

ANALISIS SENTIMEN PADA PRODUK CUSHION DI WEBSITE FEMALE DAILY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

¹⁾**Soultan Ariqoh, ²⁾M. Agus Sunandar, ³⁾Yusuf Muhyidin**

^{1,2,3)}Teknik Informatika, Sekolah Tinggi Teknologi Wastukancana

¹⁾soultanariqoh11@wastukancana.ac.id, ²⁾agoes.61@wastukancana.ac.id,

³⁾yusufmuhyidin@wastukancana.ac.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 28 Juli 2023

Disetujui : 29 Juli 2023

Kata Kunci :

Text Mining, Female Daily, Pyton,
Support Vector Machine (SVM)

ABSTRAK

Saat ini produk kosmetik sudah menjadi kebutuhan utama kaum wanita yang merupakan target utama dari industri kosmetik. Salah satu merek produk kosmetik yang popular saat ini ialah produk kosmetik merek pixy, saat ini cushion pixy mempunyai tiga jenis shades yaitu 301 Medium Beige, 201 Natural Beige, dan 101 Light Beige. Namun tidak semua jenis shades memiliki kualitas yang baik sesuai kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen sebelum konsumen memutuskan untuk membeli produk cushion merek pixy. Dalam penelitian ini penulis menggunakan metode Text mining Dengan tahap-tahap yang dilakukan adalah Scrapping Data, Labelling, Cleaning, Preprocessing text (Tranformation, Tokenization, Filtering). Algoritma yang digunakan yaitu Support Vector Machine (SVM) karena teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Pembobotan kata menggunakan pehitungan TF-IDF dan pengujian data menggunakan Confusion Matrix pada Python. Berdasarkan proses pengumpulan data diperoleh sebanyak 2044 data, setelah melakukan proses Text Mining dihasilkan sebanyak 1999 data.

ARTICLE INFO

Article History :

Received : July 28, 2023

Accepted : July 29, 2023

Keywords:

Text Mining, Female Daily, Pyton,
Support Vector Machine (SVM)

ABSTRACT

Currently cosmetic products have become the main needs of women who are the main target of the cosmetic industry. One of the most popular cosmetic product brands today is the cosmetic product brand Pixy because the price is affordable and also has good quality. The pixy cushion products currently being produced have three types of shades, namely 301 Medium Beige, 201 Natural Beige, and 101 Light Beige. However, not all types of shades have good quality according to consumer needs and this is what consumers must pay attention to before consumers decide to buy pixy brand cushion products. In this study the authors used the Text mining method with the steps taken were Data Scrapping, Labeling, Cleaning, Preprocessing (Transformation, Tokenization, Filtering). The algorithm used is Support Vector Machine (SVM) because it is a relatively new technique for making predictions, both in the case of classification and regression. Word weighting uses TF-IDF calculations and data testing uses Confusion Matrix in Python. Based on the data collection process, 2044 data were obtained, after carrying out the Text Mining process, 1999 data were produced..

1. PENDAHULUAN

Saat ini produk kosmetik sudah menjadi kebutuhan utama kaum wanita yang merupakan target utama dari industri kosmetik. Semakin banyak merek kosmetik yang bermunculan baik dari dalam negeri maupun luar negri, Salah satu merek produk kosmetik yang popular saat ini ialah produk kosmetik merek Pixy dikarenakan harganya yang terjangkau dan juga memiliki kualitas baik(Indrayuni 2019).

Produk kosmetik Pixy merupakan sebuah merek produk kecantikan yang ada di Indonesia, produk ini berasal dari Jepang yang mampu merebut hati konsumen yang ada di Indonesia. Produk cushion pixy yang diproduksi saat ini mempunyai tiga jenis shades. Namun tidak semua jenis shades memiliki kualitas yang baik sesuai kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen. Sebelum konsumen memutuskan untuk membeli produk cushion merek pixy, sebaiknya konsumen mengetahui detail produk cushion dan jenis shades yang akan dibeli, hal ini dapat dipelajari dari testimoni dan opini atau hasil review dari konsumen yang sudah membeli dan menggunakan produk cushion merek pixy tersebut sebelumnya,

Jumlah review dan opini sangat banyak sehingga akan menyulitkan dan memakan waktu untuk membaca secara keseluruhan. Oleh karena itu dapat dirancang sistem yang secara otomatis akan mengelompokkan opini dan review yang ada sesuai kelasnya. Kelas sentimen dibagi menjadi kelas positif, negatif, dan netral.

Berdasarkan latar belakang yang telah dijabarkan, maka peneliti tertarik untuk mengelompokkan opini dan review pada produk cushion pixy di website female daily menggunakan metode Support Vector Machine (SVM)

2. METODE

Pada perancangan sistem, dibutuhkan metode yang menjadi acuan dan pedoman untuk diimplementasikan pada rancangan yang akan dibuat. Berikut teori yang nantinya akan digunakan yaitu Text Mining, Analisis Sentimen, Algoritma Support Vector Machine (SVM), TF-IDF, Confusion Matrix, dan Pengambilan Data.

2.1 Text Mining

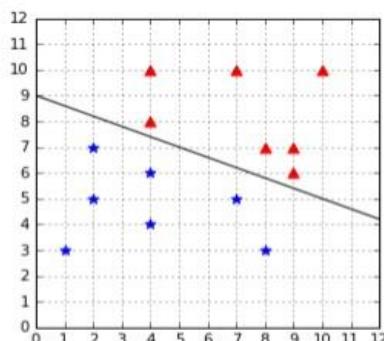
Text mining adalah suatu teknik yang digunakan untuk mengklasifikasi dokumen dimana text mining merupakan sebuah variasi dari data mining yang berusaha menemukan pola yang menarik dari kumpulan data tekstual yang memiliki jumlah besar (Sunoto and Wasito 2014).

2.2. Analisis Sentimen

Analisis Sentimen adalah suatu proses untuk menentukan opini atau sikap seseorang atas suatu produk, layanan, atau organisasi. Sasaran dari analisis sentimen adalah menemukan opini, lalu menandai sentimen yang telah diungkapkan kemudian mengklasifikasikannya (Classifier et al. 2021).

2.3. Support Vector Machine (SVM)

Metode Support Vector Machine (SVM) merupakan suatu sistem pembelajaran dengan menggunakan ruang hipotesis yang berupa fungsi-fungsi linear didalam sebuah sifat yang memiliki dimensi yang tinggi dan dilatih memakai algoritma pembelajaran yang berlandaskan teori optimasi (Monika and Furqon 2018). SVM linear merupakan data yang dipisahkan secara linear yaitu memisahkan dua kelas pada *hyperplane* dengan *soft margin*. Sedangkan non linear yaitu merupakan fungsi dari kernel *trick* terhadap ruang yang berdimensi tinggi. *Support Vector Machine* menggunakan 2 titik (*vector*) yang selanjutnya dua titik tersebut akan membentuk garis pembatas (sisi pembatas jika 3 dimensi atau lebih) garis pembatas yang dibentuk dari dua buah *vector* ini disebut *hyperplane*. Contoh bisa dilihat pada gambar berikut :



Gambar 1 Hypperlan memisahkan dua kelas

Permasalahan non linear dapat diatasi dengan memodifikasi *trick* kernel ke dalam SVM yang akan menjadi pemisah kelas atau *hyperplane* menjadi dua kelas didalam ruang *vector* d

$$f(x) = \sum a_i y_i K(x, x') + b$$

Rumus.1 Kernel Linier

Keterangan :

a_i = alfa ke-i y_i = kelas

data latih ke-i M = jumlah data

$K(x, x')$ = fungsi kernel yang dipakai,

dengan x = data uji

x_i = data latih ke-i b = bias

2.4. TF- IDF

Term Frequency - Inverse Document Frequency (TF- IDF) merupakan suatu proses penyelesaian masalah yang melakukan penggabungan dua metode yaitu konsep frekuensi kemunculan term dalam dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut, dan mampu meningkatkan proporsi jumlah dokumen yang dapat ditemukan kembali dan relevan,. Sehingga kriteria term yang paling tepat adalah term yang selalu muncul dalam dokumen tersebut (Nurjannah and Fitri Astuti 2013).

Rumus TF :

$$TF(k) = \sum F_{f,d}$$

Rumus.2 Term Frequency (TF)

Keterangan :

F = frekuensi kemunculan

k = kata

d = dokumen

Rumus IDF :

$$IDF(k) = \log \frac{D}{D_f(k)}$$

Rumus.3 Inverse Document Frequency (IDF)

Keterangan :

D = total jumlah dokumen

$DF(k)$ = total dokumen yang mengandung frekuensi kemunculan kata k

2.5. Confusion Matrix

Confusion Matrix ialah tabel yang memvisualkan performa dari sebuah model atau algoritma secara khusus dari setiap baris dari

matrix tersebut, merepresentasikan kelas sebenarnya dari data, dan setiap kolom merepresentasikan kelas perkiraan dari data atau sebaliknya (Saputro and Sari 2020).

Matrix tersebut dijelaskan pada table berikut :

Tabel.1 Rumus Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative True	Negative (TN)	False Positive (FP)
Actual Positive False	Negative (FN)	True Positive (TP)

1. *True Positive* = Berarti seberapa banyak data yang aktual kelasnya positif, dan model juga memprediksi positif.
2. *True Negative* = Berarti seberapa banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.
3. *False Positive* = Berarti seberapa banyak data yang aktual kelasnya negatif, namun model memprediksi positif.
4. *False Negative* = Berarti seberapa banyak data yang aktual kelasnya positif, namun model memprediksi negative.

3. HASIL DAN PEMBAHASAN

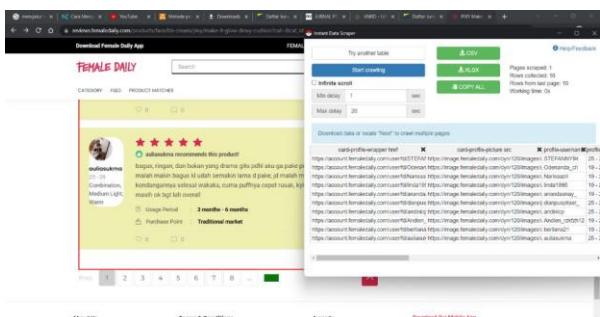
Analisis yang diterapkan kedalam bentuk aplikasi berbasis website memiliki sistem untuk menganalisis sentimen masyarakat terhadap produk cushion pixy pada website female daily. Sistem yang akan dibuat meliputi penarikan data, labelling data, cleaning Data, preprocessing text, classification data, evaluasi dan visualisasi data.Untuk penelitian kualitatif, bagian hasil memuat bagian-bagian rinci dalam bentuk sub topik-sub topik yang berkaitan langsung dengan fokus penelitian dan kategori-kategori.

3.1. Scrapping Data

Pada penelitian ini dilakukan proses pengumpulan data, dimana data yang diambil merupakan data ulasan dari pengguna produk cushion pixy pada website female daily. Proses

pengambilan data dilakukan dengan menggunakan teknik scraping melalui Instant Data Scraper versi 1.0.8 yang dimiliki oleh Google Chrome.

Terdapat beberapa langkah penggeraan dalam melakukan scraping data. Langkah pertama yang dilakukan adalah menginstal aplikasi dan mengaktifkan ekstensi Data Scraper pada Google Chrome. Dari hasil Scrapping data di website female daily diperoleh 2044 data. Proses scrapping bisa dilihat pada gambar berikut :



Gambar.2 Tahapan Scrapping

3.2. Processing Text

Preprocessing Text ialah suatu proses awal yang mengubah data masukan menjadi data dengan format yang sesuai dan siap untuk diproses. Proses yang tersedia terkait dari tujuan yang akan dicapai dalam preprocessing tersebut. Pemilihan proses yang tepat perlu dilakukan mengingat karena proses yang sesuai dalam tahapan preprocessing data akan meningkatkan tingkat pencapaian hasil (Setyoahadi, Kristiawan, and Ernawati 2017).

Setelah dilakukan cleaning data selanjutnya dilakukan tahap processing Text. Tahapan processing text meliputi transformation, tokenizing, dan filtering. Tahapan transformation, tokenizing, dan filtering dapat dilihat pada table berikut :

Tabel.2 Hasil Transformation

Sesudah	Sebelum
Aga love hate sih sama produk ini	aga love hate sih sama produk ini

Pada tabel 2 yaitu menunjukkan hasil transformation dimana data tersebut sebelum di transformation terdapat huruf kapital, Lalu

setelah melewati tahapan *transformation* semua huruf kapital diubah menjadi huruf kecil.

Tabel.3 Hasil Tokenizing

Sebelum	Sesudah
baru sebulan pakai pixy dewy cushion langsung jatuh cinta	['baru', 'sebulan', 'pakai', 'pixy', 'dewy', 'cushion', 'langsung', 'jatuh', 'cinta']

Pada table 3 yaitu menunjukkan hasil *tokenizing* dimana pada tahap ini dilakukan tokenisi untuk memecah teks atau memisahkan per-kata atau bisa disebut pemenggalan kalimat menjadi kata.

Tabel.4 Hasil Filtering

Sebelum	Sesudah
wajib di coba cushion yang ini	['wajib', 'coba', 'cushion', 'ini']

Pada Tabel 4 yaitu menunjukkan hasil filtering dimana pada tahap ini dilakukan penghapusan ataupun menaruh opsi kata simbol yang tidak dibutuhkan dalam proses berikutnya (analisis sentiment).

3.3. TF-IDF

Setelah melakukan *preprocessing* data (*transformation*, *tokenization*, dan *filtering*), maka selanjutnya melakukan proses term weighting yang nantinya akan diberikan bobot atau nilai. Perhitungan bobot tiap term dicari pada setiap data bertujuan untuk dapat mengetahui ketersediaan dan kemiripan suatu term di dalam data, semakin banyak kata yang muncul pada data, maka semakin tinggi bobot atau nilai term tersebut. Dalam pembobotan kata metode yang digunakan adalah metode TF-IDF. Untuk mendapatkan nilai TF-IDF dengan cara kalikan bobot tiap kata dari seluruh dokumen dengan nilai idf kata tersebut

Tabel.5 Tabel Hasil TF-IDF

Kata	Positif	Negatif	Netral
Cushion	1	0	0
Suka	1	0	0
Pakai	1	0	0
Muncul	1	0	0
Bisa	1	0	0
	5	0	0

3.4. Klasifikasi (SVM)

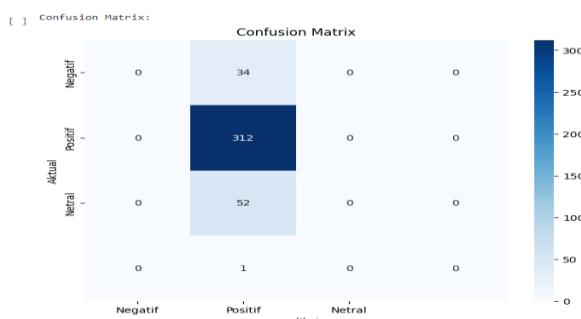
Setelah melakukan proses TF-IDF selanjutnya melakukan pengujian metode SVM pada bahasa pemrograman python, lalu membagi data training dan testing dengan perbandingan 80:20. Setelah didapatkan data training dan data testing untuk pengujian model tersebut dapat dilihat pada gambar berikut :

precision recall f1-score support				
0	0.00	0.00	0.00	34
1	0.78	1.00	0.88	312
2	0.00	0.00	0.00	52
11	0.00	0.00	0.00	1
accuracy			0.78	399
macro avg		0.20	0.25	399
weighted avg		0.61	0.78	399

Gambar.3 Hasil Klasifikasi Menggunakan Python

3.5. Confusion Matrix

Setelah dilakukannya pengujian metode SVM kemudian dilakukan pengujian menggunakan *confusion matrix* melalui library SVM. Pada tahap ini hasil akan didapatkan setelah melakukan pengujian dari keseluruhan data komentar menggunakan bahasa pemrograman *python*. Hasil dapat dilihat pada gambar berikut :



Gambar.4 Confusion Matrix

3.6. Visualisasi

Visualisasi adalah suatu rekayasa dalam pembuatan gambar, diagram atau animasi untuk penampilan suatu informasi. Visualisasi digambarkan dengan word cloud pada sentimen keseluruhan dokumen negatif positif dan netral tentang produk cushion merek pixy pada website female daily. Hasil dari Visualisasi ditampilkan pada gambar dibawah ini :



Gambar.5 Hasil Visualisasi.

4. PENUTUP

4.1. Kesimpulan

Dari hasil penelitian mengenai analisis sentimen produk cushion merek pixy pada website female daily yang berjumlah 2044 data dan sudah melewati tahap preprocessing text yang terdiri dari filtering, tokenization, transformation dan klasifikasi menggunakan algoritma support vector machine serta evaluasi data dengan confusion matrix menggunakan tools Google Collab dapat disimpulkan bahwa hasil dari tanggapan masyarakat terhadap cushion merek pixy pada website female daily tergolong positif dengan hasil persentase 78% pada akurasi (accuracy), 90% pada nilai Precision, dan tingkat keberhasilan (Recall) 100%. Nilai ini menunjukan bahwa klasifikasi algoritma support vector machine dinilai cukup baik dalam pemerosesan ini, dikarenakan hasil persentase akurasi sebesar 78%. Berdasarkan nilai tersebut membuktikan sentimen pada produk cushion merek pixy pada website female daily tergolong positif

4.2. Saran

Saran yang dapat diberikan penulis untuk penelitian analisis sentimen ini yaitu menambahkan jumlah dataset agar menghasilkan hasil yang lebih maksimal lagi. Serta menggunakan algoritma lainnya dalam menganalisis sentimen seperti Algoritma K-Nearest Neighbor dan Naïve bayes.

5. DAFTAR PUSTAKA

- Classifier, S V M, Nadifa Fadila Putri, Said Al Faraby, and Mahendra Dwifebri. 2021. “1 , 2 , 3.” 8(5): 10068–79.
- Indrayuni, Elly. 2019. “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes.” *Jurnal Khatulistiwa Informatika* 7(1): 29–36.
- Monika, Indri Parapat, and Muhammad Tanzil Furqon. 2018. “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak.” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 2(10): 3165–66. <http://j-ptiik.ub.ac.id>.
- Nurjannah, Musfiroh, and Inda Fitri Astuti. 2013. “PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman.” *Jurnal Informatika Mulawarman* 8(3): 110–13.
- Saputro, Irkham Widhi, and Bety Wulan Sari. 2020. “Uji Performa Algoritma Naïve Bayes Untuk Prediksi Masa Studi Mahasiswa.” *Creative Information Technology Journal* 6(1): 1.
- Setyohadi, Djoko Budiyanto, Felix Ade Kristiawan, and Ernawati Ernawati. 2017. “Perbaikan Performansi Klasifikasi Dengan Preprocessing Iterative Partitioning Filter Algorithm.” *Telematika* 14(01): 12–20.
- Sunoto, Yonathan, and Budi Wasito. 2014. “Analisis Testimonial Wisatawan Menggunakan Text Mining Dengan Metode Naive Bayes Dan Decision Tree , Studi Kasus Pada Hotel – Hotel Di Jakarta.” *Jurnal Informatika dan Bisnis ANALISIS* 3(2): 39–49.