

Sentimen Analisis Overclaim Skincare Skintific Menggunakan Algoritma BERT

Kharlina Hingis¹, Umi Karolina² and Rizky Pribadi^{3*}

¹ Universitas Multi Data Palembang; kharlinahingis@mhs.mdp.ac.id

² Universitas Multi Data Palembang; umikarolina@mhs.mdp.ac.id

³ Universitas Multi Data Palembang; rizky@mdp.ac.id

* Universitas Multi Data Palembang; rizky@mdp.ac.id

Info Artikel:

Dikirim: 22 Desember 2024

Direvisi: 21 Desember 2024

Diterima: 27 Desember 2024

Abstract: This year, the use of Skintific skincare products has experienced a significant surge in sales percentages. Based on observations in the beauty and cosmetics market, Skintific has recorded revenues reaching billions to trillions of rupiahs annually. This indicates that business opportunities in the beauty industry are highly promising, especially in recent years. However, intense competition in the beauty industry has given rise to numerous new cosmetic brands that often promote their products excessively, a practice known as overclaim. This behavior is often carried out unconsciously, yet it misleads targeted consumers and constitutes fraudulent practices. In this study, the author conducted a scraping process of customer reviews on Skintific cosmetics from Instagram, one of the popular social media platforms. The approach used was a Deep Learning method with the BERT (Bidirectional Encoder Representations from Transformers) algorithm. The research analyzed 1,000 customer reviews from the official Instagram account of Skintific. The findings reveal that the phenomenon of overclaim in Skintific products was only 0.0015% of the total reviews, which is starkly lower compared to other cosmetic products in the beauty industry.

Keyword: Deep Learning, BERT Algorithm, Beauty Industry, Instagram Reviews Analysis.

Intisari: Pada tahun ini, penggunaan skincare merek Skintific mengalami lonjakan signifikan dalam persentase penjualannya. Berdasarkan pengamatan di pasar kecantikan dan kosmetik, Skintific mencatat pendapatan yang mencapai miliaran hingga triliunan rupiah setiap tahunnya. Hal ini menunjukkan bahwa peluang bisnis di industri kecantikan sangat menjanjikan, terutama dalam beberapa tahun terakhir. Namun, persaingan yang ketat di industri kecantikan telah memunculkan berbagai merek kosmetik baru yang seringkali mempromosikan produk mereka secara berlebihan, atau yang dikenal dengan istilah overclaim. Praktik ini tanpa disadari merupakan tindakan yang merugikan konsumen, karena mengandung unsur penipuan terhadap target pasar. Dalam penelitian ini, penulis melakukan proses scraping ulasan pelanggan terhadap produk kosmetik Skintific di media sosial Instagram. Pendekatan yang digunakan adalah metode Deep Learning dengan algoritma BERT (Bidirectional Encoder Representations from Transformers). Penelitian ini berhasil menganalisis 1.000 ulasan pelanggan yang diambil dari akun resmi Instagram Skintific. Hasil analisis menunjukkan bahwa fenomena overclaim pada produk Skintific hanya ditemukan sebesar 0,0015% dari total ulasan, yang sangat berbanding terbalik dengan produk-produk kosmetik lainnya di industri kecantikan.

Kata Kunci: Deep Learning, Algoritma BERT, Industri Kecantikan, Analisis Ulasan Instagram.

1. Pendahuluan

Dalam era di mana konten digital menjadi semakin melimpah, tantangan utama yang dihadapi adalah bagaimana mengidentifikasi dan mengelola konten yang sensitif secara efisien. Konten sensitif dapat berupa materi yang menyinggung secara sosial, mencakup mulai dari diskriminasi hingga kekerasan, dan bisa memiliki dampak yang merugikan jika tidak dikelola dengan baik. Semakin berkembangnya zaman masyarakat mulai mengenal internet marketing yang menjadi salah satu cara meluaskan berbagai macam informasi mengenai skincare. Internet marketing atau sering juga disebut dengan digital marketing adalah sekumpulan kegiatan pemasaran yang menggunakan teknologi digital, yaitu Internet sebagai alat pemasaran (e-marketing). Pertumbuhan pemasaran digital di Indonesia dinilai sangat cepat, yang juga tercermin dari peningkatan aplikasi yang menjadi media massa Informasi, media massa untuk pelaksanaan pemasaran produk sampai dengan selesainya penjualan. Hal ini menunjukkan bahwa industri kecantikan menjadi salah satu sektor bisnis yang sangat menguntungkan [1]. Di sinilah pentingnya pengembangan model Deep Learning yang canggih untuk mendeteksi konten-konten tersebut.

Teks Komentar dapat dikategorikan menjadi komentar positif atau komentar negatif ada pun kita sendiri bisa membedakan beberapa komentar yang menurut kita baik maupun buruk secara tak langsung, tetapi jika kita melakukan pencarian teks komentar itu menggunakan terkomputerisasi akan membawa permasalahan baru itu tersendiri. Begitu pula dengan mencari konten sensitif di sebuah teks komentar akan dapat dengan mudahnya memisahkan komentar positif, negatif, dan netral.

Dalam konteks ini, algoritma BERT menonjol sebagai alat yang efektif untuk memahami dan menganalisis teks dengan lebih baik. BERT, varian dari model BERT yang dioptimalkan khusus untuk mendeteksi konten-konten sensitif dalam bahasa, telah menunjukkan keunggulannya dalam berbagai aplikasi NLP (Natural Language Processing).

Dalam jurnal ini, kami mengusulkan dan mengevaluasi model Deep Learning yang didasarkan pada algoritma BERT untuk mendeteksi konten sensitif dalam teks komentar Over Klaim Skincare. Kami menjelajahi kemampuan model untuk mengenali dan mengklasifikasikan konten yang sensitif dengan akurat, memberikan solusi yang dapat diandalkan untuk mengelola konten-konten yang berpotensi merugikan di berbagai platform digital.

Melalui penelitian ini, kami bertujuan untuk tidak hanya meningkatkan kemampuan deteksi konten sensitif, tetapi juga untuk memberikan kontribusi pada pengembangan model Deep Learning yang lebih adaptif dan responsif terhadap perubahan dalam bahasa dan konteks sosial yang terus berkembang. Dengan demikian, kami berharap penelitian ini dapat menjadi kontribusi yang berharga dalam menjaga lingkungan digital yang lebih aman dan inklusif bagi semua penggunanya.

2. Tinjauan Pustaka

Dengan Algoritma BERT Bagian ini memuat teori-teori pendukung dari metode yang diusulkan untuk pemecahan suatu masalah dan pengembangan dari metode tersebut, yang didasarkan dari referensi yang jelas (buku, jurnal, prosiding dan artikel ilmiah lainnya).

2.1 Deep Learning

Dengan menggunakan artificial neural network yang berlapis-lapis, atau multilayer, deep learning adalah teknik pembelajaran. Artificial Neural Network ini mirip dengan otak manusia karena jaringan neuronnya sangat rumit. Deep learning, juga dikenal sebagai deep structured learning, hierarchical learning, atau deep neural, adalah teknik pembelajaran yang menggunakan berbagai transformasi non-linier. Deep learning dapat dianggap sebagai kombinasi pembelajaran mesin dengan AI (artificial neural network) [2].

2.2 Algoritma Bert

Metode pemodelan topik BERT menggunakan embedding BERT dan c-TF-IDF untuk membuat cluster yang padat yang memungkinkan tafsiran mudah sambil mempertahankan kata-kata penting dalam deskripsi[3]. Metode ini melakukan tiga langkah, yaitu membuat dokumen embedding, mengubah cluster ke dalam bentuk yang serupa secara semantic, dan kemudian membuat representasi topik dari masing-masing cluster. Pada penelitian ini, penulis menggunakan paket BERT, yang merupakan open-source python, untuk melakukan pemodelan topic menggunakan metode BERT [4]. Paket OCTIS (Optimizing and Comparing Topic Models is Simple), yang merupakan open-source python, digunakan untuk menjalankan eksperimen, memvalidasi hasil, dan memproses data sebelumnya[5].

2.3 Data Mining

Menurut Gartner Group, Data mining adalah proses penemuan makna, pola, dan pola baru dengan memilih jumlah data yang disimpan pada media penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik komputer yaitu matematika dan statistika. Penambangan data adalah kombinasi dari berbagai disiplin ilmu yang menggabungkan desain dan pembelajaran mesin, basis data, statistik, pengenalan pola, dan

proses visualisasi untuk memecahkan masalah pengambilan informasi dari basis data besar. Pada fase ini, proses data mining terdiri dari serangkaian proses, dimana proses ini berfungsi untuk memeriksa nilai dari setiap kumpulan data ketika data tersebut masih berupa informasi yang tidak dapat dipahami oleh manusia secara manual karena masih ada bahasa mesin [6].

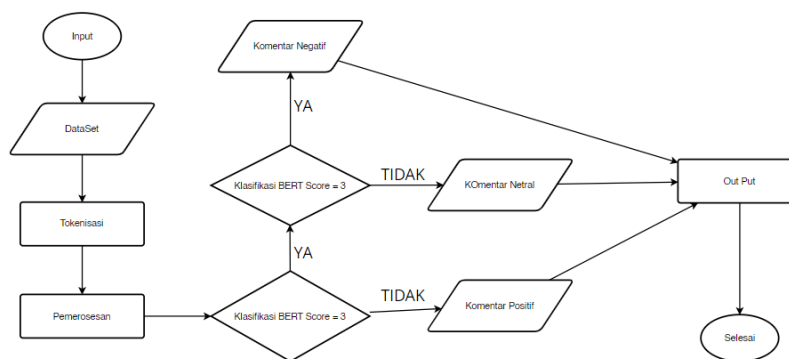
2.4 Data Set

Pada penelitian ini penulis mendapatkan dataset menggunakan apify.com. Apify adalah sebuah software yang sering digunakan untuk mescraper dataset yang di inginkan di sini penulis menggunakan Apify Instagram Comments Scraper untuk menarik dataset yang di butuhkan.

3. Metode dan Hasil Pembahasan

3.1 Metode Penelitian

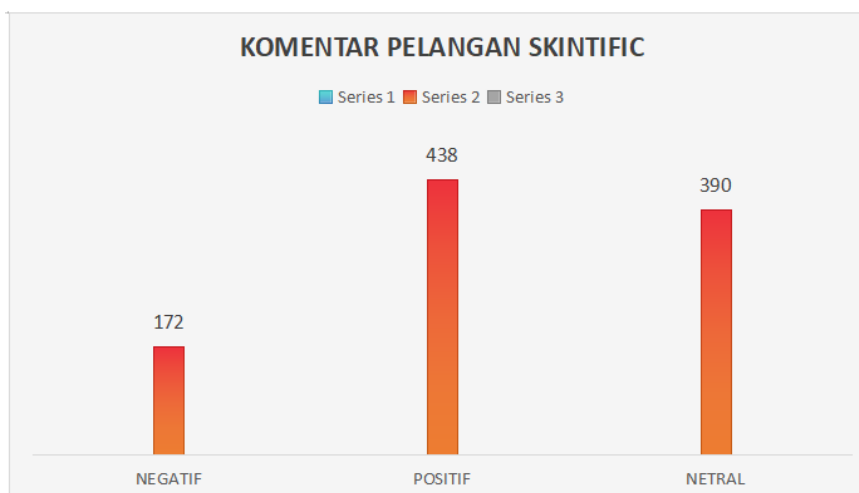
Penulis menggunakan pendekatan *Deep Learning* dengan menggunakan Algoritma BERT [7] untuk mengklasifikasikan komentar sensitif. Data yang menjadi objek penelitian penulis adalah komentar produk skincare *scientific* di sosial media Instagram yang [8]. Cara pencarian yang dilakukan dalam komentar teks atau data teks untuk mendapatkan informasi yang diinginkan pada rana text mining sehingga perlu dilakukan pemrosesan terlebih dahulu untuk mempersiapkan data agar siap dikelola [9]. pada setiap step dapat mempermudah dalam eksekusi data yang diinginkan. dilanjutkan dengan memasukan data dalam topik modeling berikut proses pada flowchart:



Gambar 1. diagram desain sistem

3.2 Data Preparation

Data yang akan kita uji dengan cara mengambil komentar Skincare skintifick di sosial media Instagram. pengumpulan data diambil dari tahun 2023 sampai 2024. data yang telah terkumpul akan dibedakan menjadi Negatif, Positif, dan Netral dari pelabelan manual yang kita lakukan. berikut gambaran dari distribusi dan informasi komentar di Instagram



Gambar 2. Komentar Produk Skintific

3.3 Pre-Processing

Pembersihan Data: Tahap ini melibatkan identifikasi serta penanganan data yang bersifat Null, outlier, atau tidak [10]. Data yang bersifat Null dapat dihapus atau akan diisi dengan nilai yang sesuai yang kita butuhkan, Pada tahap ini juga akan dilakukan penghapusan data yang memiliki duplikasi.

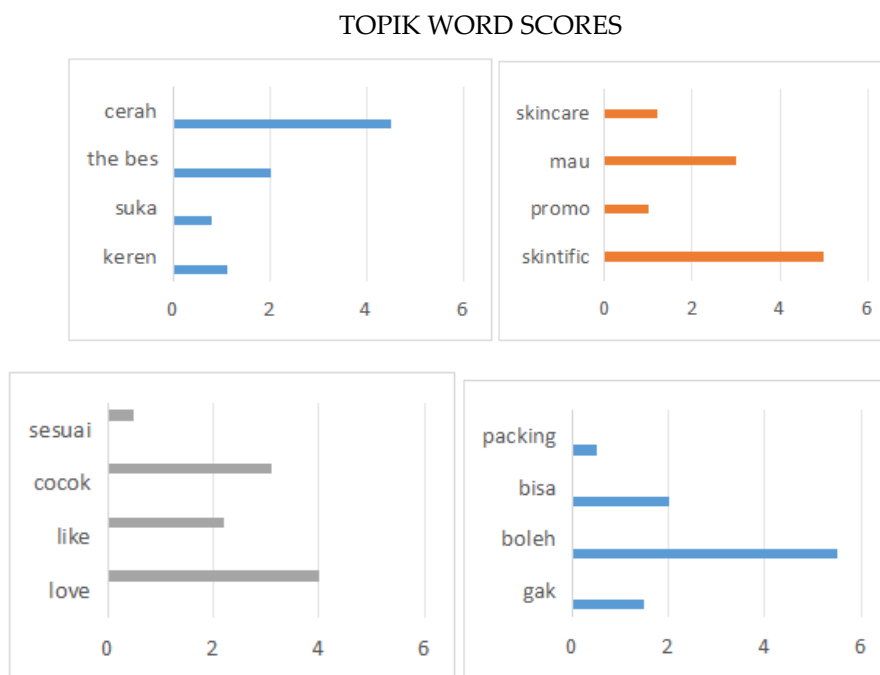
Transformasi Data: dalam dataset mungkin perlu diubah skala atau bentuk distribusinya untuk memenuhi persyaratan analisis atau pemodelan, Normalisasi dan standarisasi ialah teknik yang umum digunakan untuk mengubah skala atribut. Selain itu, transformasi logaritmik dapat digunakan untuk mengubah atribut yang cenderung mengikuti pola eksponensial.

Seleksi Fitur: Tahap ini melibatkan identifikasi atribut yang paling relevan untuk bertujuan menganalisis serta fitur yang baik dapat meningkatkan efisiensi dan kualitas model. Atribut yang tidak relevan atau memiliki korelasi yang tinggi.

Pengurangan Dimensi: pada Tahap ini bertujuan untuk mengurangi kompleksitas data dengan mengurangi jumlah atribut yang digunakan dalam analisis Teknik yang umum digunakan termasuk Analisis Komponen untuk data numerik dan Analisis Diskriminan Linier pemodelan kelas [11].

3.4 Konfigurasi Metode Queue Tree Pada MikroTik

Dalam hasil pembahasan ini Dataset yang penulis gunakan yaitu hasil review dari produk skincare scientific yang didapatkan dari scraping media sosial Instagram Official Skintific. terdapat 1 kategori dalam dataset ini yaitu teks komentar atau review produk scientific. dan penilaian dari komentar berdasarkan ahli bahasa yang mengidentifikasi review negatif, positif, dan netral. selanjutnya melakukan proses SMOTE, kali ini mengambil topik dari komentar teks review skincare scientific. Pada langkah sebelumnya, penulis dapat membuat visualisasi seperti LDA untuk memahami setiap topik. Atau, penulis juga dapat memvisualisasikan istilah yang dipilih untuk beberapa topik dengan membuat diagram batang dari skor c-TF-IDF untuk setiap representasi topik. Bagan batang akan memberi penulis kata kunci paling representatif teratas untuk setiap topik. Misalnya, jika penulis merujuk ke Topik 1 di bawah ini, dengan kata kunci: glow, keren, suka, cukup masuk akal untuk mengatakan bahwa topik ini cenderung memiliki ulasan yang terkait dengan "terkait layanan". Dari sini, penulis dapat memiliki pemahaman yang baik tentang setiap topik. Visualisasi ini juga berguna bagi penulis untuk dengan mudah membandingkan topik satu sama lain. Untuk memvisualisasikan Hasil topik yang diambil dapat dilihat pada gambar 3 di bawah ini



Gambar 3. Topik Word Score

4. Kesimpulan

Pada penelitian ini penulis menggunakan metode Deep Learning untuk menyeimbangkan dataset scraping review dari instagram dan topic modelling BERT untuk mengelompokkan dataset review. Berdasarkan hasil dan pembahasan penelitian didapatkan empat topik yang mana topik tersebut yang berlabel positif pada hasil pemakaian produk skincare dari skintific. Selain dari itu tidak ditemukan topik dengan kondisi pembeli menulis review untuk membuktikan keunggulan dan hasil dari pemakaian produk yang diklaim oleh produk skincare tersebut. Maka didapatkan hasil penelitian dari Instagram yang berjumlah 1000, Dimana jumlah suara positif mendominasi dan menunjukkan bahwa skintific merupakan skincare terpercaya dalam keaslian hasil pada pemakaiannya.

Daftar Pustaka

- [1] N. Nadya, "Peran Digital Marketing Dalam Eksistensi Bisnis Kuliner Seblak Jeletet Murni," JRMB (Jurnal Riset Manajemen Dan Bisnis), Fakultas Ekonomi UNIAT, Oct. 2016, doi:10.36226/jrmb.v1i2.17.
- [2] Bengio, Y.; Simard, P.; Frasconi, P. IEEE transactions on neural networks, 5, 157–166,1994.
- [3] Primartha, Rifkie. Belajar Machine Learning Teori Dan Praktik. Palembang, 2018. 1.
- [4] Grootendorst, M., "BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure" arXiv preprintarXiv:2203.05794, 2022.
- [5] Grootendorst, M., "BERTopic", 11 September 2022. [Online]. Available:https://github.com/MaartenGr/BERTopic.
- [6] Terragni, S., Fersini, E., Galuzzi, B. G., Tropeano, P., dan Candelieri, A., "OCTIS: Comparing and Optimizing Topic Models is Simple!" in Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, 2021, p. 263–270. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [7] Kurniawan, Bayu, A. Ari Aldino, and A. Rahman Isnain. "Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (Pse) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (Bert)." J. Teknol. dan Sist. Inf 3.4 (2022): 98-106.
- [8] Nayla, A., Setianingsih, C., & Dirgantoro, B. (2023). Deteksi Hate Speech Pada Twitter Menggunakan Algoritma BERT. *eProceedings of Engineering*, 10(1).
- [9] Junita, E. A., & Suryono, R. R. (2024). Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(4), 2042-2053.
- [10] Kusuma, R. M. R. W. P., & Yustanti, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 2(3).
- [11] Afifa, N., Saputra, R. E., & Nugrahaeni, R. A. (2023). Implementasi NLP Pada Chatbot Layanan Akademik Dengan Algoritma Bert. *eProceedings of Engineering*, 10(1).