

Analisis Sentimen Ulasan Herborist Sistem Pengambilan Keputusan Menggunakan Klasifikasi Neighbor dan TF-ID

Genoveva Andrasthea¹, Heryn Febuariyanti²

^{1,2} Universitas Stikubank, Indonesia

e-mail: ¹genovev4.22.1@gmail.com, ²herynfeb@edu.unisbank.ac.id

Abstract

Penggunaan teknologi dalam analisis sentimen terhadap ulasan produk telah menjadi fokus penelitian yang signifikan dalam beberapa tahun terakhir. Penelitian ini bertujuan untuk menganalisis sentimen ulasan produk Herborist menggunakan pendekatan klasifikasi Neighbor dan TF-ID sebagai sistem pengambilan keputusan. Analisis opini konsumen untuk menghasilkan wawasan baru tentang persepsi pengguna terhadap produk skincare Herborist, memberikan kontribusi unik pada bidang analisis sentimen di industri kecantikan. Analisis sentimen menggunakan algoritma K-Neighbor dapat mengklasifikasikan opini berdasarkan kedekatan fitur dengan data terlatih untuk menghasilkan sentimen positif atau negatif. Metode ini memungkinkan untuk menggali pola sentimen dari ulasan konsumen secara efektif. Penelitian ini diawali dengan crawling data dan menghasilkan 1006 data yang akan di proses dalam perhitungan dengan Ms.excel lalu melanjutkan proses dengan menggunakan Bahasa pemrograman *Python* dengan aplikasi google collabs, dibantu dengan library sastrawi. Nilai K yang digunakan adalah 5. Penelitian ini mencakup pengukuran performa (akurasi, presisi, recall dan f-measure) metode KNN dengan berbagai macam nilai K pada objek 1006 data ulasan minyak zaitun herborist. Dapat disimpulkan nilai performa paling baik pada K=5, Dimana tingkat akurasi mencapai 96%, presisi 97%, recall 1.00, dan F-Measure sebesar 99%

Keywords: K-Nearest Neighbor, Klasifikasi, Teks Mining, Ulasan Produk, Analisis sentiment

Abstract

The use of technology in sentiment analysis of product reviews has been a significant research focus in recent years. This research aims to analyze the sentiment of Herborist product reviews using the Neighbor classification approach and TF-ID as a decision making system. Consumer opinion analysis to generate new insights into user perceptions of Herborist skincare products, providing a unique contribution to the field of sentiment analysis in the beauty industry. Sentiment analysis using the K-Neighbor algorithm can classify opinions based on the proximity of features to the trained data to produce positive or negative sentiment. This method makes it possible to mine sentiment patterns from consumer reviews effectively. This research began with data crawling and produced 1006 data which would be processed in calculations with Ms. Excel and then continued the process using the Python programming language with the Google Collabs application, assisted by the literary library. The K value used is 5. This research includes performance measurements (accuracy, precision, recall and f-measure) of the KNN method with various K values on 1006 herborist olive oil review data objects. It can be concluded that the best performance value is at K=5, where the accuracy level reaches 96%, precision 97%, recall 1.00, and F-Measure of 99%

Keywords: K-Nearest Neighbor, Classification, Text Mining, Product Reviews, Sentiment Analysis

1. Pendahuluan

Minyak zaitun dalam perkembangan jaman menjadi sangat terkenal karena berbagai manfaat dan khasiat yang ada di dalamnya, Herborist sebagai salah satu

merek terkemuka dalam industry perawatan kecantikan dan kesehatan, tidak hanya memiliki basis konsumen yang luas tetapi juga mendapat perhatian secara signifikan dalam bentuk ulasan online

Promosi (promotion) adalah usaha atau upaya untuk memajukan atau meningkatkan; misalnya untuk meningkatkan perdagangan atau memajukan bidang usaha. (Putri Fitrianti et al., 2019) Herborist sebagai salah satu merek perawatan kulit favorit perempuan Indonesia yang bernaung di bawah PT Victoria Care Indonesia Tbk (VICI) dan dikenal dengan produk-produk yang memanfaatkan kekayaan alam Indonesia. Kali ini, dengan ini penelitian ini dikerjakan dengan menggunakan analisis sentimen ulasan-ulasan dan komentar dari pengguna agar Herborist bisa dipilih oleh pengguna pengguna baru. (Amalia, 2021)

Analisis sentimen merupakan salah satu bagian dari studi text mining yang bias disebut studi komputasi untuk melakukan klasifikasi terhadap pendapat, emosi dan sikap seseorang terhadap entitas. (Leidiyana, 2013)

Penelitian ini menggunakan Klasifikasi Neighbor karena metode ini tangguh terhadap data noise. Menurut salah satu penelitian yaitu penelitian Puspita dan Widodo tahun 2020, Algoritma KNN ini merupakan algoritma yang sudah populer. (Nurjanah et al., 2017)

Tujuan dari penelitian ini adalah untuk menganalisis sentimen ulasan pengguna skincare Herborist menggunakan algoritma K-Nearest Neighbor (K-NN). Penelitian ini bertujuan mengklasifikasikan ulasan pengguna menjadi sentimen positif, negatif, atau netral. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi performa K-NN dalam pengolahan teks ulasan, mengidentifikasi faktor-faktor yang mempengaruhi sentimen pengguna, serta memberikan wawasan bagi perusahaan terkait persepsi konsumen terhadap produk skincare Herborist.

Pengklasifikasian nilai sentimen terbagi menjadi positif dan negatif dengan menggunakan algoritma KNN dengan pembobotan kata TF-IDF dan fungsi Cosine Similarity. Penelitian tersebut menghasilkan nilai akurasi terbesar adalah 67,5% ketika k=5, presisi tertinggi 56,94% ketika k=5, dan recall 78,24% dengan k=15. (Rahayu et al., 2022)

Penelitian yang dilakukan adalah mengumpulkan data pada *official akun* dari herborist Twitter adalah sebuah situs jejaring sosial yang sedang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari

komputer ataupun perangkat mobile mereka dari manapun dan kapanpun. Setelah diluncurkan pada Juli 2006, jumlah pengguna Twitter meningkat sangat pesat. Pada September 2010, diperkirakan jumlah pengguna Twitter yang terdaftar sekitar 160 juta pengguna. (Ernawati & Wati, 2018)

2. Metode Penelitian

2.1 Analisis Sentiment

Analisis sentimen merupakan salah satu bagian dari studi text mining yang bisa disebut studi komputasi untuk melakukan klasifikasi terhadap pendapat, emosi dan sikap seseorang terhadap entitas. (Pandegani, 2023).

2.2 Text Mining

Text mining merupakan bagian dari data mining yang merupakan proses untuk ekstraksi pengetahuan dan informasi dari pola pada sekumpulan dokumen berupa text dengan menggunakan metode analisis tertentu. (Sutrisno & Amini, 2023)

Text mining juga memiliki fungsi untuk mencari kata yang dapat mewakili apa saja yang terdapat pada dokumen sehingga dilakukan analisis keterhubungan pada dokumen. (Tanggu Mara et al., 2021) dalam hal ini text mining memiliki beberapa tahapan antara lain :

1. Data Cleaning

Tahapan ini dilakukan untuk membersihkan data yang dimana data teks akan diambil kolom yang relevan dengan topik dan dibersihkan, setelah data dibersihkan maka didapat 1006 data, hasil pembersihan data menggunakan excel disimpan dalam format CSV (comma separated values) dengan kolom komentar. (Ernawati & Wati, 2018)

2. Tokenizing

Tokenizing dilakukan untuk memecahkan sebuah kata dan melakukan penghapusan terhadap delimiter serta karakter angka bersama entitas tweet seperti halnya hastag, retweet dan mention. (Adzhan & Yusup, 2022)

3. Stopword

Stopword ini dilakukan untuk menghilangkan kata yang bisa disebut tidak memiliki arti. Sebuah daftar stopwords akan dibuat sebelum melakukan proses stopwords removal

4. Stemming

Tahapan Stemming ini merupakan sebuah proses yang dilakukan pemetaan dan penguraian dari berbagai varian dari suatu

kata menjadi bentuk kata dasar(Furqan et al., 2022)

2.4 Pembobotan TF-IDF

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode yang melakukan perhitungan pada tiap bobot kata yang terdapat didalam data dokumen.(Legito et al., 2023)

$$idf = \log \frac{N}{df}$$

$$w(k,d) = tf(k,d) * idf$$

$$w(k,d) = tf(k,d) * \log \frac{N}{df}$$

Dengan W(k,d) adalah bobot kata (term) yang tersedia dalam dokumen, tf (k,d) adalah jumlah frekuensi munculnya kata didalam dokumen, N adalah jumlah seluruh dokumen yang tersedia didalam database dan df adalah jumlah dokumen yang mengandung term.(Kosasih & Alberto, 2021)

2.5 K-Nearest Neighbor

K-Nearest Neighbor (K-NN) ialah metode sederhana mudah diimplementasikan, data yang digunakan memiliki label sehingga memudahkan dalam proses pengelompokkan ke dalam kelas yang paling sesuai dan memiliki keunggulan dapat mengklasifikasikan data dengan adanya data latih dan data uji serta memiliki kemudahan menerjemahkan hasil dan akurasi dari prediksi dengan secara akurat memilah terlebih dulu nilai k terdekat dengan tepat.(Yoga Pratama et al., 2021) K-NN akan dihitung memakai jarak Euclidean dengan persamaan sebagai berikut :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

$$= \sqrt{\sum (x_i - y_i)^2}$$

Dengan d(x, y) sebagai jarak antar variabel x1 dan y2, x sebagai variabel dan k sebagai jumlah variabel(Ramadhan, 2019)

3. Hasil dan Pembahasan

Hasil dan pembahasan yang didapatkan terdiri dari hasil tahapan pre-processing,

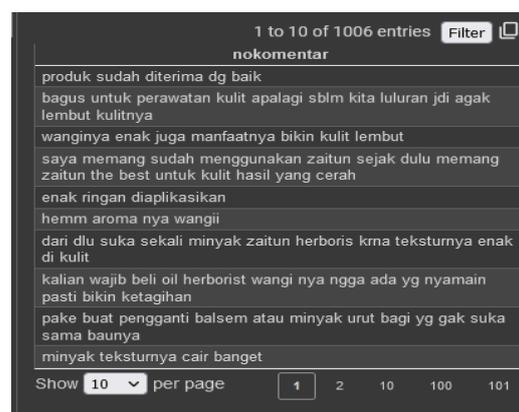
perhitungan TF-IDF, dan hasil dari klasifikasi K-Nearest Neighbor.

3.1 Pre=processing

Data yang sudah diambil tersebut akan diolah menggunakan excel untuk pertama kali untuk memudahkan pada saat proses pengolahan menggunakan Google collabs.

3.1.1 Data Cleaning

Tahapan ini adalah tahapan yang akan membersihkan data mentah dan menghasilkan data bersih ditunjukkan pada gambar 3.1



Gambar 3.1 Hasil cleaning data

1.Case folding

Pada tahap ini huruf akan dibuat seragam ke bentuk huruf kecil karena proses dilakukan oleh komputer maka case sensitive dengan huruf besar kecil akan berpengaruh pada proses ini seluruh huruf akan dibuat low case. Ditunjukkan pada tabel 3.1

Tabel 3.1 Case folding

INPUT	OUTPUT
Wangi minyak zaitunnya sangat khas .. Cocok untuk pijat & urut :)	wangi minyak zaitunnya sangat khas..cocok untuk & pijat urut :)

2. Normalisasi

Pada tahap normalisasi segala bentuk karakter khusus angka dan simbol dihilangkan dengan menyisakan kata yang dibutuhkan untuk dilanjutkan pada proses tokenizing. Ditunjukkan tabel 3.2

Tabel 3.2 Normalisasi

INPUT	OUTPUT
Wangi minyak	wangi
zaitunnya	minyak
sangat khas ..	zaitunnya
Cocok untuk	sangat khas
pijat &urut :)	cocok untuk
	pijat urut

3.1.2 Tokenizing

Proses tokenizing dilakukan dengan tujuan memecah kalimat menjadi komponen kata terpisah satu sama lain. Sebagai contoh karakter whitespace, seperti enter, tabulasi, spasi dianggap sebagai pemisah kata. Namun untuk karakter petik tunggal ('), titik (.), semikolon (;), titik dua (:) atau lainnya ditunjukkan pada tabel 3.3

Tabel 3.3 Tokenizing

INPUT	OUTPUT
wangi minyak	"wangi"
zaitunnya	"minyak"
sangat khas	"zaitunnya"
cocok untuk	"sangat"
pijat urut	"khas"
	"cocok"
	"untuk" "pijat"
	"urut"

3.1.3 Stopword Removal

Pada tahap ini dapat dilakukan dengan cara membuang kata yang kurang penting (stopword) atau dapat menyimpan kata penting (wordlist). Contoh stopwords adalah "yang", "dan", "di", "dari", dan lain sebagainya, ditunjukkan pada tabel 3.4

Tabel 3.4 Stopword

INPUT	OUTPUT
wangi	"wangi"
minyak	"minyak"
zaitunnya	"zaitun"
sangat khas	"khas"
cocok untuk	"cocok"
pijat urut	"pijat" "urut"

3.1.4 Stemming

Stemming merupakan proses untuk mendapatkan root/stem atau kata dasar dari satu kata yang memisahkan masing-masing kata dari bahasa dasar dan

imbuhan nya baik awalan (prefiks) maupun akhiran (sufiks). Ditunjukkan pada tabel 3.5

Tabel 3.5 Stemming

INPUT	OUTPUT
wangi	"wangi"
minyak	"minyak"
zaitun khas	"zaitun" "asli"
cocok pijat	"pas" "pijat"
urut	"urut"

3.2. TF-IDF (Term Frequency-Inverse Dense Frequency)

Tahap ini akan berlangsung setelah text-processing diselesaikan, pada tahap ini akan diberlakukan hitung bobot kata dengan pendekatan TF-IDF. Dapat dilihat di coding berikut ditampilkan program untuk menerapkan TF-IDF pada teks review seperti pada gambar 3.2 dibawah ini

```
# Terapkan TF-IDF pada teks review
matrix_tfidf = tfidf_vectorizer.fit_transform(teks_review)
```

Gambar 3.2 Proses TF-IDF

Dengan proses diatas maka akan dihasilkan seperti pada gambar 3.3 dibawah ini:

```
Dokumen ke-323:
Kata: khasiatnya, Skor TF-IDF: 0.6684
Kata: bikin, Skor TF-IDF: 0.4301
Kata: lembut, Skor TF-IDF: 0.4184
Kata: kulit, Skor TF-IDF: 0.3097
Kata: bagus, Skor TF-IDF: 0.3120

Dokumen ke-324:
Kata: okeeee, Skor TF-IDF: 0.5153
Kata: syukaa, Skor TF-IDF: 0.5459
Kata: oke, Skor TF-IDF: 0.3888
Kata: banget, Skor TF-IDF: 0.2161
Kata: sekali, Skor TF-IDF: 0.3758
Kata: enak, Skor TF-IDF: 0.2196
Kata: wanginya, Skor TF-IDF: 0.2216

Dokumen ke-325:
Kata: banyak, Skor TF-IDF: 0.5728
Kata: melembabkan, Skor TF-IDF: 0.4735
Kata: manfaatnya, Skor TF-IDF: 0.5804
Kata: kulit, Skor TF-IDF: 0.3328
```

Gambar 3.3 Hasil TF-IDF

Sedangkan untuk hasil presentase TF-IDF bisa dilihat pada gambar 3.4 dibawah ini

```
Dokumen ke-1006:
Kata: bagussss, Skor TF-IDF: 1.0000

Dokumen ke-1007:
Kata: suka, Skor TF-IDF: 0.7342
Kata: wanginya, Skor TF-IDF: 0.6789

Jumlah total kata seluruhnya: 5818
```

Gambar 3.4 Presentase TF-IDF

- S., & Erkamim, M. (2023). Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 324–330. <https://doi.org/10.57152/malcom.v3i2.893>
- Leidiyana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, 1(1), 65–76.
- Nurjanah, W. E., Setya Perdana, R., & Fauzi, M. A. (2017). Analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial twitter menggunakan metode k-kearest neighbor dan pembobotan jumlah retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1750–1757. <http://j-ptiik.ub.ac.id>
- Pandegani, D. (2023). *Analisis Sentimen Algoritma K-Nearst Neighbor (Knn) Pada Ulasan Google Play Aplikasi Tokopedia*. [http://repository.unas.ac.id/id/eprint/8658%0Ahttp://repository.unas.ac.id/8658/2/BAB II.pdf](http://repository.unas.ac.id/id/eprint/8658%0Ahttp://repository.unas.ac.id/8658/2/BAB%20II.pdf)
- Putri Fitrianti, R., Kurniawati, A., Agustien, D., Sistem Informasi, J., & Ilmu Komputer dan Teknologi Informasi, F. (2019). A-27 Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 1907–5022.
- Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Ramadhan, F. W. (2019). *Analisis Sentimen Komentar Warganet Terhadap Layanan It Pada Bank Mandiri Menggunakan Algoritma K-Nearest Neighbor (K-Nn) Berdasarkan Kriteria Itsm*.
- Sutrisno, E. P., & Amini, S. (2023). Implementasi Algoritma K-Nearest Neighbor Pada Analisis Sentimen Ulasan Pengguna Aplikasi Digital Korlantas Polri. *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2(2), 687–695.
- Tangu Mar, A., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. *Jointer - Journal of Informatics Engineering*, 2(01), 24–31. <https://doi.org/10.53682/jointer.v2i01.30>
- Yoga Pratama, A., Umidah, Y., Voutama, A., Informatika, T., Ilmu Komputer, F., Singaperbangsa Karawang Ds Paseurjaya, U., Telukjambe Timur, K., Karawang, K., & Barat, J. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 897–910.