

Analisis Sentimen Pada *Review Skincare Female Daily* Menggunakan Metode *Support Vector Machine (SVM)*

Riszki Wijayatun Pratiwi ^{#1}, Sharfina Febbi H^{#2}, Dairoh^{#3}, Dwi Intan Af'idah^{#4}, Qirani Rifa A^{#5}, Amaliyah Gian F ^{#6}

^{#1,2,3,4,5,6} Prodi Sarjana Terapan Teknik Informatika, Politeknik Harapan Bersama
Jalan Mataram No 9 Pesurungan Lor, Margadana Kota Tegal 52147

¹riszkipratiwi@poltektegal.ac.id, ²sharfina.handayani@poltektegal.ac.id, ³dairoh@poltektegal.ac.id,
⁴dwiintanafidah@poltektegal.ac.id, ⁵qiranileo@gmail.com, ⁶amaliyahgian.12@gmail.com

accepted on 29-11-2021

Abstract

Penampilan menarik merupakan hal yang diinginkan oleh sebagian perempuan, terutama perawatan kulit. Saat ini tingkat kepedulian merawat kulit sudah tinggi pada perempuan. Sebelum melaksanakan perawatan kulit hal yang utama yaitu mengetahui jenis kulit, untuk membantu perempuan agar lebih mudah mengetahui jenis produk skincare yang cocok pada kulit sehingga penelitian ini membahas mengenai analisis sentimen *review skincare* pada *website* kecantikan *female daily*. Metode yang digunakan pada penelitian ini adalah *Support Vector Machine (SVM)*. SVM merupakan *machine learning* yang sudah banyak digunakan pada penelitian-penelitian sebelumnya dan memberikan hasil yang maksimal. Penelitian ini mendapatkan hasil akurasi sebanyak 87% dengan *recall* sebesar 90%, *precision* sebesar 84,90%, dan *f1 score* sebesar 87,37%.

Keywords: *skincare*, analisis sentimen, SVM

I. INTRODUCTION

Mempunyai penampilan yang bersih dan mempesona dan wajah yang enak dipandang adalah hal yang diidam-idamkan kaum hawa. Akan tetapi untuk mendapatkan penampilan yang mempesona tidak hanya dilihat dari pakaian atau aksesoris yang dipakai saja, melainkan ada penunjang lainnya yakni kulit yang sehat. Kulit merupakan salah satu bagian yang penting untuk dirawat. Saat ini tingkat kepedulian merawat kulit pada perempuan telah berkembang, adapun langkah-langkah untuk melakukan perawatan kulit yang dilakukan sesuai kebutuhan. Saat ini banyak perempuan yang mempelajari dan memahami pengetahuan terkait jenis kulit

sebelum melaksanakan perawatan kulit. Oleh karena itu pada penelitian ini penulis ingin mengkaji analisis sentimen *review skincare* pada website kecantikan *female daily*.

Analisis sentimen merupakan sebuah prosedur untuk menganalisa pendapat seseorang terhadap suatu isu atau tema tertentu [1]. Secara umum analisis sentimen dikelompokkan menjadi dua tipe yaitu, analisis sentimen berbasis polaritas dan analisis sentimen berbasis emosi [2]. Analisis sentimen berbasis polaritas melakukan pengambilan informasi pada *review* berdasarkan kelasnya. Yakni, positif negatif dan netral. Jika Analisis sentimen berbasis emosi pengambilan informasi pada *review* didasarkan pada emosi dasar manusia [3].

Penelitian terkait analisis sentimen pada bahasa Indonesia sudah dilakukan, metode yang digunakannya juga bermacam- macam. Yakni *naïve bayes*, KNN dan *Support Vector Machine* (SVM) yang digunakan untuk menemukan model atau fitur yang sesuai dengan target. Pada penelitian ini metode yang digunakan adalah SVM. SVM merupakan metode pembelajaran yang meninjau data serta mengidentifikasi pola yang digunakan untuk klasifikasi. SVM merupakan salah satu metode yang baik untuk memperoleh hasil akurasi yang optimal karena SVM memproses data dengan efektif dan efisien [4].

II. LITERATUR REVIEW

A. Penelitian Sebelumnya

Penelitian mengenai analisis sentimen ini menggunakan referensi dari penelitian sebelumnya. Penelitian [5] membahas mengenai analisis sentimen pada *review* pelanggan toko online JD.ID menggunakan metode *naïve bayes*, pada penelitian ini memperoleh akurasi sebesar 96.44%. Sedangkan penelitian [6] membahas mengenai analisis sentimen pada toko online juga akan tetapi pada penelitian ini terdapat algoritma penanganan negasi. Algoritma yang digunakan untuk penanganan negasinya adalah *First Sentiment Word* (FSW) dan *Fixed Window Length* (FWL) sedangkan metodenya adalah *naïve bayes*, data yang digunakan pada penelitian ini merupakan *review* film yang berasal dari twitter. Hasilnya terjadi kenaikan akurasi ketika ditambahkan algoritma FSW dan FWL dibandingkan hanya *Naïve Bayes* saja.

Penelitian Analisis sentimen yang menggunakan algoritma SVM juga dilakukan oleh [7] penelitian ini membahas mengenai analisis sentiment masyarakat terhadap tindakan vaksinasi dalam upaya mengatasi pandemi covid-19. Pada penelitian ini membandingkan 2 metode yakni *naïve bayes* dan SVM dengan hasil akurasi *naïve bayes* 85.59% dan akurasi SVM 84.41%. Adapun penelitian Isnin membahas mengenai klasifikasi berita *hoax* covid-19 mendapatkan hasil akurasi dari metode SVM yang cukup bagus yakni 97.06%. berdasarkan literature rreview pada penelitian sebelumnya maka penelitian ini menggunakan metode SVM.

B. Analisis Sentimen

Analisis sentimen merupakan bidang studi yang mempelajari terkait opini, sentiment, evaluasi, sikap dan emosi *reviewer* yang ditumpahkan kedalam bahasa tulisan. Adapun cakupan pada analisis sentiment yaitu *emotion analysis*, *opinion extraction*, *sentiment mining*, *subjectivity analysis*, *affected analysis*, *emotion analysis*, *review mining* [1].

Tujuan dari analisis sentimen yaitu cara agar komputer bisa mengenali tingkat kelas emosi positif atau negatif dari sampel teks. Terdapat polaritas sentiment fungsinya untuk mengenali ciri khusus dri teks tersebut. Sebuah teks secara umum diklasifikasikan ke polaritas sentiment positif dan sentiment negatif. Terdapat juga polaritas campuran yakni kisaran dimana suatu dokumen dapat berisi pernyataan [8].

C. Text Mining

Text mining adalah proses untuk mendapatkan informasi dari user dengan cara berinteraksi sekumpulan dokumen dari waktu ke waktu dengan menggunakan alat-alat untuk menganalisa. Text mining merupakan sebuah perkembangan dari ilmu komputer untuk menangani masalah mengenai *Natural Language Processing*. Text mining mengadopsi bermacam metode dan teknik dari beberapa cakupan yakni *Information Retrieval*, *Information Extraction*, dan komputasi linguistik berbasis korpus [9].

D. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode yang sering digunakan pada klasifikasi data berupa text dengan tingkat akurasi yang lebih baik [10]. SVM adalah salah satu metode pada *supervised learning* yang digunakan untuk proses klasifikasi. Pada model klasifikasi, SVM mempunyai konsep yang lebih kuat dan lebih jelas secara matematis. SVM digunakan untuk mencari hyperplane optimal dengan cara memaksimalkan jarak antar kelas. *Hyperplane* adalah fungsi yang bisa dipakai untuk memisahkan antar kelas. Pada 2-D fungsi yang dipakai untuk klasifikasi antar kelas disebut *line whereas*, fungsi yang digunakan untuk klasifikasi antar kelas pada 3-D disebut *plane similarity*, sedangkan fungsi yang dipakai dalam klasifikasi ruang kelas adalah dimensi yang lebih tinggi dinamakan *hyperplane*. Berdasarkan penelitian [11] rumus SVM yang sering dipakai adalah sebagai berikut.

1) Polynomial Kernel

Polynomial Kernel merupakan proses yang harus dilakukan untuk mengklasifikasikan dataset training yang sudah normal. Rumus ini dapat di implementasikan menggunakan rumus persamaan 1 sebagai berikut :

$$K(\vec{X}_i, \vec{X}_j) = (\vec{X}_i \cdot \vec{X}_j + 1)^d \quad (1)$$

2) Radial Bias Function (RBF)

RBF merupakan salah satu rumus fungsi dalam SVM yang digunakan untuk klasifikasi dataset yang tidak terpisah secara linier. Pada proses ini RBF mempunyai kelebihan tingkat akurasi *training* dan *prediction* yang sangat baik. Rumus ini dapat di implementasikan menggunakan persamaan 2 berikut :

$$K(\vec{X}_i, \vec{X}_j) = \exp(-\gamma \|\vec{X}_i - \vec{X}_j\|^2) \quad (2)$$

3) Sigmoid Kernel

Sigmoid merupakan proses pengembangan dari jaringan saraf tiruan. Rumus ini dapat diimplementasikan menggunakan persamaan 3 berikut :

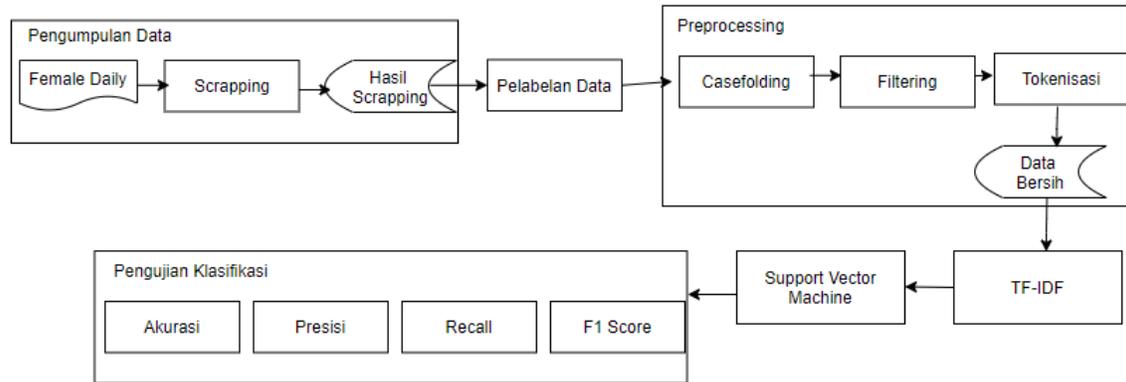
$$K(\vec{X}_i, \vec{X}_j) = \tanh(\alpha \vec{x}_i \cdot \vec{x}_j + \beta) \quad (3)$$

Keterangan :

- K = Kernel
- X & Y = Vector input space
- d = quadratic
- α = scalar parameter

III. RESEARCH METHOD

Data yang digunakan pada penelitian ini sebanyak 1.249 review dari website female daily melalui teknik scrapping. Sebanyak 649 review YA dan 600 review TIDAK. Kemudian data tersebut dibagi menjadi data training dan data testing. Detail proses analisis sentimen dapat dilihat pada Gambar 1.



Gambar. 1. Proses Analisis Sentimen dengan Metode SVM

A. Preprocessing

Setelah data diberi label maka langkah selanjutnya preprocessing pada penelitian ini preprocessingnya menggunakan casefolding, filtering dan tokenisasi. Casefolding adalah perubahan semua karakter huruf pada review menjadi karakter huruf kecil semua, kemudian berikutnya filtering berfungsi untuk menghilangkan karakter khusus seperti \$,%,#, * dan lainnya serta filtering juga berfungsi menghilangkan kata yang tidak sesuai dengan hasil parsing seperti username @, hashtag, #, *Uniform Resource Locator* (URL) dan emoticon. Adapun simbol, tanda atau angka juga dihapus karena tidak memiliki banyak pengaruh penentuan label. Proses preprocessing yang terakhir adalah tokenisasi, tokenisasi berfungsi untuk membagi ulasan dari kalimat menjadi per kata. Pada proses ini dilakukan dengan melihat per spasi tersebut kalimat bisa dipisahkan menjadi per kata.

B. Term Frequency Inverse-Document Frequency (TF-IDF)

TF-IDF merupakan pengalihan kedua algoritma yakni *Term Frequency* dan *Inverse Document Frequency*. *Term Frequency* adalah banyaknya jumlah term yang muncul pada suatu dokumen, sedangkan *document frequency* adalah pengurangan term yang sering muncul diberbagai dokumen, dengan perhitungan dari kebalikan frekuensi dokumen yang mengandung suatu kata.

C. Klasifikasi menggunakan Metode SVM

Data yang sudah selesai proses preprocessing selanjutnya di split menjadi 2 bagian, yakni data training dan data testing, dengan prosentase 80% data training dan 20% data testing. Kemudian setelah pemisahan data training dan data testing selanjutnya pembobotan kata menggunakan TF-IDF dan setelah itu diklasifikasikan menggunakan metode SVM. Pervorma model akan diujikan menggunakan nilai akurasi, *presisi*, *recall* dan *f-measure*.

IV. RESULT AND DISCUSSION

A. Perhitungan menggunakan metode SVM

Data yang digunakan diperoleh dari scrapping website *female daily* berjumlah 1.249. Adapun data yang diperoleh dari website tersebut merupakan produk lokal kecantikan di Indonesia dan atribut yang digunakan dalam penelitian ini yaitu nama akun dan nama produk. Berdasarkan hasil eksperimen yang telah dilakukan

untuk memecahkan masalah pada klasifikasi review produk kecantikan mendapatkan kesimpulan bahwa hasil analisis sentimen menggunakan *Support Vector Machine* (SVM) menghasilkan akurasi sebesar 87% dengan data split training sebesar 80% dan data testing sebesar 20%. Berikut contoh data yang telah dilakukan *labeling* pada Tabel 1:

TABEL I
 REVIEW PADA WEBSITE FEMALE DAILY

NO	Review	Kelas Sentimen
1.	teksturnya gel gitu jadi cepat terserap ke dalam kulit ga ninggalin efek basah atau lengket dikipun komposisinya juga bagus banget jadi bisa dianggap sunscreen dan skincare sekaligus gitu, intinya bagus banget dan membuat kulitku mulus dan terhidrasi	Ya
2.	Suka banget tekstore serumnya ringan cepat meresap ke kulit jadi gak ngeganggu pake serum di siang hari, aromanya juga seger sih menurut aku, jadi tinggal tetesin aja ke jerawat besokannya udah kering, rekomended banget deh buat kulit yang gampang jerawat	Ya
3.	Agak kurang suka sih sama tekstur dan after pemakaiannya. bikin muka ngerasa dempul gitu whiteningnya juga ga ngefek di muka aku walaupun sebenarnya seneng karena spfnya 35 yang ngebuat cream ini bisa ngelindungin paparan sinar matahari	Tidak
4.	Pernah nyoba produk ini karena adanya ini di toko dekat rumah, cuma sekali pakai maaf banget langsung muncul jerawat sepertinya memang tidak cocok. kelebihannya sih gampang dibeli di toko mana aja dan ukurannya ekonomis	Tidak

B. Pengujian *Confusion Matrix*

Confusion matrix merupakan suatu metode yang biasanya digunakan untuk melakukan hasil perhitungan nilai tingkat akurasi pada data mining. Detail tabel confusion matri dapat dilihat pada Tabel II, sedangkan untuk hasil pengujian pada penelitian ini terdapat 4 parameter yang akan dihitung yaitu akurasi, *presisi*, *recall*, dan *F1 score* dapat dilihat pada Tabel III.

TABEL II
 CONFUSION MATRIKS

Predicted Values	Actual Values	
	1 (Positive)	0 (Negative)
1 (Positive)	TP (True Positive)	FP (False Positive)
0 (Negative)	FN (False Negative)	TN (True Negative)

TABEL III
HASIL PENGUJIAN SVM

Parameter	Hasil Pengujian
Data Training	80%
Data Testing	20%
Hasil Akurasi	87,00 %
Hasil <i>Recall</i>	90, 00%
Hasil <i>Precision</i>	84,90 %
Hasil <i>F1score</i>	87,37 %

V. CONCLUSION

Berdasarkan hasil analisis dan pengujian terhadap analisis sentiment pada *review* produk kecantikan menggunakan metode *Support Vector Machine* maka didapatkan hasil kesimpulan bahwa Metode *Support Vector Machine* dapat melakukan prediksi kelas sentimen pada *review* produk kecantikan sesuai analisa yang disiapkan. Kemudian analisis sentimen ini belum sepenuhnya relevan untuk memprediksi kelas sentimen yang sesuai terhadap pemberian kelas berdasarkan rating atau bintang pada *review* tersebut, dan analisis sentimen pada produk kecantikan menggunakan SVM menghasilkan nilai akurasi menggunakan dataset 80% data *training* dan 20% data *testing* sehingga mendapatkan akurasi sebesar 87% dengan recall sebesar 90%, *precision* sebesar 84,90%, dan *f1 score* sebesar 87,37%.

ACKNOWLEDGMENT

Program Studi Sarjana Terapan Teknik Informatika Politeknik Harapan Bersama Kota Tegal yang telah membantu dan memberikan kemudahan fasilitas sarana dan prasarana dalam pelaksanaan penelitian ini.

REFERENCES

- [1] B. Liu, *Sentiment analysis: mining opinions, sentiments, and emotions*. New York, NY: Cambridge University Press, 2015.
- [2] A. Amolik, N. Jivane, M. Bhandari, and D. M. Venkatesan, "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques.," *International Journal of Engineering and Technology*, vol. 7, no. 6, p. 7, 2016.
- [3] M. J. Power and T. Dalgleish, *Cognition and Emotion: From Order to Disorder*. London: Psychology Press.
- [4] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," vol. 5, no. 1, p. 9.
- [5] F. V. Sari and A. Wibowo, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," vol. 10, no. 2, p. 6, 2019.
- [6] A. D. Rahmawan, "Analisis Emosi Pada Twitter Review Film Dalam Bahasa Indonesia," Gadjah Mada University, Yogyakarta, 2018.
- [7] B. Laurensz and Eko Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *JNTETI*, vol. 10, no. 2, pp. 118–123, May 2021, doi: 10.22146/jnteti.v10i2.1421.
- [8] Y. Mejova, "Sentiment Analysis: An Overview," p. 34.
- [9] *Advanced Approaches in Analyzing Unstructured Data*. Inggris: Cambridge University Press, 2006.
- [10] O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *SJI*, vol. 3, no. 1, pp. 34–45, Jun. 2016.
- [11] B. Pamungkas and M. E. Purbaya, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," p. 11, 2021.