

ANALISIS SENTIMEN PRODUK SKINCARE SCARLETT WHITENING DENGAN BRAND AMBASSADOR EXO MENGGUNAKAN METODE NAÏVE BAYES

Rahajeng Audi Nurianingrum¹, Muhammad Eka Purbaya²

¹ Program Studi Sistem Informasi, Fakultas Informatika, Institut Teknologi Telkom Purwokerto, Jl. DI Panjaitan No.128, Karangreja, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah 53147, Indonesia

² Program Studi S1 Bisnis Digital, Telkom University, Jl. DI Panjaitan No.128, Karangreja, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah 53147, Indonesia

Email: 20103113@itttelkom-pwt.ac.id¹, mekapur@telkomuniversity.ac.id²

Received: March 03, 2024 / Revised: August 29, 2024 / Accepted: Sept 30, 2024

Abstrak

Kebutuhan akan produk perawatan kulit (*skincare*) semakin meningkat, termasuk di Indonesia. Salah satu produk lokal yang populer adalah *Scarlett Whitening*, yang pada 14 September 2023 berkolaborasi dengan grup EXO sebagai Brand Ambassador. Kolaborasi ini menimbulkan beragam tanggapan dari masyarakat Indonesia, terutama di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kolaborasi tersebut dengan menggunakan data komentar dari YouTube channel resmi EXO. Data dianalisis menggunakan metode klasifikasi Naïve Bayes dan *Support Vector Machine* (SVM) melalui proses *text mining* dan *sentiment analysis*. Hasil penelitian menunjukkan bahwa secara umum, mayoritas komentar pengguna bersentimen **positif**, dengan total 869 komentar positif, diikuti oleh 137 komentar netral dan 179 komentar negatif. Hal ini menunjukkan bahwa masyarakat Indonesia cenderung memberikan tanggapan yang baik terhadap kolaborasi EXO dengan *Scarlett Whitening*. Naïve Bayes menunjukkan performa lebih baik dengan akurasi 73% dan F1 score 75%, dibandingkan SVM yang hanya mencapai akurasi 72% dan F1 score 60%. Berdasarkan hasil tersebut, Naïve Bayes dinilai lebih sesuai untuk digunakan dalam analisis sentimen terhadap komentar media sosial YouTube. Secara keseluruhan, hasil analisis menunjukkan bahwa strategi pemasaran *Scarlett Whitening* yang menggandeng EXO sebagai Brand Ambassador cukup efektif dan mendapat respon positif dari masyarakat, yang dapat berkontribusi dalam peningkatan citra dan penjualan produk.

Kata kunci: Analisis Sentimen, Brand Ambassador, Naïve Bayes, Scarlett Whitening, YouTube

Abstract

The demand for skincare products continues to rise globally, including in Indonesia. One of the popular local brands is Scarlett Whitening, which collaborated with the K-pop group EXO as its Brand Ambassador on September 14, 2023. This collaboration sparked various public reactions, particularly on social media. This study aims to analyze public sentiment toward the collaboration using comment data from EXO's official YouTube channel. The data was analyzed using Naïve Bayes and Support Vector Machine (SVM) classification methods through text mining and sentiment analysis techniques. The results show that, in general, most user comments were **positive**, with a total of 869 positive comments, followed by 137 neutral and 179 negative comments. This indicates that Indonesian audiences tended to respond positively to the partnership between EXO and Scarlett Whitening. The Naïve Bayes algorithm demonstrated better performance with an accuracy of 73% and an F1 score of 75%, compared to SVM, which achieved an accuracy of 72% and an F1 score of 60%. Based on these findings, Naïve Bayes is considered more suitable for sentiment analysis of social media comments on YouTube. Overall, the analysis suggests that Scarlett Whitening's marketing strategy involving EXO as Brand Ambassador was effective and well-received by the public, potentially enhancing the brand's image and sales.

Keywords: Brand Ambassador, EXO, Naïve Bayes, Sentiment Analysis, YouTube

1. Pendahuluan

Kebutuhan akan produk perawatan kulit (*skincare*) semakin meningkat di seluruh dunia, termasuk Indonesia. Kini masyarakat mulai memiliki kesadaran yang tinggi akan pentingnya memiliki kulit yang sehat untuk kenyamanan dan kecantikan menjadi faktor pendorongnya. *Skincare* sendiri memiliki peran penting

dalam merawat kulit sejak usia remaja hingga dewasa. Salah satu industri yang paling menjanjikan bagi perusahaan adalah *skincare*, terutama bagi negara berkembang seperti Indonesia (Mas & Mansyur, 2023). Peningkatan kesadaran akan perawatan kulit dan permintaan konsumen yang tinggi telah membuat industri ini menjadi sangat kompetitif (Fauzan & Aisyah,

2023). Pendapatan yang diperoleh dari penjualan perawatan kulit dan kosmetik di Indonesia di tahun 2019 mendekati US\$6,9 miliar. Diprediksi, pendapatan ini akan terus meningkat hingga mencapai 10% pada tahun 2024, yang sebagian besar berasal dari penjualan online (D. N. Sari & Feranita, 2022).

Scarlett Whitening merupakan salah satu produk perawatan wajah dan tubuh lokal yang didirikan oleh Felicya Angelista pada tahun 2017. Meskipun terbilang baru, *Scarlett Whitening* mampu menjadi salah satu kompetitor terbaik di Indonesia. Hal ini karena *Scarlett Whitening* memiliki produk yang berkualitas dan menggunakan strategi pemasaran yang tepat (Tamonsang, Matheous; Putri, 2023). Produk *Scarlett* disukai banyak orang karena mampu membuat kulit terlihat lebih bersih dan cerah. Salah satu strategi pemasaran yang digunakan untuk mempromosikan produknya adalah dengan menggandeng celebrity Indonesia sejak tahun 2017. Pada tahun 2021, *Scarlett Whitening* untuk pertama kalinya menggandeng selebriti asing asal Korea Selatan. Strategi pemasaran yang efektif dari *Scarlett* telah menjadikannya salah satu produk perawatan kulit yang paling populer di Indonesia (Puspita et al., n.d.).

Scarlett Whitening terus berinovasi untuk menarik minat konsumen dan bersaing dengan kompetitor. Melihat banyaknya penggemar K-pop di Indonesia, *Scarlett Whitening* menggandeng EXO sebagai *brand ambassador*. Para penggemar K-pop rela mengeluarkan uang untuk membeli produk yang diiklankan oleh idolanya, termasuk produk *Scarlett Whitening*. Pemilihan EXO sebagai *brand ambassador* produk *Scarlett* karena group tersebut memiliki popularitas, pengaruh, kredibilitas, dan daya tarik pasar yang luas. Hal ini dapat membantu *Scarlett* meningkatkan kesadaran merek, memperluas basis konsumen, dan membangun hubungan yang lebih kuat dengan penggemar setia EXO serta penggemar produk tersebut (Suryaningsih & Ningtias, 2023).

Saat ini, internet adalah wadah bagi berbagai jenis media sosial. Berdasarkan data HootSuite Indonesia pada tahun 2021, YouTube menempati posisi pertama sebagai media sosial dengan jumlah pengguna terbesar, yaitu 107 juta pengguna. Angka-angka tersebut menunjukkan bahwa YouTube dapat menjadi tempat untuk memberikan komentar, masukan, dan opini terkait video yang telah diposting. Cara alternatif untuk menilai sentimen pendapat masyarakat terhadap sebuah konten adalah dengan menghitung jumlah komentar netral, positif, dan negatif terhadap konten tersebut di media sosial (Abdulloh & Pambudi, 2021). Salah satu contohnya adalah kolaborasi yang dilakukan brand *Scarlett* dengan menjadikan EXO sebagai brand ambassador. Seperti diketahui, kolaborasi *Scarlett* dengan EXO cukup ramai dibicarakan oleh masyarakat Indonesia (Suryaningsih & Ningtias, 2023).

Berdasarkan hal tersebut untuk mengetahui sentimen masyarakat terhadap kolaborasi tersebut perlu dilakukan analisis sentimen terhadap hal tersebut. Analisis sentimen adalah proses untuk mengetahui nada emosional dari teks, seperti apakah teks tersebut positif, negatif, atau netral. Analisis sentimen bertujuan untuk mengekstraksi informasi tentang sentimen dari teks, seperti apakah teks tersebut berisi pujian, kritik, atau informasi netral. Dengan dilakukannya analisis sentimen, dapat dengan mudah untuk membantu memahami bagaimana orang merespons terhadap suatu topik, produk, layanan, atau merek tertentu (Syah & Witanti, 2022). Analisis yang dilakukan pada penelitian ini diambil dan dilakukan *crawling* data dari postingan promosi yang ada pada channel resmi EXO melalui sosial media Youtube melalui tautan berikut : <https://youtu.be/2SHC3y08Dhw?si=doBpMfCKueyV3S50> yang diupload tanggal 14 September 2023 dengan jumlah komentar mencapai 2.668.

Oleh karena banyaknya jumlah data komentar yang tidak memungkinkan untuk dianalisis satu per satu, maka pendekatan *Text mining* digunakan untuk mengolah data komentar tersebut. Data komentar tersebut diklasifikasikan sebagai komentar positif, negatif, atau netral. Terdapat metode pendekatan klasifikasi yang dapat digunakan untuk melakukan analisis sentimen. Metode utama yang digunakan pada penelitian ini adalah Naïve Bayes, karena metode tersebut mampu memberikan tingkat akurasi yang tinggi dalam mengategorikan ulasan ke dalam pendapat positif, netral, dan negatif. Naïve Bayes juga dikenal sebagai metode klasifikasi yang paling sederhana (Indrawan et al., 2023). Ada beberapa penelitian terkait yang sudah dilakukan dengan menggunakan algoritma Naïve bayes. Pada penelitian (Zhafira et al., 2021) yang telah dilakukan sebelumnya memperoleh hasil tingkat akurasi terbaik sebesar 91,8% dengan nilai precision, recall, dan *F1 Score* sebesar 90.35%, 93.6%, 91.95%.

Selain Naïve Bayes, penelitian ini juga mencoba untuk menggabungkan metode *Support Vector Machines* sebagai metode usulan. *Support Vector Machines* adalah algoritma pembelajaran mesin yang dapat menghasilkan klasifikasi yang akurat, bahkan dengan jumlah data yang relatif kecil. *Support Vector Machines* memiliki konsep dan formulasi yang mudah dipahami, sehingga mudah untuk dimodifikasi sesuai dengan kebutuhan. *Support Vector Machines* juga dapat diimplementasikan dengan relatif mudah (Rahayu et al., 2022). Terdapat penelitian terkait yang sudah dilakukan dengan menggunakan algoritma *Support Vector Machines*. Pada penelitian (Asrianto & Herwinanda, 2022) yang telah dilakukan sebelumnya diperoleh hasil tingkat keakuratan terbaik dengan nilai akurasi tertinggi yaitu 86.33%. Hal ini diikuti oleh nilai presisi 75%, nilai recall 66.67%, dan nilai *f1-score* 70.59%.

Berdasarkan penjelasan diatas, peneliti ingin mengetahui cara yang digunakan untuk melakukan

analisis sentimen, hasil yang diperoleh, dan mengetahui algoritma yang terbaik untuk digunakan dalam analisis sentimen. Harapannya hasil yang diperoleh dapat menjadi bahan referensi untuk penelitian selanjutnya untuk melakukan analisis sentimen dengan studi kasus yang serupa.

2. Dasar Teori

2.1 Analisis sentimen

Analisis sentimen adalah teknik yang digunakan untuk mengubah data opini menjadi bentuk yang lebih terstruktur, sekaligus memahami dan mengolah data teks secara otomatis guna mengidentifikasi sentimen di dalam suatu opini (F. V. Sari & Wibowo, 2019). Melalui metode ini, informasi yang sebelumnya tidak terorganisir dapat disusun menjadi data yang lebih sistematis (Nitamia & Februariyanti, 2022). Analisis sentimen bertujuan untuk memetakan pendapat pengguna sesuai dengan topik yang telah ditentukan, serta membantu memperoleh informasi yang lebih rinci dan akurat dari berbagai sudut pandang ulasan pengguna (Rambe et al., 2023). Fokus utama dari analisis sentimen adalah pada opini-opini yang mencerminkan atau menunjukkan kecenderungan sentimen positif maupun negatif (Gusriani et al., 2016).

2.2 Brand Ambassador

Brand Ambassador adalah individu yang memiliki ketertarikan serta pemahaman yang baik terhadap suatu produk atau merek. Peran Brand Ambassador berpotensi memengaruhi sikap dan keputusan konsumen dalam memilih serta menggunakan produk tersebut. Oleh karena itu, perusahaan perlu selektif dalam menentukan Brand Ambassador, yakni memilih sosok yang mampu memberikan pengaruh positif dan sesuai dengan karakteristik target pasar (Brestilliani, 2020). Kehadiran Brand Ambassador yang tepat tidak hanya mendorong minat beli konsumen, tetapi juga berperan dalam menciptakan tren di pasar. Umumnya, selebriti kerap dipilih sebagai Brand Ambassador karena dinilai memiliki daya tarik psikologis yang kuat di mata konsumen. Selebriti dapat berfungsi sebagai perantara, juru bicara, sekaligus penghubung dalam sebuah promosi, sehingga pesan produk atau jasa dapat tersampaikan dengan lebih efektif. Selain itu, kepribadian Brand Ambassador juga turut membentuk citra merek di benak konsumen, yang pada akhirnya memengaruhi persepsi masyarakat terhadap merek tersebut dan mendorong keputusan untuk membeli (Yusuf, 2021).

2.3 YouTube

Youtube adalah sebuah *platform* situs web video sharing yang sangat populer di seluruh dunia dan menawarkan berbagai macam konten video, mulai dari film pendek, klip film, klip musik, hingga konten amatir seperti vlog, video pendek original, dan video pendidikan (Nurhadi et al., 2020). YouTube adalah layanan berbagi

video dari Google yang memungkinkan penggunaanya mengunggah dan menonton video tanpa biaya. YouTube memiliki jumlah dan variasi video yang sangat luas, YouTube dikenal sebagai salah satu basis data video paling populer di dunia maya (Hurley et al., 2016).

2.4 Orange Data Mining

Orange adalah salah satu perangkat lunak data mining berbasis *open source* yang berfungsi untuk melakukan analisis sekaligus visualisasi data. *Software* ini memiliki fitur-fitur yang lengkap dan mudah dipahami. *Widget* yang disediakan dapat digunakan untuk berbagai macam tugas *data mining*, mulai dari eksplorasi data hingga prediksi. Hasil analisis juga ditampilkan secara jelas, sehingga pengguna dapat dengan mudah memahaminya (Info et al., 2023).

2.5 Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dalam *machine learning* yang berfungsi untuk menyelesaikan permasalahan klasifikasi data. Algoritma ini banyak digunakan karena kemudahannya dalam menggunakannya dan sering menghasilkan tingkat akurasi yang sebanding daripada algoritma yang lain. Selain itu, Naïve Bayes juga dipandang sebagai algoritma yang efektif dan efisien. Algoritma Naïve Bayes terdiri dari dua tahapan yaitu pelatihan dan klasifikasi. Pada tahap pelatihan, kata dalam dataset akan diuji dataset pelatihan untuk menghitung probabilitas masing-masing kategori sentimen (positif dan negatif). Perlu mempelajari pola dalam dataset pelatihan terlebih dahulu sebelum melakukan klasifikasi. Selanjutnya, pada langkah klasifikasi, probabilitas setiap kategori sentimen dihitung terhadap data yang dimasukkan dan hasilnya adalah probabilitas dari setiap kategori sentimen yang diberikan pada dataset (Tauhid & Ruldeviyani, 2020).

2.6 Support Vector Machine (SVM)

Support Vector Machine adalah salah satu algoritma dalam *machine learning* yang memanfaatkan *hyperplane* untuk memisahkan data ke dalam beberapa wilayah kelas. *Hyperplane* sendiri berperan sebagai garis batas pemisah antar kelas data. Saat melakukan prediksi, SVM akan menentukan label kelas suatu data berdasarkan posisi wilayah kelas tempat data tersebut berada. Algoritma ini umumnya digunakan dalam pengolahan dataset berukuran besar yang sering bersumber dari situs online, dan SVM cukup populer berkat efektivitasnya dalam tugas-tugas klasifikasi teks. Inti dari metode SVM adalah membentuk *hyperplane* dengan margin yang seimbang, di mana posisi *hyperplane* tidak condong ke salah satu sisi kelas. Proses ini dilakukan dengan menghitung margin dan mencari nilai maksimumnya. Pencarian *hyperplane* terbaik sebagai pemisah antar kelas menjadi fokus utama dalam penerapan algoritma SVM (Fikri et al., 2020).

2.7 Text mining

Text mining adalah penerapan konsep dan teknik data mining untuk mengidentifikasi pola dalam teks untuk menemukan informasi yang bermanfaat untuk tujuan tertentu. Proses *text mining* dapat dipandang sebagai dua tahapan utama. Tahap pertama adalah pemberian struktur pada data teks mentah, sedangkan tahap kedua melibatkan pengambilan informasi dan pengetahuan penting dari data yang telah terstruktur dengan memanfaatkan teknik dan alat yang serupa dengan yang digunakan dalam data mining (D. N. Sari et al., 2020). *Text mining* sendiri mencakup berbagai aktivitas seperti pengkategorian, pengelompokan teks, ekstraksi konsep atau entitas, analisis sentimen, pembuatan ringkasan dokumen, serta pemodelan hubungan antar entitas (Gusriani et al., 2016).

2.8 Pre-processing Text

Pre-processing Text adalah langkah awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan menghapus informasi yang tidak relevan, sehingga data siap digunakan dalam proses analisis sentimen (Putri et al., 2022).

2.9 Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF adalah salah satu metode pembobotan yang populer dan banyak digunakan karena mampu memberikan tingkat akurasi dan recall yang cukup baik. Metode ini umumnya diterapkan dalam proses pencarian informasi (*information retrieval*). Dalam TF-IDF, tingkat kepentingan sebuah kata dihitung berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen. Semakin tinggi frekuensinya di dalam dokumen, semakin besar bobotnya. Namun, jika kata tersebut muncul di banyak dokumen, bobotnya justru akan berkurang. TF-IDF sendiri terdiri dari dua elemen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF) (Yutika et al., 2021).

2.10K Fold Cross Validation

K-fold cross validation adalah metode yang digunakan untuk mengestimasi tingkat kesalahan prediksi pada model *machine learning*. Teknik ini membagi dataset menjadi beberapa subset berukuran sama, lalu model akan dilatih dan diuji sebanyak jumlah subset yang telah ditentukan. Pada setiap iterasi, satu subset dipilih sebagai data pengujian, sementara bagian lainnya digunakan sebagai data pelatihan (Mardiana et al., 2022).

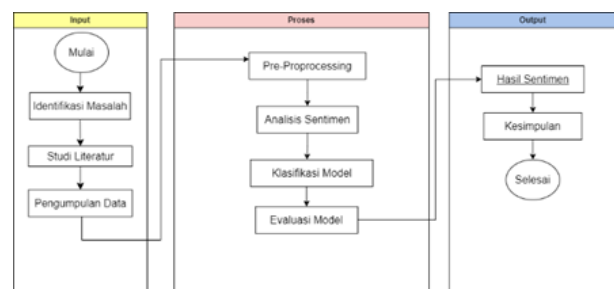
2.11Confusion matrix

Confusion matrix adalah metode yang digunakan dalam data mining untuk mengevaluasi tingkat akurasi model klasifikasi. Terdapat empat istilah yang digunakan untuk menggambarkan hasil klasifikasi, yaitu *True Positive (TP)*, yang mengacu pada nilai positif yang terdeteksi dengan benar. *True Negatif (TN)* menyatakan

jumlah data negatif yang berhasil dideteksi dengan benar. *False Positif (FP)* merujuk pada situasi di mana data yang sebenarnya negatif terdeteksi sebagai positif. Sementara itu, *False Negatif (FN)* mencerminkan kondisi ketika data yang sebenarnya negatif terdeteksi sebagai negatif (Mutawalli et al., 2019).

3. Metode Penelitian

Terdapat beberapa tahapan dalam penelitian ini antara lain mengidentifikasi masalah, studi literatur, pengumpulan data komentar yang diperoleh dari proses crawling data pada sosial media YouTube dengan Python dan Youtube API, pre-processing dengan software Orange Data Mining, analisis sentimen, klasifikasi model dengan algoritma Naïve Bayes dan *Support Vector Machine (SVM)*, evaluasi model dengan confusion matrix, hasil sentimen, dan terakhir kesimpulan. Pada gambar 1 disajikan diagram alir penelitian.



Gambar 1. Diagram alir penelitian

3.1 Identifikasi Masalah

Tahap pertama, peneliti membuat perumusan masalah dengan melihat fenomena yang sedang populer saat ini mengenai penggunaan *Brand Ambassador* idol Korea EXO dalam mempromosikan produk *skincare Scarlett Whitening*. Selanjutnya peneliti menentukan tujuan dan manfaat dari penelitian ini.

3.2 Studi Literatur

Tahap kedua, peneliti mencari referensi jurnal-jurnal yang relevan dengan penelitian ini. Peneliti menggunakan jurnal atau artikel ilmiah dengan rentang waktu 5 tahunan terakhir dari sumber yang terpercaya seperti *Google Scholar*, *Science Direct* dan *Knowbase*.

3.3 Pengumpulan Data

Pada tahap ini dilakukan dengan cara crawling di media sosial YouTube menggunakan YouTube API dengan bantuan google colab. Data yang diambil berupa data komentar pada video pada *channel* resmi EXO yang diunggah pada tanggal 14 September 2023 dengan total 2.667 komentar.

3.4 Pre-processing

Tahapan preprocessing data mengharuskan peneliti untuk memproses data komentar yang telah dikumpulkan. Hal ini karena data komentar yang

didapatkan tidak semuanya dapat digunakan dalam proses analisis, terdapat beberapa bagian yang dihilangkan atau dirubah. Dengan begitu data yang akan digunakan dapat sesuai dengan kebutuhan ada proses analisis. Proses ini terdapat beberapa tahapan yang dilakukan yaitu *transformation*, *tokenization*, *normalization* dan *filtering* yang sebelum nantinya data di simpan dalam bentuk excel.

3.5 Analisis Sentimen

Langkah lanjutan dari preprocessing data dimana setelah melakukan pemrosesan terhadap data yang diperoleh peneliti kemudian dilakukan sentimen analisis. Dalam penelitian ini, digunakan teknik Lexicon Based, yang mencocokkan kata dasar dari ulasan dengan kamus kata positif dan negatif yang telah disusun oleh peneliti. Tahap ini juga dilakukan labeling untuk mengklasifikasikan komentar-komentar tersebut untuk mengetahui komentar yang bernilai positif, negatif dan netral menggunakan *tools orange data mining*.

3.6 Klasifikasi Model

Tahapan berikutnya yaitu klasifikasi model, dimulai dari mengkombinasikan informasi awal probabilitas sebelumnya dan informasi baru probabilitas kondisional untuk menghasilkan rumus yang menghitung kemungkinan setiap kategori sentimen.

3.7 Evaluasi Model

Pada tahap evaluasi, *confusion matrix* digunakan untuk menghitung *akurasi*, *recall*, *precision*, dan *f1-score* dari kedua model klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine*, yang diterapkan untuk menganalisis komentar masyarakat terkait produk *Scarlett Whitening* yang menjadikan EXO sebagai *Brand Ambassador*.

3.8 Hasil Sentimen

Hasil analisis sentimen yang sudah dilakukan dapat memberikan hasil yang dapat dilihat pada *tools test and score* melalui *evaluation result*.

3.9 Kesimpulan

Hasil yang telah didapatkan dari proses pengujian keakuratan klasifikasi akan disimpulkan. Kesimpulan ini akan menjawab rumusan masalah dan pertanyaan peneliti yang telah dijelaskan sebelumnya.

4. Hasil dan Pembahasan

Berdasarkan langkah-langkah dalam metodologi penelitian yang telah dilakukan, berikut adalah penjelasan mengenai proses yang diterapkan:

3.1 Crawling Data

Dalam penelitian ini, data diperoleh dari data komentar Youtube yang diunggah pada tanggal 14 September 2023 oleh channel Youtube resmi EXO

dengan judul “For you, Sahabat Scarlett!💖 | EXO Scarlett Behind”. Terdapat 2.667 komentar yang didapatkan melalui tahapan Youtube Crawling. Komentar tersebut dipilih sebanyak 1186 komentar dalam kategori bahasa Indonesia. Pengambilan data menggunakan metode crawling data komentar Youtube dengan menggunakan YouTube API. Proses crawling data menggunakan YouTube API dapat dilihat pada Gambar 2.

```
[ ] # isikan dengan api key Anda
    api_key = 'AIzaSyAjlGkl2kjFOHLVNFGoPbbkpp1_RTqW08BI'

    # Enter video id
    # contoh url video = https://www.youtube.com/watch?v=2SHC3y08Dhw
    video_id = "2SHC3y08Dhw" #isikan dengan kode / ID video

    # Call function
    comments = video_comments(video_id)

    comments
```

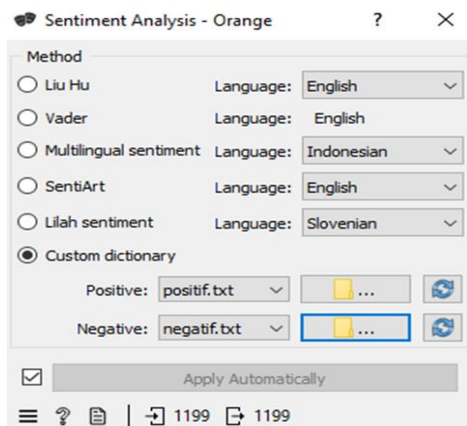
Gambar 2. Proses crawling data

Setelah proses *crawling* data selesai, maka disimpan dengan format .csv agar dapat diolah ke tahap berikutnya menggunakan *software Orange Data Mining*. Data yang disimpan dapat dilihat pada tabel 1.

Tabel 1. Hasil *Crawling* Data YouTube

No	Coments
1	sehat selalu kesayangan, semoga bisa bertemu aaamiin
2	kalian semua sangat tampan ; bahagia, terimakasih banyak sudah bekerja keras ; terimakasih Scarlett sudah memilih EXO
3	terimakasih Scarlett sudah memilih EXO
4	AAAAAA SUKA BANGETTTT... SEMOGA BISA KETEMU MEREKA SEMUA YAAALLAH AAAMIIN
5	Nanti konser di Indonesia full member yaa kalo bisa OT9
6	Lay Katanya Gak Bisa Hanya Bisa Mengantarkan Exo Internasional Supaya Menang diajang Award Dunia
7	sepanjang nonton konyten ini senyum senyum sendiri kekk bangga bangeett produk lokall kollab ama ayang2kuu
8	terima kasih banyak scarlett, lancar selalu
9	sehat selalu kesayangan, semoga bisa bertemu aaamiin
...	
1186	MAKASIH EXO KERANA TELAH DATANG KE INDONESIA!!! WE LOVE YOU EXO!!!

Tabel 1 menunjukkan hasil *crawling* data komentar Youtube yang terdiri dari variabel *content*, yang merupakan isi cuitan dari masyarakat yang menanggapi terkait EXO yang menjadi *brand ambassador* produk *scarlett*.



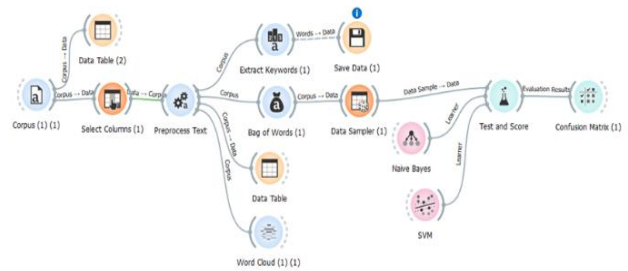
Gambar 9. Sentiment Analysis

3.4 Klasifikasi Model

Pada tahap ini, kedua metode klasifikasi yang digunakan adalah Naïve Bayes dan Support Vector Machine, yang dapat dilihat pada Gambar 10. Data hasil analisis sentimen yang telah diberi label kemudian diimpor ke Orange Data Mining melalui widget corpus. Selanjutnya, kolom sentimen dan komentar dipilih untuk dianalisis lebih lanjut. Setelah itu, data sentimen diproses kembali dengan langkah yang serupa dengan sebelumnya, termasuk pre-processing ulang. Namun, kali ini ditambahkan komponen filter most frequent tokens untuk membatasi kata-kata yang relevan untuk diproses. Proses berikutnya adalah penerapan metode scoring dengan menggunakan TF-IDF untuk memberikan bobot yang lebih besar pada kata-kata yang penting dan lebih kecil pada kata yang kurang penting. Kata-kata dengan nilai TF-IDF yang lebih tinggi akan dianggap sebagai kata kunci penting, seperti yang terlihat pada gambar 11. Selanjutnya, data teks diubah menjadi vektor yang dapat dipahami oleh komputer dengan menggunakan metode bag of words, yang membantu menghitung frekuensi kemunculan kata dalam seluruh dokumen.

3.5 Evaluasi Model

Evaluasi dilakukan untuk mengetahui performa dalam melakukan klasifikasi analisis sentiment dengan menggunakan metode Cross Validation dengan 10 fold. Data yang telah didapatkan dari proses crawling kemudian diklasifikasikan ke dalam kategori positif, negatif, dan netral menggunakan ketiga algoritma tersebut dengan bantuan software Orange Data Mining. Hasil klasifikasi kemudian dianalisis menggunakan *confusion matrix* untuk mengetahui performansi model seperti pada tabel 2 dan tabel 3.



Gambar 10. Model Klasifikasi Model

Word	TF-IDF
terimakasih	0.108
exo	0.082
indonesia	0.053
scarlett	0.046
banquet	0.036
sayang	0.029
ya	0.029
qanteng	0.028
bqt	0.026
konser	0.026
kanzen	0.022
br	0.022
tunggu	0.022
nya	0.020
qa	0.019
makasih	0.018
ditunggu	0.017
lucu	0.015

Gambar 11. Proses TF-IDF

Tabel 2 menunjukkan hasil confusion matrix dari model klasifikasi Naive Bayes yang digunakan untuk mengklasifikasikan data ke dalam tiga kategori: Negatif, Netral, dan Positif. Berdasarkan tabel tersebut, model mampu mengklasifikasikan sebagian besar data dengan cukup baik. Sebanyak 101 data dengan label aktual Negatif berhasil diklasifikasikan dengan benar, sementara sisanya salah diklasifikasikan sebagai Netral (34) dan Positif (11). Untuk label Netral, hanya 51 data yang diklasifikasikan dengan benar, sedangkan 13 diklasifikasikan sebagai Negatif dan 56 sebagai Positif. Kategori Positif menunjukkan performa terbaik, dengan 540 data diklasifikasikan secara tepat, meskipun masih terdapat kesalahan klasifikasi ke Negatif (22) dan Netral (120). Total data yang diuji berjumlah 948, dan secara umum, Naive Bayes menunjukkan kinerja yang cukup seimbang, meski masih terdapat misclassifications yang cukup signifikan terutama antara Netral dan Positif.

Pada tabel 2 dapat dijelaskan bahwa hasil confusion matrix algoritma *Support Vector Machine* terdapat 682 data berhasil diprediksi dengan benar sebanyak 682 data, sedangkan dari 120 data dan 146 data diprediksi positif tetapi sebenarnya negatif.

Tabel 2. Hasil Confusion Matrix Metode Naïve Bayes

Actual \ Predicted	Negatif	Netral	Positif	Σ (Total Actual)
Negatif	101	34	11	146
Netral	13	51	56	120
Positif	22	120	540	682
Σ (Total Predicted)	136	205	607	948

Tabel 3. Hasil Confusion Matrix Metode Support Vector Machine

Actual \ Predicted	Negatif	Netral	Positif	Σ (Total Actual)
Negatif	0	0	146	146
Netral	0	0	120	120
Positif	0	0	682	682
Σ (Total Predicted)	0	0	948	948

Tabel 3 menampilkan confusion matrix dari model klasifikasi SVM (Support Vector Machine) untuk tugas klasifikasi yang sama. Namun, hasil yang ditampilkan menunjukkan bahwa model SVM mengalami overfitting ekstrem atau kegagalan dalam proses pelatihan. Hal ini ditandai dengan seluruh data dari ketiga kelas aktual (Negatif, Netral, dan Positif) diklasifikasikan sebagai Positif, tanpa ada satu pun yang diklasifikasikan sebagai Negatif atau Netral. Sebanyak 146 data Negatif, 120 data Netral, dan 682 data Positif semuanya diprediksi sebagai Positif, menghasilkan distribusi prediksi yang sangat timpang. Kondisi ini menunjukkan bahwa model SVM tidak mampu membedakan fitur dari ketiga kelas, dan hanya membuat prediksi tunggal untuk semua input. Hal ini bisa disebabkan oleh ketidakseimbangan data, parameter model yang tidak tepat, atau kurangnya fitur yang informatif.

Tabel 3. Hasil Klasifikasi Tiap Algoritma

Metode	Accuracy	Precision	Recall	F1 Score
Naïve Bayes	73%	79%	79%	75%
Support Vector Machine	72%	52%	72%	60%

3.6 Hasil Sentimen

Hasil analisis sentimen dapat diketahui melalui tools *test and score*, yang akan menampilkan performa analisis sentimen dalam bentuk presentase, yaitu

accuracy, *precision*, dan *recall* seperti yang disajikan pada tabel 3.

Berdasarkan persentase dari klasifikasi pada Tabel 2. menunjukkan pada algoritma *Naïve Bayes* menghasilkan nilai *Accuracy* adalah 73%, *Precision* adalah 79% , *Recall* adalah 79%, dan *F1 Score* adalah 75%. Sedangkan pada algoritma *Support Vector Machine* menghasilkan nilai *Accuracy* adalah 72%, *Precision* adalah 52% , *Recall* adalah 72%, dan *F1 Score* adalah 60%. Berdasarkan nilai dari akurasi yang sudah disebutkan, dari kedua algoritma yang digunakan untuk analisis produk *Scarlett Whitening* dengan EXO, diperoleh bahwa *Naïve Bayes* lebih baik dan optimal dibandingkan dengan *Support Vector Machine*.

5. Kesimpulan

Kesimpulan dari penelitian ini adalah dari analisis sentimen terhadap ulasan komentar di channel YouTube resmi EXO mengenai produk *Scarlett Whitening*, yang melibatkan tahapan crawling data, pre-processing, analisis sentimen, klasifikasi model, evaluasi menggunakan confusion matrix, menunjukkan bahwa mayoritas ulasan bersifat positif dengan jumlah ulasan positif mencapai 869, diikuti ulasan netral sebanyak 137, dan ulasan negatif hanya 179. Hasil ini menunjukkan bahwa masyarakat Indonesia cenderung memberikan respons positif terhadap produk *Scarlett Whitening* yang dipromosikan oleh EXO sebagai Brand Ambassador. Analisis sentimen yang dilakukan, menunjukan algoritma *Naïve Bayes* memiliki performa yang lebih baik dibandingkan dengan *Support Vector Machine*, sehingga algoritma *Naïve Bayes* lebih cocok digunakan untuk melakukan analisis sentimen terhadap produk *Scarlett Whitening* dengan EXO sebagai Brand Ambassador. Selain itu, hasil analisis penelitian menunjukkan bahwa strategi pemasaran yang melibatkan EXO sebagai Brand Ambassador berhasil diterima dengan baik oleh masyarakat Indonesia dan mampu meningkatkan penjualan produk *Scarlett Whitening*.

Daftar Pustaka

- Abdulloh, F. F., & Pambudi, I. R. (2021). Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19. *CSRID (Computer Science Research and Its Development Journal)*, 13(3), 141–148.
- Asrianto, R., & Herwinanda, M. (2022). Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 431–440.
- Brestilliani, L. (2020). PENGARUH BRAND AWARENESS , BRAND AMBASSADOR , DAN HARGA TERHADAP KEPUTUSAN PEMBELIAN ONLINE PADA MARKETPLACE SHOPEE (Studi Pada Mahasiswa Stiesia). *Jurnal Ilmu Dan Riset*

- Manajemen*, 9(2461–0593).
- Fauzan, M., & Aisyah, S. (2023). Strategi Pemasaran Digital pada Industri *Skincare* melalui Platform TikTok. *Jurnal Ekonomi Dan Bisnis*, 1(6), 977–982.
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan *Support Vector Machine* pada Analisis Sentimen Twitter. *SMATIKA Jurnal Volume*, 10(2087–0256), 71–76.
- Gusriani, S., Wardhani, K. D. K., & Zul, M. I. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka). *4th Applied Business and Engineering Conference*, 1(1), 1–7.
- Hurley, C., Chen, S., & Umumnya, K. (2016). YOUTUBE SEBAGAI SARANA KOMUNIKASI BAGI KOMUNITAS MAKASSARVIDGRAM. *Jurnal Komunikasi KAREBA*, 5, 259–272.
- Indrawan, I. G. A., Ayu, D., Cahya, I., Ayu, I., & Ananda, P. (2023). Analisis Sentimen Terhadap Presidensi G20 2022 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes. *Kajian Ilmiah Informatika Dan Komputer*, 4(1), 553–561.
- Info, A., Gestasional, D., & Handoko, A. C. (2023). Perbandingan metode supervised learning untuk prediksi diabetes gestasional. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(4), 1238–1247.
- Mardiana, L., Kusnandar, D., & Satyahadewi, N. (2022). ANALISIS DISKRIMINAN DENGAN K FOLD CROSS VALIDATION UNTUK KLASIFIKASI KUALITAS AIR DI KOTA PONTIANAK. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, 11(1), 97–102.
- Mas, J. K. H., & Mansyur, N. (2023). Optimalisasi Strategi Promosi Digital di Era Digitalisasi : Sebuah Kajian Komprehensif pada PT Glamourix AFA Indonesia *Skincare* 2022-2023. *Journal of Emerging Business Management and Entrepreneurship Studies*, 3, 119–134.
- Mutawalli, L., Taufan, M., Zaen, A., Bagye, W., Sains, S. D., Lombok, S., Basuki, J., No, R., & Lombok, P. (2019). KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *Jurnal Informatika & Rekayasa Elektronika*, 2(2), 43–51.
- Nitamia, M. T., & Februariyanti, H. (2022). Analisis Sentimen Ulasan Ekpedisi J&T Ekspres Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, 5(1), 20–29.
- Nurhadi, Z. F., Salamah, U., Yuniar, A. A., Studi, P., Komunikasi, I., & Garut, U. (2020). MOTIF PENGGUNAAN YOUTUBE SEBAGAI MEDIA INFORMASI KECANTIKAN GENERASI MILLENIAL. *Jurnal Komunikasi Dan Media*, 4(2527–8673), 170–190.
- Puspita, A. D., Desrianti, D., & Permana, E. (n.d.). STRATEGI PEMASARAN DIGITAL SCARLETT WHITENING. *Jurnal Riset Pendidikan Ekonomi (JRPE)*, 8(2540–9247), 180–187.
- Putri, A. J., Syafira, A. S., & Purbaya, M. E. (2022). Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine. *Jurnal Trinistik*, 2829–4971(1), 16–21.
- Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi Algoritma Naïve Bayes Dan *Support Vector Machine* (SVM) Pada Analisis Sentimen Spotify. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 349–354.
- Rambe, T. S., Hasibuan, M. N. S., & Dar, M. H. (2023). Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method. *Sinkron*, 8(2), 980–989. <https://doi.org/10.33395/sinkron.v8i2.12303>
- Sari, D. N., & Feranita, N. V. (2022). Pengaruh Inovasi Produk Dan Citra Merek Terhadap Keputusan Pembelian Konsumen. *Ilmu Manajemen*, 10(2), 233–243.
- Sari, D. N., Sari, D. N., Adelia, F., Rosdiana, F., Butar, B. B., & Hariyanto, M. (2020). Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier. *JIKA (Jurnal Informatika)*, 4(3), 109. <https://doi.org/10.31000/jika.v4i3.3086>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.
- Suryaningsih, S., & Ningtias, A. (2023). Komunikasi Pemasaran Digital Produk *Skincare Scarlett Whitening* dengan Brand Ambassador EXO. *Bandung Conference Series: Communication Management*, 1074–1082.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67.
- Tamonsang, Matheous; Putri, Y. D. A. (2023). Pengaruh Celebrity Endorser dan Electronic Word Of Mouth Terhadap Minat Beli Dengan Brand Image. *Reviu Akuntansi Komputer Indonesia*, 4(2), 89–101.
- Tauhid, S. M., & Ruldeviyani, Y. (2020). Sentiment Analysis of Indonesians Response to Influencer in Social Media. *7th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2020 - Proceedings*, 1, 90–95. <https://doi.org/10.1109/ICITACEE50144.2020.9239218>
- Yusuf, M. (2021). Pengaruh Promosi dan Brand Ambassador terhadap Keputusan Pembelian Pengguna Market Place X dengan Brand Image sebagai Variabel Intervening. *Jurnal Bisnis, Manajemen, Dan Keuangan*, 2(2), 445–458.
- Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*,

5(2), 422. <https://doi.org/10.30865/mib.v5i2.2845>

Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021).

Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 2(1), 55–63.