

Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare

Thomas Setiawan^{1*}, Steven Liem² dan M. Rizky Pribadi³

¹ Informatika, Universitas Multi Data Palembang; thomassetiawan13@mhs.mdp.ac.id;
stevenliem115@mhs.mdp.ac.id; rizky@mdp.ac.id

* Korespondensi: thomassetiawan13@mhs.mdp.ac.id;

Info Artikel:

Dikirim: 12 Juni 2024

Direvisi: 24 Juni 2024

Diterima: 28 Juni 2024

Intisari: Riset ini membandingkan performa algoritma Support Vector Machine (SVM) dan Naïve Bayes dalam analisis sentimen komentar TikTok tentang produk skincare, dengan menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketidakseimbangan data. Hasil evaluasi menunjukkan bahwa SVM memiliki performa yang lebih unggul dibandingkan Naïve Bayes dengan akurasi 59,43% berbanding 47,65%. Selain itu, SVM juga unggul dalam metrik F1 Score (60,37% dibandingkan 54,74%), meskipun Naïve Bayes memiliki presisi yang sedikit lebih tinggi (67,96% berbanding 62,76%). Dengan demikian, SVM lebih efektif dalam mengklasifikasikan sentimen komentar, menjadikannya algoritma yang lebih direkomendasikan untuk tugas analisis sentimen pada domain produk skincare di TikTok.

Kata Kunci: Analisis Sentimen; TikTok; Support Vector Machine (SVM); Produk Skincare; Naïve Bayes; SMOTE;

1. Pendahuluan

Analisis sentimen adalah proses mencari tahu sikap, emosi, atau opini seseorang terhadap suatu entitas. Entitas ini bisa berupa individu, kegiatan, atau topik [1]. Sikap atau opini ini biasanya dikategorikan sebagai positif, negatif, atau netral. Penelitian terkait sentimen merupakan salah satu bidang riset paling diminati dalam ilmu komputer dengan lebih dari 7.000 artikel telah ditulis [2].

Media sosial adalah wadah online yang memungkinkan individu untuk membuat, berbagi, dan terlibat dengan berbagai konten, sambil berkomunikasi dengan sesama pengguna. Ini telah menjadi sarana komunikasi dan pemasaran yang vital bagi perusahaan, lembaga, dan entitas, termasuk yang aktif dalam ranah politik [3]. Beberapa media sosial yang populer di Indonesia ialah Tiktok, Instagram, dan Facebook. Menurut data yang dilaporkan oleh We are Social dan Hootsuite, pada bulan Oktober 2023, jumlah pengguna Tiktok di Indonesia mencapai 106,52 juta orang. Jumlah tersebut meningkat sebesar 6,74% jika dibandingkan dengan tiga bulan sebelumnya [4].

Sentimen tentang suatu produk bisa diketahui melalui komentar pada postingan di Tiktok. Ini relevan karena setiap pengusaha tentu ingin mengetahui pandangan masyarakat atau pelanggan mengenai produk mereka [5]. Analisis sentimen digunakan untuk mengetahui bagaimana pengguna merespon sebuah topik atau sebuah merek di media sosial [6]. Analisis sentimen juga dapat membantu perusahaan meningkatkan produk mereka dan membangun pemasaran yang lebih efektif [7].

Ada banyak algoritma yang bisa dipakai untuk menganalisa sentimen. Beberapa diantaranya adalah SVM [8], Naïve Bayes Classifier [7], dan KNN [9]. Naïve Bayes merupakan algoritma yang sering digunakan saat jumlah data training sedikit [10]. Algoritma ini banyak digunakan dalam klasifikasi teks karena kesederhanaannya, kepraktisannya, dan tingkat akurasi yang tinggi [11]. Sedangkan beberapa penelitian menunjukkan SVM memiliki akurasi yang lebih tinggi dari Naïve Bayes [12][13].

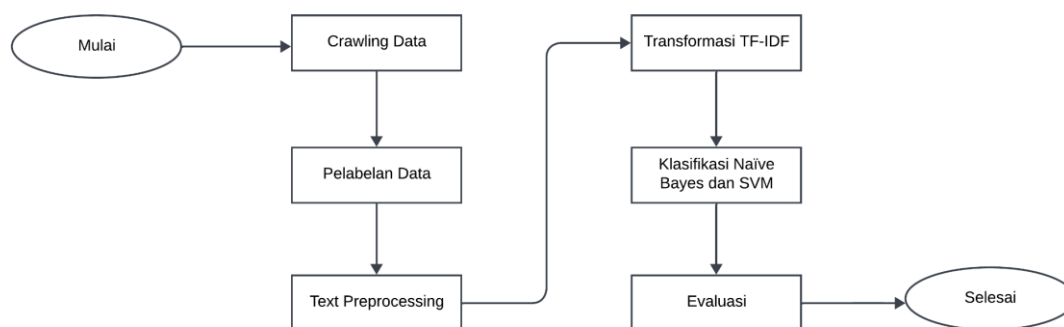
Dalam analisis sentimen, ketidakseimbangan kelas pada data dapat terjadi. Untuk mengatasi hal ini, salah satu strategi yang paling banyak digunakan adalah metode Synthetic Minority Oversampling Technique (SMOTE). Dasar dari prosedur SMOTE adalah melakukan interpolasi di antara instance kelas minoritas yang berdekatan. Dengan demikian, metode ini mampu meningkatkan jumlah instance kelas minoritas dengan memperkenalkan contoh-contoh

baru dari kelas minoritas di lingkungan tersebut, sehingga membantu classifier dalam meningkatkan kapasitas generalisasinya [14].

Riset ini membahas tentang cara memanfaatkan opini pengguna tiktok terhadap produk skincare untuk analisis sentimen melalui komentar pada postingan. Opini tersebut akan diklasifikasikan menjadi sentimen positif, atau sentimen negatif, atau sentimen netral. Sebelum diklasifikasikan, data akan mengalami preprocessing terlebih dahulu. Salah satunya menggunakan metode SMOTE untuk mengatasi ketidakseimbangan data. Pengklasifikasian tersebut nantinya akan dilakukan dengan 2 metode klasifikasi yang berbeda, yaitu Naïve Bayes Classifier dan SVM. Hasil akhir berupa akurasi, presisi, recall, dan f1-score dari kedua metode yang akan dibandingkan guna mengetahui metode dengan performa terbaik.

2. Metodologi Penelitian

Riset ini dijelaskan pada Gambar 1 secara berurutan.



Gambar 1. Alur Riset

2.1 Crawling data

Sumber data dikumpulkan dari platform media sosial yang bernama Tiktok. Penggunaan Tiktok dipilih karena banyaknya pengguna aktif yang sering membagikan opini dan reaksi mereka terhadap produk skincare, termasuk komentar mengenai pengalaman pribadi mereka dan pemikiran tentang produk tertentu. Data yang dikumpulkan mencakup komentar dalam Bahasa Indonesia pada postingan <https://www.tiktok.com/@eminacosmeticsid/video/7311565116120501509>. Data yang diambil sebanyak dibagi menjadi 70% untuk train dan 30% untuk test.

2.2. Pelabelan data

Pelabelan data dilakukan secara manual oleh 2 orang, yaitu penulis sendiri. 1 orang berperan memberikan label, dan 1 orang melakukan validasi. Label yang diberikan adalah positif, atau negatif, atau netral.

2.3 Text Preprocessing

Preprocessing dilakukan melalui 4 tahap, yaitu:

1. Case folding, yaitu proses mengubah semua huruf kapital dalam teks menjadi huruf kecil.
2. Tokenize, merupakan langkah memecah teks menjadi unit-unit yang lebih kecil yang disebut token. Token ini dapat berupa kata, frasa, atau karakter.
3. Penghapusan Stop words, yaitu menghilangkan kata-kata yang banyak muncul pada teks tetapi tidak mengandung informasi yang penting
4. Stemming, yaitu mengurangi kata-kata ke bentuk dasarnya atau akar katanya.

2.4 Transformasi TF-IDF

Dalam tahap ini, dilakukan penentuan bobot untuk setiap term dalam dokumen. Kata-kata yang sering muncul akan diberi bobot yang lebih tinggi, namun jika kata-kata tersebut juga umum dan muncul di banyak dokumen, maka nilai TF-IDF akan menurun.

2.5 Klasifikasi Naïve Bayes

Naive Bayes adalah sebuah metode klasifikasi sederhana yang berbasis probabilitas, dimana diasumsikan bahwa variabel penjelas bersifat independen. Algoritma Naive Bayes mengoperasikan dengan menggunakan teorema Bayes, yang berprinsip pada probabilitas bersyarat. Probabilitas bersyarat adalah probabilitas dari suatu peristiwa dengan memperhitungkan peristiwa terkait yang telah terjadi. Rumus Naïve Bayes sebagai berikut :

$$P(C | X) = \frac{P(X | C) P(C)}{P(X)} \tag{1}$$

Keterangan:
 $P(C | X)$ = $P(X | C)P(C)$
 $P(C | X)$ = probabilitas bersyarat C yang diberikan oleh X
 $P(X | C)$ = probabilitas bersyarat X yang diberikan oleh C
 $P(C)$ = probabilitas kejadian C
 $P(X)$ = probabilitas kejadian X

2.6 Klasifikasi Support Vector Machine (SVM)

Fungsi klasifikasi dalam SVM direpresentasikan oleh $f(x)$, yang berperan dalam memprediksi kelas target y berdasarkan fitur input x . Dalam rumus SVM, W merupakan vektor bobot, X adalah vektor fitur input, dan b adalah bias.

$$f(x) = W^T X + b \tag{2}$$

2.7 Evaluasi

Hasil klasifikasi kedua algoritma akan dievaluasi menggunakan confusion matrix. Confusion matrix adalah tabel yang menunjukkan jumlah data yang diklasifikasikan dengan benar dan yang salah, seperti yang terlihat pada Tabel 1.

Tabel 1. Confusion matrix

| Komentar | Prediksi Data | |
|----------|---------------|---------|
| | Positif | Negatif |
| Positif | TP | FP |
| Negatif | FN | TN |

3. Pembahasan

Visualisasi data dilakukan melalui word cloud seperti yang ditunjukkan Gambar 2.



Gambar 2. Word cloud

Hasil klasifikasi untuk metode naive bayes ditunjukkan pada Tabel 2, dan hasil klasifikasi untuk metode SVM ditunjukkan pada Tabel 3.

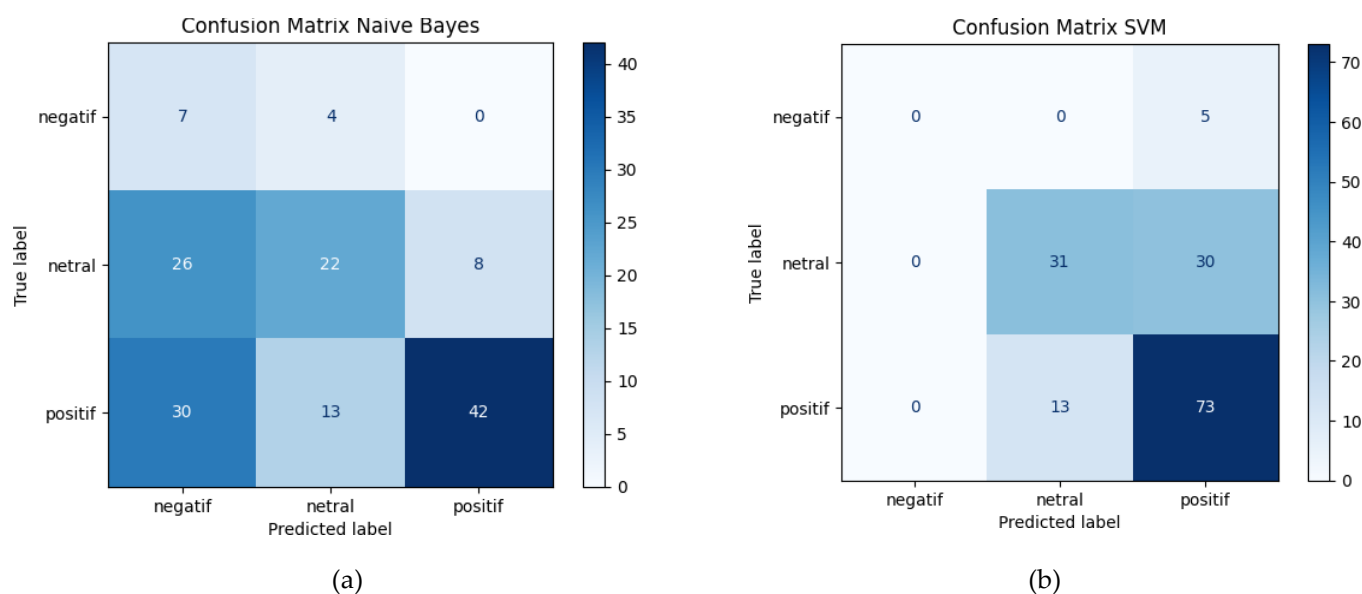
Tabel 2. Hasil klasifikasi menggunakan Naïve Bayes

| K-Fold | Akurasi | Presisi | Recall | F1-score |
|-----------|---------|---------|--------|----------|
| k=3 | 0.4731 | 0.6834 | 0.4871 | 0.5531 |
| k=5 | 0.4712 | 0.6821 | 0.4871 | 0.5508 |
| k=7 | 0.4850 | 0.6672 | 0.4751 | 0.5395 |
| k-10 | 0.4770 | 0.6857 | 0.4794 | 0.5463 |
| Rata-rata | 0.4765 | 0.6796 | 0.4821 | 0.5474 |

Tabel 3. Hasil klasifikasi menggunakan SVM

| K-Fold | Akurasi | Presisi | Recall | F1-score |
|-----------|---------|---------|--------|----------|
| k=3 | 0.5761 | 0.6477 | 0.6475 | 0.6032 |
| k=5 | 0.5900 | 0.6263 | 0.6435 | 0.6033 |
| k=7 | 0.6058 | 0.6200 | 0.6357 | 0.5970 |
| k-10 | 0.6056 | 0.6166 | 0.6437 | 0.6116 |
| Rata-rata | 0.5943 | 0.6276 | 0.6426 | 0.6037 |

Berikut adalah confusion matrix metode naive bayes pada Gambar 3, sedangkan metode svm pada Gambar 4.



Gambar 3. (a) Confusion matrix untuk metode Naïve Bayes; (b) Confusion matrix untuk metode SVM

4. Kesimpulan

Riset ini membandingkan kinerja dua algoritma pembelajaran mesin, yaitu Support Vector Machine (SVM) dan Naïve Bayes, dalam melakukan analisis sentimen terhadap komentar TikTok mengenai produk skincare.. SVM

menunjukkan akurasi yang lebih tinggi (59,43% dibandingkan 47,65%) dan F1 Score yang lebih baik (60,37% dibandingkan 54,74%). Meskipun presisi Naïve Bayes sedikit lebih tinggi (67,96% dibandingkan 62,76%), recall dan metrik F1 Score SVM lebih baik, menunjukkan bahwa SVM lebih efektif dalam menangkap pola-pola sentimen dalam data komentar yang digunakan. Secara keseluruhan, SVM lebih direkomendasikan untuk tugas analisis sentimen pada domain ini karena memberikan kinerja yang lebih konsisten dan akurat.

Daftar Pustaka

- [1] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- [2] Rita Apriani, Dudih Gustian, Jurnal Rekayasa Teknologi Nusa Putra. Vol. 6, No. 1, September 2019: Hal 54 - 62
- [3] Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48(1), 79–95. <https://doi.org/10.1007/s11747-019-00695-1>
- [4] "dataIndonesia.id" . (2023, Desember 12). *Data Jumlah Pengguna TikTok di Indonesia hingga Oktober 2023*. <https://dataindonesia.id/ekonomi-digital/detail/data-jumlah-pengguna-tiktok-di-indonesia-hingga-oktober-2023>
- [5] Muhammad Arief Rahman, Herman Budianto, & Setiawan, E. I. (2019). Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network. *INSYST: Journal of Intelligent System and Computation*, 1(2), 50–57. <https://doi.org/10.52985/insyst.v1i2.83>
- [6] Kaharudin, A., Agus Supriyadi, A., Baitika, H., & Derryanur, M. (2023). OKTAL : Jurnal Ilmu Komputer dan Science Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes: Kajian Literatur Review. 2(6). <https://harzing.com/resources/publish-or-perish>
- [7] Amelia, E. E., & Yustiana, I. (2024). Analisis Sentimen Pada Ulasan Produk UNIQLO dengan Algoritma Naive Bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 8, Issue 1).
- [8] Oktavia, D., & Ramadahan, Y. R. (2023). Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Media Online*, 4(1), 407–417. <https://doi.org/10.30865/klik.v4i1.1040>
- [9] Lillah, M. R. R. L., Maylawati, D. S. ., Zulfikar, W. B. ., Uriawan, W. ., & Wahana, A. . (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Analisis Sentimen Pengguna Aplikasi Tokopedia . *Intellect : Indonesian Journal of Learning and Technological Innovation*, 2(2), 171–184.
- [10] Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, 87, 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>
- [11] Daffa Muafa Lizda Iswari, M. (n.d.). *Pengembangan Aplikasi Berbasis Web dengan Rshiny untuk Data Klasifikasi Menggunakan Metode Naive Bayes*.
- [12] Noviana, R., & Rasal, I. (2023). Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter
- [13] Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine(SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375-384.
- [14] Fern´andez, A., Garcia, S., Herrera, F. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, 61. <https://doi.org/10.1613/jair.1.11192>