengan fitur seleksi menggunakan algoritma PSO, yang menghasilkan akurasi sebesar 78,55% [12].

Penelitian yang dilakukan oleh Christine Dewi et al dalam penelitian berjudul “*Improvement of Support Vector Machine for Predicting Diabetes Mellitus with Machine Learning Approach*” membandingkan dua metode klasifikasi yaitu SVM dan *Naïve Bayes* untuk memprediksi diabetes melitus. Hasilnya menunjukkan bahwa SVM dengan *kernel radial basis function (RBF)* menghasilkan klasifikasi terbaik dengan tingkat akurasi mencapai 98,25% [13].

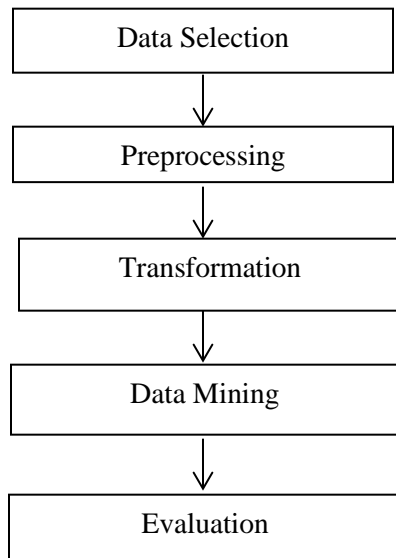
Dari penelitian sebelumnya menunjukkan bahwa *Support Vector Machine (SVM)* memiliki akurasi tertinggi dibandingkan metode klasifikasi *Naive Bayes* dan *K-Nearest Neighbor (SVM)* dikenal sebagai salah satu metode yang digunakan untuk klasifikasi dan regresi [14]. Dalam konteks klasifikasi, SVM memiliki konsep yang lebih matang dan sistematis [15]. Penelitian ini akan mengevaluasi analisis sentimen dari ulasan produk kecantikan wajah jenis *moisturizer* menggunakan metode klasifikasi SVM untuk menganalisis apakah ulasan di aplikasi X mengenai produk *moisturizer* lebih banyak mengandung sentimen negatif atau positif. Penelitian ini dilakukan dengan mengambil data dari X dari tanggal 1 Januari 2022 hingga 31 Oktober 2023 menggunakan kata kunci (*keywords*) “*moisturizer*” dengan *tweet* berbahasa Indonesia dan dianalisis menggunakan metode klasifikasi *Support Vector Machine (SVM)*. Penelitian ini bertujuan untuk tingkat ketepatan yang diperoleh metode klasifikasi SVM dan juga mengidentifikasi seberapa

berpengaruhnya penggunaan produk perawatan wajah berjenis *moisturizer* berdasarkan X.

2. METODE PENELITIAN

Dalam penelitian ini, pendekatan yang digunakan adalah *Knowledge Discovery in Database (KDD)*, yang dikenal karena kemampuannya dalam mengungkapkan pola terstruktur dari kumpulan data yang kompleks, sehingga memudahkan pemahaman data tersebut [7].

Tahapan metode penelitian yang dilakukan berupa *data selection* (pemilihan data), *preprocessing* (pra pemrosesan), *transformation* (transformasi), *data mining* (penambangan data), dan *evaluation* (evaluasi) seperti yang ditampilkan pada gambar 1.

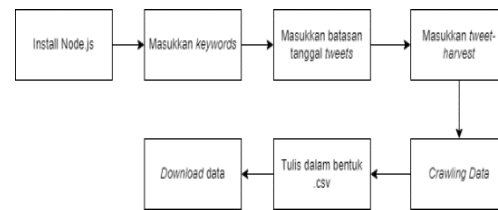


Gambar 1. Metode Penelitian

2.1 Data Selection

Proses *Data Selection* merupakan proses mengumpulkan dengan cara melakukan *data crawling* di aplikasi X. Proses pertama adalah menentukan *keywords* yang sesuai kemudian mengidentifikasi kata kunci (*keywords*) yang relevan, kemudian memanfaatkan *Google Collaboratory* sebagai alat untuk melakukan pengambilan data. Pada penelitian ini menggunakan *keywords: moisturizer*. Data

dengan *keywords* tersebut diambil dari tanggal 1 Januari 2022 hingga 31 Oktober 2023 berbahasa Indonesia. Aplikasi X memberikan batasan dalam *scraping data* Twitter sebanyak 8000 data untuk akun yang telah terverifikasi, 800 data untuk akun lama tidak terverifikasi, dan 400 data untuk akun baru tidak terverifikasi. Ketentuan ini menyebabkan keterbatasan dalam *crawling data* dimana harus menetapkan batasan tanggal *tweets* untuk mengambil data tersebut perhari. Proses *data selection* ditampilkan dalam gambar 2.



Gambar 2. Proses Crawling Data Twitter

Setelah itu, data yang berhasil diambil dari proses *crawling* disimpan ke dalam *CSV (Comma Separated Values)* sehingga dapat digunakan pada tahap selanjutnya.

2.2 Preprocessing

Setelah selesai melakukan pemilihan data pada tahap *data selection*, langkah selanjutnya adalah tahap *preprocessing*. Pada tahap ini, data dibersihkan untuk menghilangkan *noise* dan disiapkan untuk analisis sehingga hasil perhitungan optimal dapat dibaca [16]. Proses pada *preprocessing data* terdapat beberapa tahapan, yaitu:

2.2.1 Cleaning Data

Data yang telah diberi label akan dilakukan pembersihan dari duplikat, karakter, maupun tanda baca (punctuation). Contoh tanda baca tersebut adalah tanda seru, titik, koma, tanda tanya, dan lain sebagainya. [17]

2.2.2 Case Folding

Case Folding adalah proses mengubah huruf dalam teks menjadi huruf kecil (*lowercase*) [7].

2.2.3 Stopword Removal

Stopword removal merupakan proses menghilangkan kata-kata yang tidak memberikan makna signifikan dalam sebuah teks, seperti kata-kata sambung “di”, “dan”, “yang”, dll.

2.2.4 *Tokenization*

Tokenization merupakan tahap pemecahan sebuah kalimat menjadi kata. Proses ini bertujuan untuk mengurai kata-kata dalam sebuah kalimat dengan memisahkan kalimat tersebut menjadi unit-unit kata, sehingga memungkinkan analisis lebih lanjut terhadap setiap kata tersebut [18].

2.2.5 *Stemming Data*

Stemming adalah proses mengubah kata-kata dalam sebuah teks menjadi bentuk dasarnya dengan menghilangkan imbuhan atau afiksasi yang melekat pada kata tersebut [19].

2.2.6 *Labeling*

Langkah berikutnya adalah melakukan pelabelan pada data latihan ulasan menggunakan *TextBlob*. *TextBlob library* yang dimiliki *Python* dapat digunakan untuk memproses *Natural Language Processing (NLP)*. [20] Pelabelan data ini dibagi menjadi 3 kelas : data dengan sentimen negatif, sentimen positif, dan sentimen netral. Dengan menggunakan analisis sentimen yang membedakan antara ulasan positif, negatif, dan netral, akan mempermudah konsumen yang ingin menemukan ulasan produk, karena mereka tidak perlu lagi membaca semua ulasan secara menyeluruh. [21]

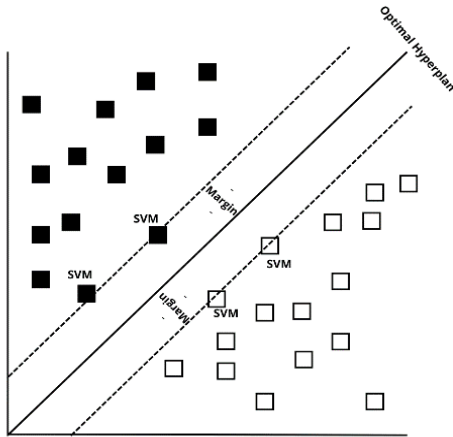
2.3 *Transformation*

Tahap *transformation* merupakan proses vektorisasi yang bertujuan untuk memberikan bobot pada kata-kata dalam sebuah *tweet* sehingga *tweet* tersebut dapat dikenali polanya saat masuk ke dalam tahap data mining.

2.4 *Data Mining*

Pada tahap ini dilakukan penerapan Model SVM. SVM merupakan metode pembelajaran yang memanfaatkan hipotesis dalam bentuk fungsi dan fitur yang memiliki dimensi tinggi [23]. Proses pelatihan SVM dilakukan dengan menggunakan algoritma pembelajaran berdasarkan teori optimasi [24]. SVM memiliki keunggulan dalam kinerjanya yang efektif dalam ruang berdimensi tinggi, efisiensi yang tinggi ketika jumlah fitur lebih besar dari jumlah sampel, dan penggunaan memori yang efisien karena hanya menggunakan subset titik pelatihan [25].

Tahap awal pada proses ini adalah pembagian data menjadi 2, yaitu 70% digunakan untuk data latih dan 30% digunakan untuk data uji [26]. Data latih digunakan untuk mengidentifikasi pola yang diinginkan, sementara data uji digunakan untuk menguji validitas penemuan yang dihasilkan dari data latih [17]. Pada dasarnya, metode ini berfungsi dengan menetapkan batas antara dua kelas dengan memastikan jarak maksimal dari data yang terdekat. Untuk mencapai batas maksimal antara kelas-kelas tersebut, diperlukan pembentukan *hyperplane* terbaik di dalam ruang input, yang didasarkan pada pengukuran margin *hyperplane* dan penemuan titik maksimumnya [27]. Margin adalah jarak antara *hyperplane* dan titik terdekat dari setiap kelas dalam data. Ini merupakan konsep inti dari SVM [28]. SVM memiliki kemampuan untuk melakukan klasifikasi data yang dapat dipisahkan baik secara *linear* maupun *non-linear* [17]. Visualisasi dari klasifikasi menggunakan model SVM tersedia pada gambar 3.



Gambar 3. Metode SVM

Proses pengklasifikasian data uji akan menggunakan *library SVM*.

2.4 Evaluation

Tahap evaluasi dilakukan untuk mengevaluasi kualitas dari suatu *classifier*. Suatu sistem klasifikasi perlu dievaluasi performanya untuk mengukur akurasi dari prediksi klasifikasi yang dibuat [25]. Setelah tahap tersebut, perbandingan hasil dari dua dataset yang berbeda dievaluasi menggunakan *confusion matrix* untuk mengukur presisi (*precision*), *recall*, *f1-score*, dan akurasi (*accuracy*) [26].

Hasil dari *confusion matrix* diilustrasikan pada tabel 1.

Tabel 1. ilustrasi hasil *confusion matrix*

AKTUAL	PREDIKSI		
	NEGATIF	NETRAL	POSITIF
NEGATIF	TN	FNet	FP
NETRAL	FNetN	TNet	FNetP
POSITIF	FPN	FPNet	TP

Dari kolom aktual di tabel 1 menjelaskan sentimen dari data uji yang telah diberikan label, sedangkan kolom prediksi berisi hasil label sentimen dari model yang telah kita latih sebelumnya. Umumnya semakin tinggi nilai TN, TP, dan TNet, maka semakin baik pula hasil klasifikasi dengan model tersebut [29], [30], [33].

3. HASIL DAN PEMBAHASAN

Data Selection

Data Selection dilakukan dengan cara melakukan *crawling data* berupa *tweet* yang berisi opini dari aplikasi X menggunakan *keywords 'moisturizer'* berbahasa Indonesia. *Crawling data* ini memanfaatkan *tweet-harvest* yang di bangun menggunakan *Node.js*. Proses ini mengumpulkan 50.868 data opini yang didapat dari tanggal 1 Januari 2022 hingga 31 Oktober 2023. Data yang dihasilkan mencakup *created_at*, *full_text*, *username*, dan *tweet_url*.

Preprocessing Data

Setelah data terkumpul, selanjutnya data masuk ke dalam tahap *preprocessing* yang berfungsi untuk membersihkan data dari duplikasi, kata-kata dan symbol yang tidak diperlukan, serta pemberian label pada data. Data yang masuk ke dalam tahap *preprocessing* ini adalah data dari kolom '*tweet*'.

Cleaning Data

Cleaning data membersihkan data "*full_text*" yang berisi *tweet* dari *url*, *html*, *emoji*, penomoran, dan simbol.

Case Folding

Dalam proses ini, *tweet* yang telah berhasil dibersihkan dan mengandung huruf kapital dan huruf kecil diubah menjadi huruf kecil seluruhnya. Proses ini dapat mempermudah mesin dalam memproses data.

Tokenization

Proses *Tokenization* merupakan proses pemecahan kalimat opini menjadi potongan-potongan kata yang membentuk opini tersebut. Secara umum setiap kata teridentifikasi atau dipisahkan dari kata lainnya dengan spasi, sehingga proses *tokenization* melibatkan karakter spasi dalam opini untuk memisahkan setiap kata. [34]

Stopword Removal

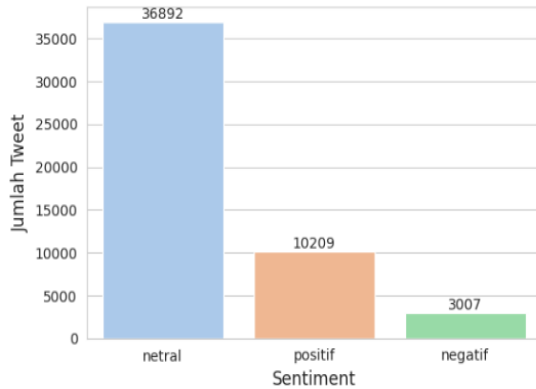
Proses ini memanfaatkan *stopwords* berbahasa Indonesia yang dimiliki oleh *library NLTK*.

Stemming Data

Proses ini memperkecil indeks dari opini yang telah melalui proses *tokenization* tanpa menghilangkan makna dari opini tersebut. Stemming dapat mengurangi beban memori selama pelatihan model dan klasifikasi [35].

Labeling Data

Dalam melakukan *Labeling data* menggunakan *sklearn library* untuk membagi antara data latih dan data uji. Dengan menggunakan *library* tersebut, diperoleh jumlah data latih sebanyak 35.075 dan data uji sebanyak 15.033 yang diambil secara acak dengan menggunakan fungsi *'train_test_split'*. Hal ini memastikan data yang diambil adalah representatif dari seluruh dataset. Hasil *data labeling* dengan *TextBlob* dapat dilihat pada gambar 4.

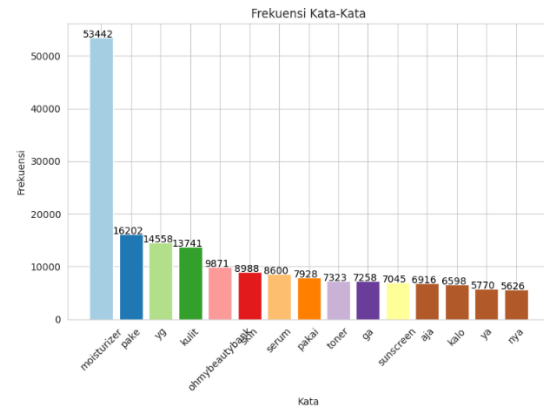


Gambar 4. Visualisasi Hasil Labeling Data

Gambar 4 menampilkan gabungan label dari data latih dan data uji. Dari data-data tersebut, dapat divisualisasikan menggunakan *word cloud* untuk melihat kata-kata yang sering muncul. Hasil visualisasi *word cloud* dapat dilihat pada gambar 5, sementara diagram batang yang menunjukkan frekuensi kata yang sering muncul dapat dilihat pada gambar 6.



Gambar 5. Visualisasi Word cloud



Gambar 6. Kata yang Sering Muncul

3.2 Transformation

Tahap *transformation* mencakup tahap vektorisasi atau memberikan nilai pada kata dalam *tweet* opini agar setiap kata dapat dicari polanya ketika masuk ke dalam tahap klasifikasi. Vektorisasi ini menggunakan rumus TF-IDF yang telah tertera pada persamaan (1). Hasil perhitungan TF-IDF dapat dilihat dari tabel 1 dimana perhitungan tersebut diambil dari teks sampel yang terdapat pada tabel 2.

Tabel 1. Teks Sampel

Tweet Hasil Preprocessing	
D1	butuh toner hatamogi serum the ordinary moisturizer skintific
D2	sbyfess mending moisturizer bright stuff emina yg spfnya bagus
D3	pake sunscreen ga sih dipake umur rekomendasi sunscreen bagus moisturizer remaja

Tabel 2. Sampel data yang telah ditransformasi

Term	D F	IDF log(N/DF)	TF-IDF		
			D1	D2	D3
butuh	1	0.698	0.069	0	0
		97	897		
toner	1	0.698	0.069	0	0
		97	897		
hatam	1	0.698	0.069	0	0
		97	897		
ogi	1	0.698	0.069	0	0
		97	897		
serum	1	0.698	0.069	0	0
		97	897		
the	1	0.698	0.069	0	0
		97	897		

ordinar	1	0.698	0.069	0	0
y		97	897		
moistu	3	0.221	0.022	0.018	0.017
rizer		849	185	487	065
skintifi	1	0.698	0.069	0	0
c		97	897		
sbyfes	1	0.698	0	0.058	0
s		97		248	

Untuk melakukan vektorisasi dengan TF-IDF menggunakan kode program 1.

Kode Program 1. Vektorisasi dengan TF-IDF

```
from sklearn.feature_extraction.text
import TfidfVectorizer

#inisialisasi TF_IDF Vectorizer

tfidf_vectorizer = TfidfVectorizer()

X_train_tfidf =
tfidf_vectorizer.fit_transform(X_train)

X_test_tfidf =
```

Dengan melakukan kode program 1 dapat melakukan transformasi menggunakan TF-IDF Vectorizer sehingga menghasilkan representasi numerik dari teks yang siap digunakan untuk melatih model SVM.

3.3 Data Mining

Proses Data Mining di sini adalah proses pelatihan dari model *Support Vector Machine* (SVM). Langkah pertama dalam proses pelatihan SVM ini adalah menginisialisasi model SVM menggunakan kernel linear. Kernel linear dipilih karena memungkinkan model untuk menciptakan pemisah linear dalam ruang fitur yang tinggi, yang dapat berguna untuk menangani masalah klasifikasi teks. Model SVM kemudian dilatih menggunakan data pelatihan yang telah divektorisasi menggunakan TF-IDF yang telah dilakukan pada tahap *transformation*. Proses pelatihan ini bertujuan untuk mempelajari pola dan struktur data yang relevan dalam teks dan memungkinkan model SVM untuk membangun batas keputusan yang memisahkan kelas-kelas

yang berbeda dengan optimal. Dalam proses pelatihan model SVM sendiri tertera pada kode program 2.

Kode Program 2. Melatih model SVM

```
svm_model = SVC(kernel='linear',
random_state=42)

svm_model.fit(X_train_tfidf, y_train)

#inisialisasi model SVM

svm_model = SVC(kernel='linear',
random_state=42)

#Melatih model dengan data pelatihan yang
sudah di vektorisasi

svm_model.fit(X_train_tfidf, y_train)

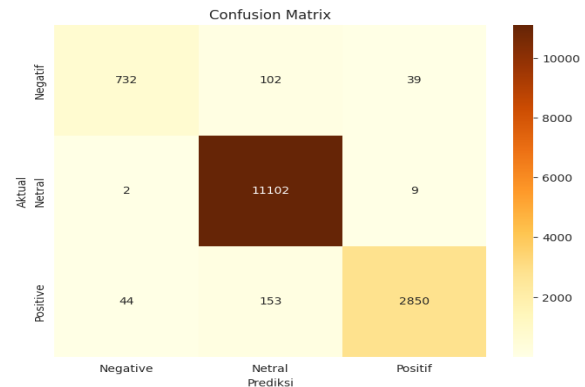
print(f"Support Vectors:
{svm_model.support_vectors}")
```

Pada kode program 2, parameter dari model SVM yang telah dilatih, dievaluasi. Parameter-parameter ini mencakup kernel yang digunakan, nilai C (parameter regulasi yang mengontrol *trade-off* antara *margin* dan kesalahan klasifikasi), *intercept* (bias), dan *support vectors* (titik-titik data yang paling dekat dengan batas keputusan). Untuk nilai C (*Cost*) sendiri menggunakan 1.0.

Dari kode program 2 menghasilkan *output* seperti yang ada pada gambar 7.

```

Parameter model SVM:
Kernel: linear
C: 1.0
Intercept: [-0.95608763 -0.36014963 0.88681814]
Support Vectors: (0, 1906) 0.17147651653963625
(0, 4310) 0.17916844541675384
(0, 4655) 0.25340427881269817
(0, 6539) 0.20695509326676953
(0, 10319) 0.206629972487371
(0, 12520) 0.1220174043662384
(0, 13244) 0.16624368430211567
(0, 13793) 0.2094837037963438
(0, 17225) 0.21899923626172796
(0, 18944) 0.09968739943819223
(0, 19394) 0.18059768467061676
(0, 20146) 0.2224818532510782
(0, 20782) 0.20299954606519816
(0, 21028) 0.25824408163018814
(0, 22491) 0.039890479960049366
(0, 25385) 0.16271967445721772
(0, 26043) 0.28940557624713464
(0, 26071) 0.21499133408485685
(0, 27676) 0.21361250685551084
(0, 33705) 0.28214386134488934
(0, 34578) 0.3423684050203568
    
```



Gambar 7. Visualisasi *Confusion Matrix*

Gambar 7 Hasil klasifikasi dengan SVM

3.4 Evaluation

Berdasarkan pembagian antara data latih dan data uji, dilakukan klasifikasi dan evaluasi menggunakan model klasifikasi *Support Vector Machine*. Tahap *evaluation* ini dilakukan untuk menilai seberapa akurat model *Support Vector Machine* tersebut. Untuk menentukan keakuratan sebuah model dibutuhkan *confusion matrix*. Hasil dari perhitungan *confusion matrix* dapat dilihat pada tabel 3.

Tabel 3. Hasil *Confusion Matrix*

Aktual	Prediksi		
	Negatif	Netral	Positif
Negatif	732	102	39
Netral	2	11102	9
Positif	44	153	2850

Dari tabel 3 dapat disimpulkan bahwa nilai True Negatif sebesar 732, nilai True Netral sebesar 11.102, dan jumlah True Positif sebesar 2.850. Angka ini didapatkan dari hasil pengujian 15.033 data. Adapun visualisasi dari *confusion matrix* terdapat pada gambar 7.

Dari gambar 7 dapat terlihat bahwa nilai true paling banyak terletak pada True Netral dimana mendapatkan 11.102 data. Adapun opini sebanyak 102 dan 39 yang seharusnya masuk ke dalam kategori negatif tetapi diprediksi netral dan positif oleh model. Kemudian terdapat 2 dan 9 opini yang dikategorikan netral tetapi diprediksi negatif dan positif oleh model. Selanjutnya terdapat 44 dan 153 opini yang dikategorikan sebagai positif tetapi diprediksi negatif dan netral oleh model.

Setelah mengetahui *confusion matrix*, selanjutnya dapat mengetahui nilai precision, recall, dan f1-score. Adapun hasil perhitungannya terdapat pada tabel 4.

Tabel 4. Hasil klasifikasi dengan SVM

	Precisio n	Recal l	F1- scor e	Suppor t
Negatif	0.94	0.84	0.89	873
Netral	0.98	1.00	0.99	11113
Positif	0.98	0.94	0.96	3047
Accuracy			0.98	15033
Macro avg	0.97	0.92	0.94	15033
Weighte d avg	0.98	0.98	0.98	15033

Dari tabel 4 dapat dilihat bahwa model SVM untuk melakukan klasifikasi sentimen mendapatkan nilai *precision* sebesar 97%, nilai *recall* sebesar 93%, nilai *f1-score* sebesar 95%, dan nilai *accuracy* sebesar 98%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa mayoritas masyarakat memiliki kecenderungan netral terhadap produk moisturizer yang tengah beredar di masyarakat saat ini. Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk melakukan analisis sentimen terhadap ulasan produk moisturizer di platform X berdasarkan 50,868 tweet opini. Metode SVM berhasil mencapai tingkat akurasi sebesar 98%, dengan nilai tertinggi untuk matriks kebingungan dan recall mencapai 98% dan 100% masing-masing, serta f1-score sebesar 99%. Hasil ini menunjukkan bahwa SVM sangat efektif dalam mengklasifikasikan sentimen positif, negatif, dan netral terhadap produk moisturizer, memberikan wawasan yang berharga bagi produsen untuk memahami pandangan konsumen terhadap produk mereka di platform sosial media seperti X..

5. REFERENSI

- [1] D. W. R. Wati, N. Fatmawatie, and N. Fauza, "Pengaruh Customer Satisfaction Dan Customer Trust Terhadap Customer Loyalty Produk Krim Pelembab Wajah Fair & Lovely," *Istithmar J. Stud. Ekon. Syariah*, vol. 4, no. 1, pp. 50–71, 2020, doi: 10.30762/istithmar.v4i1.4.
- [2] M. E. T. Butarbutar and A. Y. Chaerunisaa, "Peran Pelembab dalam Mengatasi Kondisi Kulit Kering," *Maj. Farmasetika*, vol. 6, no. 1, pp. 56–69, 2020, doi: 10.24198/mfarmasetika.v6i1.28740.
- [3] I. L. Situmorang, "Pengaruh Kualitas Produk dan Iklan terhadap Citra Merek dan Keputusan Pembelian Produk Kecantikan Merek Pond's," *Suparyanto dan Rosad (2015)*, vol. 5, no. 3, pp. 248–253, 2020, [Online]. Available: <https://jom.unri.ac.id/index.php/JOMFE/KON/article/view/12298>
- [4] S. Fatimah, "Pengaruh Kesadaran Merek, Persepsi Kualitas, Asosiasi Merek, Dan Loyalitas Merek Terhadap Keputusan Pembelian Pelembab Wardah Pada Konsumen Al Yasini Mart Wonorejo," *Sketsa Bisnis*, vol. 1, no. 2, 2014, doi: 10.35891/jsb.v1i2.75.
- [5] E. Putra, "Pengaruh Promosi Melalui Sosial Media Dan Review Produk Pada Marketplace Shopee Terhadap Keputusan Pembelian (Studi Pada Mahasiswa Stie Pasaman) the Influence of Promotion Through Social Media and Product Review on the Marketplace Shopee on Purchase Dec," *J. Apresiasi Ekon.*, vol. 8, no. 3, pp. 467–474, 2020.
- [6] I. Zukhrufillah, "Gejala Media Sosial Twitter Sebagai Media Sosial Alternatif," *Al-I'lam J. Komun. dan Penyiaran Islam*, vol. 1, no. 2, p. 102, 2018, doi: 10.31764/jail.v1i2.235.
- [7] R. Slamet, W. Gata, A. Novtariany, K. Hilyati, and F. A. Jariyah, "Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 145–153, 2022, doi: 10.31539/intecomsv5i1.3933.
- [8] C. Dewi and R. C. Chen, "Complement Naive Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13757 LNAI, pp. 81–93, 2022, doi: 10.1007/978-3-031-21743-2_7.
- [9] C. Dewi, B.-J. Tsai, and R.-C. Chen, "Shapley Additive Explanations for Text Classification and Sentiment Analysis of Internet Movie Database BT - Recent Challenges in Intelligent Information and Database Systems," 2022, pp. 69–80.
- [10] C. Dewi, R.-C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam J.*



- Comput. Sci.*, vol. 10, no. 04, pp. 485–498, 2023, doi: 10.1142/s2196888823500100.
- [11] Irbah salsabila and Yuliant Sibaroni, “Multi Aspect Sentiment of Beauty Product Reviews using SVM and Semantic Similarity,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 520–526, 2021, doi: 10.29207/resti.v5i3.3078.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [13] C. Dewi, J. Zentrato, and H. J. Christanto, “Improvement of support vector machine for predicting diabetes mellitus with machine learning approach,” *J. Auton. Intell.*, vol. 7, no. 2, pp. 1–12, 2023, doi: 10.32629/jai.v7i2.888.
- [14] C. Dewi and R. C. Chen, “Random forest and support vector machine on features selection for regression analysis,” *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 6, pp. 2027–2037, 2019, doi: 10.24507/ijicic.15.06.2027.
- [15] N. W. S. Agustini, D. Priadi, and R. V. Atika, “Profil Kimia dan Aktivitas Antibakteri Fraksi Aktif Nannochloropsis sp. sebagai Senyawa Penghambat Bakteri Penyebab Gangguan Kesehatan Mulut,” *J. Pascapanen dan Bioteknol. Kelaut. dan Perikan.*, vol. 17, no. 1, p. 19, 2022, doi: 10.15578/jpbkp.v17i1.781.
- [16] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [17] H. Nurrin 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.*, vol. 7, no. 2, pp. 57–64, 2016.
- [18] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [19] J. W. Iskandar and Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [20] R. Azhar, A. Surahman, and C. Juliane, “Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, 2022.
- [21] B. M. Iqbal, K. M. Lhaksmana, and E. B. Setiawan, “2024 Presidential Election Sentiment Analysis in News Media Using Support Vector Machine,” *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 397–404, 2023, doi: 10.47065/josyc.v4i2.3051.
- [22] M. Hamka, N. Alfatari, and D. Ratna Sari, “Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 64, 2022, doi: 10.30865/json.v4i1.4740.
- [23] I. P. Monika and M. T. Furqon, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3165–3166, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [24] A. R. Isnain, A. I. Sakti, D. Alita, and N.

- S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [25] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.
- [26] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [27] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [28] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [29] O. Arifin and T. B. Sasongko, "Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier," *Semin. Nas. Teknol. Inf. dan Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018.
- [30] Trivusi, "Metriks Evaluasi Sistem Menggunakan Confusion Matrix," *Trivusi.web.id*, 2022. <https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html> (accessed Feb. 26, 2024).
- [31] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means," *Techno.Com*, vol. 16, no. 1, pp. 1–8, 2017, doi: 10.33633/tc.v16i1.1263.
- [32] Anggreany Maria Susan, "Confusion Matrix," *Bina Nusantara University*, 2022. <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> (accessed Feb. 26, 2024).
- [33] Rina, "Memahami Confusion Matrix: Accuracy, Precision, Recall, Specificity, dan F1-Score untuk Evaluasi Model Klasifikasi," *Medium.com*, 2023. <https://esairina.medium.com/memahami-confusion-matrix-accuracy-precision-recall-specificity-dan-f1-score-610d4f0db7cf> (accessed Feb. 26, 2024).
- [34] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [35] R. Novaneliza, F. Handayani, R. J. Suhandar, H. Surono, N. S. Azzahra, and D. Nadilla, "Perbandingan Algoritma Untuk Analisis Sentimen Pada Twitter Transportasi Umum Commuterline," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 13–21, 2023.