

Analisis Sentimen Kepuasan Pelanggan Parfum Scentplus dan Moris di Platform Tik Tok menggunakan Metode Regresi Logistik

Rivaldi Alwi¹, Adam Arif Budiman^{2*}

^{1,2} Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Darma Persada
^{1,2} Jl Taman Malaka Selatan No 8, Jakarta Timur 13450, Indonesia
Email : ariadam@gmail.com

Abstrak — Parfum Scentplus dan Moris, dua merek parfum lokal yang tengah meroket ketenaran nya melalui platform Tik Tok, menjadi fokus penelitian ini. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap kepuasan pelanggan terhadap kedua merek parfum ini. Metode Regresi Logistik digunakan sebagai alat utama untuk merinci dan mengeksplorasi sentimen yang terkandung dalam komentar-komentar pelanggan, Pengumpulan data dilakukan dengan memanfaatkan teknik Scraping untuk mengakses dan mengumpulkan komentar-komentar pelanggan dari platform Tik Tok, Data latih yang digunakan sebanyak 1000 data yang telah dilabel dengan masing-masing data sentimen positif 420, sentimen netral 149, dan sentimen negatif 431 untuk parfum moris dan data sentimen positif 456, sentimen netral 146, dan sentimen negatif 398 untuk parfum scentplus yang dilatih menggunakan algoritma regresi logistik, Pada penelitian ini menunjukkan model untuk parfum Moris memiliki performa terbaik dengan akurasi sebesar 93%, presisi sebesar 93%, dan recall sebesar 93%. Sedangkan model untuk parfum Scentplus memiliki akurasi sebesar 91%, presisi sebesar 91%, dan recall sebesar 91%.

Kata kunci – Analisis Sentimen, Moris, Regresi Logistik, Scentplus, TikTok

Copyright © 2024 JURNAL TIFDA
All rights reserved.

I. PENDAHULUAN

Pada era digital seperti saat ini, platform media sosial seperti TikTok telah menjadi salah satu media utama bagi masyarakat untuk berbagi pendapat, pengalaman, dan ulasan tentang berbagai produk, Salah satu produk yang sering dibahas dan direview oleh pengguna TikTok adalah parfum, Kepuasan pengguna TikTok terhadap parfum ini memiliki dampak positif yang signifikan bagi perusahaan Dengan berhasil menarik banyak pelanggan dan mencapai kepuasan pelanggan.

Kepuasan pelanggan ini menciptakan komunitas pelanggan yang setia dan terus berkembang [1], Sementara pelanggan yang tidak puas akan menyampaikan pengalaman buruknya kepada orang lain dan dapat dibayangkan betapa besarnya kerugian yang dialami perusahaan dari kegagalan memuaskan pelanggan. Oleh sebab itu, perusahaan wajib merencanakan strategi yang efektif untuk mengukur dan menganalisis kepuasan pelanggan [2], Salah satu strategi yang dapat dilakukan adalah melalui analisis sentiment.

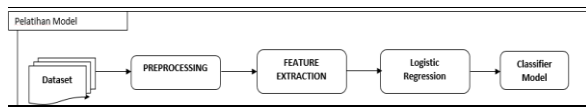
Analisis sentimen merupakan metode untuk mengekstraksi dan menganalisis pendapat, opini, atau perasaan yang terkandung dalam teks atau konten yang diposting oleh pengguna media sosial, Analisis sentimen mengklasifikasikan teks sebagai opini yang positif, negatif, atau netral [3]. Salah satu algoritma yang paling tinggi akurasi nya untuk melakukan analisis sentimen adalah regresi logistik pernyataan ini dapat di lihat pada hasil penelitian yang dilakukan oleh Mona Cindo dkk [4].

Regresi logistik merupakan salah satu model regresi yang dapat digunakan untuk menjelaskan hubungan antara variabel respon yang terdiri dari dua kategori dengan variabel prediktor yang dapat berupa variabel kontinu, diskrit, atau campuran keduanya. Algoritma ini sangat efektif dalam memprediksi kemungkinan hasil dari kategori tertentu berdasarkan data yang diberikan [5].

II. METODOLOGI

Dalam Penelitian ini metode regresi logistik akan diterapkan untuk mengklasifikasikan sentimen dari ulasan pelanggan menjadi positif, negatif, atau netral,

dan untuk memahami faktor-faktor apa saja yang mempengaruhi tingkat kepuasan pelanggan. Berikut merupakan langkah-langkah untuk menerapkan algoritma regresi logistik untuk analisis sentiment.



Gambar 1. Pelatihan Model

Tahapan yang dilakukan pada proses preprocessing diantaranya.

1. *Cleaning Text*, dimana tahap ini merupakan pembersihan atau cleansing yang mencakup Langkah-langkah untuk menghilangkan elemen-elemen yang tidak diinginkan atau tidak relevan dari teks, seperti tag HTML, URL, atau kode yang tidak terbaca.
2. Tahapan kedua yaitu case folding dimana tahap ini adalah konversi dari seluruh teks menjadi huruf kecil.
3. Stopword, dimana tahap ini melibatkan penghapusan kata umum yang sering tidak memiliki nilai semantic dalam analisis teks, seperti “dan”, “atau”, “di”, dan sebagainya, menghapus stopwords membantu mengurangi ukuran vector dan fokus pada kata-kata penting.
4. Stemming, dimana tahap ini melibatkan proses mereduksi kata-kata ke bentuk dasarnya dengan membantu mengurangi ukuran vector dan fokus pada kata-kata penting.
5. Tokenzing, dimana tahap ini melibatkan proses memecah teks menjadi unit yang lebih kecil, yang disebut toke. Token ini dapat berupa kata, frasa, atau kalimat.
6. Setelah melalui proses preprocessing kalimat yang sudah dipisahkan menjadi kata akan melalui proses feature extraction. Pada proses ini kata akan diboboti dengan metode tf-idf, data akan dibagi menjadi dua bagian, yaitu 80% data training dan 20% data testing. Proses terakhir pada training data adalah melakukan prediksi dari kata yang sudah diberi bobot dengan algoritma regresi logistik

III. LANDASAN TEORI

A. Data Mining

Data mining adalah sebuah proses pencarian informasi secara otomatis yang terdapat dalam tempat penyimpanan data berukuran besar, Teknik *data mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk mengidentifikasi pola yang tidak terlihat, hubungan, atau informasi yang berguna lainnya yang dapat membantu dalam pengambilan keputusan yang lebih baik *Metode data mining* melibatkan berbagai teknik statistik, matematika, dan kecerdasan buatan untuk menganalisis data secara mendalam dan mengekstraksi wawasan yang berguna [6].

Tahapan *data mining* meliputi penerapan algoritma untuk klasifikasi. Sebelum menerapkan algoritma, Dataset harus di bagi menjadi dua bagian, Data *training* dan data *testing*. Data *training* digunakan untuk melatih model, sementara data *testing* digunakan untuk menguji keakuratan model tersebut [7].

B. Scraping Data

Teknik *web scraping* merupakan cara untuk mengambil data dalam jumlah besar dari situs *website*. Data yang diekstraksi kemudian disimpan dalam file lokal di komputer atau database dalam format tabel, mirip dengan spreadsheet. Setelah itu, data dapat diproses menggunakan alat seperti *RapidMiner* untuk melakukan pra-pemrosesan atau pra-pengolahan data sebelum memasuki tahap berikutnya [8].

Pada penelitian ini *scraping* data yang di gunakan memakai tools yang sudah di sediakan oleh situs *website apify* untuk mengambil data ulasan atau komentar pelanggan *Tik Tok scentplus official* dan *moris Indonesia*.

C. Jupyter Notebook

Jupyter notebook merupakan *tools open-source* untuk menyusun code – code *Python* dengan kelebihan *user friendly* dan mudah digunakan, Dengan *Jupyter notebook* dapat membuat dan berbagi dokumen yang menggabungkan *live code*, *visualisasi*, narasi teks, dan elemen-elemen lainnya. Dokumen-dokumen ini dapat berisi kode dalam berbagai bahasa pemrograman, seperti *Python*, *R*, *Julia*, dan lainnya, Selain itu *jupyter notebook* dapat mendokumentasikan sebuah pekerjaan, dimana coding dan dokumentasi bisa dilakukan dalam satu page dan disimpan dalam bentuk presentasi yang menarik [9].

Dalam penelitian ini, *Jupyter notebook* digunakan untuk melatih dan membangun model menggunakan metode regresi logistik. Dengan memanfaatkan *jupyter notebook*, proses pengembangan model menjadi lebih interaktif dan efisien, Selain itu penggunaan *jupyter notebook* memungkinkan untuk mempermudah melakukan eksplorasi data dan visualisasi hasil..

D. Confusion Matrix

Confusion Matrix merupakan perhitungan yang menggambarkan performa sebuah classifier model dengan cara membandingkan hasil prediksi terhadap nilai label sebenarnya dari dataset yang dijadikan data uji [10]. Dalam perhitungan ini, prediksi kelas positif yang benar disebut True Positive (TP), sementara yang salah disebut False Positive (FP). Sedangkan, untuk kelas negatif, prediksi yang benar disebut True Negative (TN) dan yang salah disebut False Negative (FN). Confusion Matrix memberikan gambaran yang jelas tentang seberapa baik model tersebut mengklasifikasikan data, serta membantu dalam mengevaluasi kekuatan dan kelemahan model tersebut.

E. Regresi Logistik

Regresi logistik adalah salah satu jenis *regresi* yang sangat berguna dalam mengatasi masalah klasifikasi, *Regresi logistik* memungkinkan untuk memahami dan memodelkan hubungan antara *variabel input* dan *output* dengan cara yang mudah dipahami [11].

Dalam penelitian ini rumus regresi logistik multinomial akan diterapkan karena rumus regresi logistik multinomial merupakan generalisasi dari regresi logistik biner untuk masalah multikelas. Dalam regresi logistik biner, hanya ada dua kemungkinan hasil, seperti "Positif" atau "Negatif". Dalam regresi logistik multinomial, terdapat lebih dari dua kemungkinan hasil, seperti "Positif", "Negatif", atau "Netral", Cara menghitung nya adalah dengan menghitung probabilitas masing-masing kategori (positif, negatif, dan netral) berdasarkan frekuensi kemunculan kata-kata, Untuk setiap kategori, nilai probabilitas ditentukan dengan mengalikan koefisien regresi (misalnya, $B_1\text{positif}$, $B_1\text{negatif}$, $B_1\text{netral}$) dengan frekuensi kemunculan kata-kata positif, negatif, atau netral, persamaan 1 adalah rumus regresi logistik multinomial yang digunakan di penelitian ini.

$$g_j(x) = \beta_j0 + \beta_j1x_1 + \beta_j2x_2 + \dots + \beta_jpx_p \quad (1)$$

F. Use Case Diagram

Penggunaan *use case* merupakan hal yang sangat penting untuk memahami kebutuhan fungsional suatu sistem. Oleh karena itu, Penting bagi pengembang sistem untuk merancang *use case* sejak awal. Selain itu, versi *use case* yang lebih terperinci harus disusun sebelum memulai pengembangan sistem tersebut [12]. Pada penelitian ini, UML yang digunakan adalah *use case diagram* dan *activity diagram*. *Use case diagram* digunakan untuk menggambarkan interaksi antara aktor-aktor dengan sistem atau aplikasi yang sedang dikembangkan, sementara *activity diagram* digunakan untuk memodelkan alur kerja atau proses dalam aplikasi tersebut secara grafis. Kedua jenis diagram ini membantu dalam memvisualisasikan dan merancang sistem secara lebih terstruktur dan sistematis, memudahkan pengembangan dalam memahami serta mengkomunikasikan kebutuhan dan desain aplikasi kepada tim pengembangan dan pemangku kepentingan lainnya.

G. MySQL

MySQL merupakan salah satu database paling populer, karena hampir semua aplikasi berbasis *website* seperti *WordPress* memanfaatkan teknologi nya, Selain itu *MySQL* juga di tawarkan dalam berbagai versi termasuk versi gratis [13].

Pada penelitian ini, *MySQL* digunakan sebagai database untuk menyimpan hasil klasifikasi kalimat serta data akun admin pemasaran dan manajer. *MySQL* dipilih karena kehandalannya dalam penyimpanan dan

pengelolaan data, serta kemampuannya untuk menangani volume data yang besar dengan performa yang baik.

H. GitHub

GitHub merupakan platform multifungsi yang menggabungkan manajemen proyek, sistem versioning code, dan jaringan sosial bagi para developer di seluruh dunia. Dengan beragam fitur yang disediakan, GitHub mempermudah developer dalam mengelola, berbagi, dan mengembangkan proyek-proyek inovatif mereka. Melalui platform ini, kolaborasi menjadi lebih efisien dan inspirasi semakin mudah ditemukan [14].

GitHub umum nya di gunakan untuk menjadi repositori Open Source. Tujuan utamanya adalah untuk mempelajari proyek perangkat lunak sebelumnya agar dapat meningkatkan pengembangan proyek baru. Dalam beberapa tahun terakhir, GitHub telah menjadi sumber data yang kaya untuk penelitian, dengan banyak makalah yang diterbitkan berdasarkan analisis data dari platform ini, yang membantu komunitas memahami lebih dalam tentang rekayasa perangkat lunak [15].

Dalam penelitian ini, GitHub digunakan untuk menyimpan file-file analisis sentimen dan menghubungkan file-file tersebut ke Streamlit sebagai platform untuk melakukan deployment analisis sentimen. Dengan memanfaatkan GitHub, seluruh proses pengelolaan dan pengembangan proyek menjadi lebih terorganisir dan mudah diakses

I. Streamlit

Streamlit merupakan *library python* yang memungkinkan untuk membangun aplikasi dengan cepat menggunakan *Python* murni. Dengan *Streamlit*, Pengguna dapat membuat prototipe dan menguji aplikasi dengan mudah tanpa harus memiliki pengetahuan mendalam tentang alat-alat *front-end* untuk antarmuka pengguna. Hal ini merupakan solusi yang sederhana namun bermanfaat untuk menciptakan aplikasi dengan kecepatan dan efisiensi [16].

Pada penelitian ini, *Streamlit* digunakan untuk melakukan deployment model yang sudah jadi agar dapat digunakan oleh semua orang dengan mudah. *Streamlit* dapat mengubah model machine learning atau analisis data yang kompleks menjadi aplikasi web interaktif dengan cepat dan efisien.

IV. HASIL DAN PEMBAHASAN

A. ANALISIS SENTIMEN PERKALIMAT

Gambar 2 merupakan tampilan analisis sentimen perkalimat *admin pemasaran* dan *manajer* dalam situs *website* penelitian ini, Dengan antarmuka yang sudah disesuaikan, *Admin pemasaran* dan *manajer* dapat dengan mudah melakukan analisis sentimen perkalimat.

Masukkan kalimat untuk analisis sentimen:

Masukkan kalimat

motif botol moris keren keren banget

Analisis

Gambar 2. Halaman Analisis Sentimen Perkalimat

B. Analisis Sentimen profile

Gambar 3 merupakan tampilan analisis sentimen profile *admin pemasaran* dan *manajer* dalam situs *website* penelitian ini, Dengan antarmuka yang sudah disesuaikan, *admin pemasaran* dan *manajer* dapat dengan mudah melakukan analisis sentimen profile.

Unggah file untuk Prediksi Sentimen

Unggah file Excel

Drag and drop file here
Limit 200MB per file • XLSX

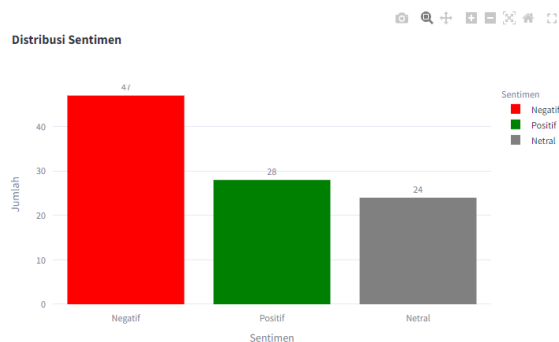
Browse files

data moris 100 no sentimen.xlsx 22.6MB

Gambar 3. Halaman Analisis Sentimen profile

C. Diagram Batang

Gambar 4 merupakan diagram batang hasil analisis sentimen yang dapat di lihat oleh *admin pemasaran* dan *manajer* dalam situs *website* penelitian ini, dengan antarmuka yang sudah disesuaikan *admin pemasaran* dan *manajer* dapat dengan mudah melihat hasil analisis sentimen dalam bentuk diagram batang.



Gambar 4. Diagram Batang

D. Word Cloud

Gambar 5 merupakan kata kata yang sering muncul atau word cloud yang dapat di lihat oleh *admin pemasaran* dan *manajer* dalam situs *website* penelitian ini, dengan antarmuka yang sudah disesuaikan *admin pemasaran* dan *manajer* dapat dengan mudah melihat kata kata yang sering muncul dalam bentuk *word cloud*.



Gambar 5. Word Cloud

V. KESIMPULAN

Penelitian ini memberikan beberapa kesimpulan diantaranya.

1. Model yang digunakan dapat secara otomatis mengklasifikasikan ulasan atau komentar tentang kedua produk parfum, yaitu *Moris* dan *Scentplus*, di platform *TikTok* ke dalam sentimen positif, netral, dan negatif.
2. Data latih yang digunakan sebanyak 1000 data yang telah dilabel dengan masing-masing data sentimen positif 420, sentimen netral 149, dan sentimen negatif 431 untuk parfum *moris* dan data sentimen positif 456, sentimen netral 146, dan sentimen negatif 398 untuk parfum *scentplus* yang dilatih menggunakan algoritma *regresi logistik*.
3. Berdasar analisis yang telah dilakukan, terlihat bahwa hasil analisis menunjukkan model untuk parfum *Moris* memiliki performa terbaik dengan akurasi sebesar 93%, presisi sebesar 93%, dan recall sebesar 93%. Sedangkan model untuk parfum *Scentplus* memiliki akurasi sebesar 91%, presisi sebesar 91%, dan recall sebesar 91%.
4. Model untuk parfum *Moris* menunjukkan performa yang lebih baik karena memiliki data yang lebih bervariasi dibandingkan dengan model untuk parfum *Scentplus*.

DAFTAR PUSTAKA

- [1] Ni Luh Nurkariani, "Pentingnya Kepuasan Konsumen Untuk Meningkatkan Loyalitas Pelanggan," *Artha Satya Dharma*, vol. 15, no. 1, pp. 27–32, 2022, doi: 10.55822/asd.v15i1.238.
- [2] M. Riadi, J. Kamase, and M. Mapparenta, "Pengaruh Harga, Promosi Dan Kualitas Layanan Terhadap Kepuasan Konsumen Mobil Toyota (Studi Kasus Pada PT. Hadji Kalla Cabang Alauddin)," *J. Manag. Sci.*, vol. 2, no. 1, pp. 41–60, 2021, doi: 10.52103/jms.v2i1.320.
- [3] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, "Comparison of Logistic Regression, Multinomialnb, Svm, and K-Nn Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1487–1494, 2023, doi: 10.52436/1.jutif.2023.4.6.863.
- [4] M. Cindo and D. P. Rini, "Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review:

- Metode Klasifikasi Pada Sentimen Analisis,” *Januari*, pp. 66–70, 2019, [Online]. Available: <https://seminar-id.com/semnas-sainteks2019.html>
- [5] A. Jalu Narendra Kisma and C. Raras Ajeng Widiawati, “Analisis Aplikasi Di Playstore Berdasarkan Rating Dan Type Menggunakan Naive Bayes Dan Logistik Regresi,” *Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 174–184, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [6] Gede Aditra Pradinyana dkk, *Data Mining Menemukan Pengetahuan Dalam Data*. Rajawali Pers, 2020.
- [7] O. Hondro, “Analisis Sentimen Pengguna Aplikasi E-Wallet Dana Melalui Postingan di Media Sosial Twitter Menggunakan Naïve Bayes,” *KETIK J. Inform.*, 2023, [Online]. Available: <https://jurnal.faatuatua.com/index.php/KETIK/article/view/18>
- [8] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [9] Teguh Wahyono, *FUNDAMENTAL OF PYTHON MACHINE LEARNING*. GAVA MEDIA, 2021.
- [10] P. Studi, T. Informasi, F. Sains, D. A. N. Teknologi, U. I. N. Ar-raniry, and B. Aceh, “Perbandingan algoritma klasifikasi naïve bayes , logistic regression , dan knn untuk analisis sentimen peminatan masyarakat terhadap kandidat bakal calon presiden 2024,” 2024.
- [11] Anis Zubair, *DATA MINING Menggunakan Orange*. TEKNOSAIN, 2022.
- [12] Martin Fowler, *Panduan Singkat Bahasa Pemodelan Objek Standar*. ANDI, 2014.
- [13] Jubilee Enterprise, *Pemrograman Database dengan Python dan MySQL*. PT. Alex Media Komputindo, 2018.
- [14] Dicoding Intern, “Apa itu GitHub? Berikut Cara Menggunakannya.” [Online]. Available: <https://www.dicoding.com/blog/apa-itu-github/>
- [15] V. Cosentino, J. L. C. Izquierdo, and J. Cabot, “A Systematic Mapping Study of Software Development with GitHub,” *IEEE Access*, vol. 5, pp. 7173–7192, 2017, doi: 10.1109/ACCESS.2017.2682323.
- [16] Partha Mishra, *A Beginner`s Guide to Streamlit for Data Science Book 1 : Foundations*. Partha Mishra, 2023.