Journal of Information Technology and Computer Science (INTECOMS)
Volume 9 Nomor 5, Tahun 2026
e-ISSN : 2614-1574
p-ISSN : 2621-3249

PENGUKURAN BRAND STRENGTH LAPTOP AXIOO DAN ADVAN MENGGUNAKAN TEXT MINING DAN SEMANTIC RETRIEVAL

MEASURING AXIOO AND ADVAN LAPTOP BRAND STRENGTH USING TEXT MINING AND SEMANTIC RETRIEVAL

Widi Anggraini[1], Lukman Sunardi[2], Cindi Wulandari[3], Budi Santoso[4]
Program Studi Informatika, Universitas Bina Insan[1,2,3,4]
widi24522@gmail.com[1], lukmanmmci@gmail.com[2],
cindi_wulandari@univbinainsan.ac.id[3], budisantoso@univbinainsan.ac.id[4]

ABSTRACT

This study aims to analyze user perceptions of Advan and Axioo laptops through YouTube comments using text mining, sentiment analysis, and embedding-based semantic retrieval. Data were collected from 21 videos, producing 3,519 comments, and 3,417 unique comments remained after duplicate removal. Initial sentiment labeling used TextBlob, while classification used TF-IDF and Multinomial Naive Bayes. The dataset was split using a stratified 80:20 procedure, and Random Oversampling was applied only to the training data to prevent data leakage. The evaluation produced an accuracy of 0.730, a weighted F1-score of 0.782, and a macro F1-score of 0.458. The label distribution consisted of 250 positive, 3,050 neutral, and 117 negative comments. The sentiment-based brand strength index was 0.031429 for Advan and 0.043860 for Axioo. Semantic retrieval was used to identify comments relevant to a query.

***Keywords**: Brand strength, TF-IDF, Multinomial Naive Bayes, YouTube, semantic retrieval.*

ABSTRAK

Penelitian ini bertujuan menganalisis persepsi pengguna terhadap laptop Advan dan Axioo melalui komentar YouTube menggunakan text mining, analisis sentimen, dan semantic retrieval berbasis embedding. Data dikumpulkan dari 21 video dan menghasilkan 3.519 komentar, kemudian diperoleh 3.417 komentar unik setelah penghapusan duplikat. Pelabelan sentimen awal menggunakan TextBlob, sedangkan klasifikasi dilakukan dengan TF-IDF dan Multinomial Naive Bayes. Pembagian data dilakukan secara stratified 80:20 dan Random Oversampling hanya diterapkan pada data latih untuk mencegah data leakage. Hasil pengujian menghasilkan accuracy 0,730, weighted F1-score 0,782, dan macro F1-score 0,458. Distribusi label terdiri atas 250 positif, 3.050 netral, dan 117 negatif. Indeks brand strength berbasis sentimen sebesar 0,031429 untuk Advan dan 0,043860 untuk Axioo. Semantic retrieval digunakan untuk menemukan komentar yang relevan terhadap query.

Kata kunci: Brand strength, TF-IDF, Multinomial Naive Bayes, YouTube, semantic retrieval.

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah cara konsumen mencari informasi, mengevaluasi produk, serta menyampaikan pengalaman setelah menggunakan suatu produk. Media sosial dan berbagai platform digital memungkinkan konsumen menghasilkan konten berupa ulasan, komentar, opini, pengalaman, dan keluhan yang dapat menjadi sumber informasi mengenai persepsi konsumen. Konten yang dihasilkan pengguna tersebut juga dapat memengaruhi konsumen lain dalam proses pencarian informasi dan pengambilan keputusan pembelian. Namun, jumlah data yang besar dan karakteristiknya yang tidak terstruktur menyebabkan analisis secara manual menjadi kurang efisien. Oleh karena itu, diperlukan pendekatan komputasional seperti *sentiment analysis* dan *opinion mining* untuk mengekstraksi informasi dari data tekstual tersebut [1]. Penelitian mengenai pemanfaatan data media sosial dalam pengambilan keputusan bisnis juga menunjukkan bahwa *sentiment*

analysis dan *opinion mining* dapat digunakan untuk memahami serta menganalisis perilaku dan opini konsumen [2].

Sentiment analysis merupakan pendekatan dalam pengolahan bahasa alami yang digunakan untuk mengidentifikasi kecenderungan opini atau sentimen yang terdapat dalam suatu teks. Dalam konteks data konsumen, pendekatan ini dapat digunakan untuk mengelompokkan opini ke dalam kategori seperti positif, negatif, dan netral sehingga kumpulan komentar yang besar dapat diringkas menjadi informasi yang lebih mudah dianalisis. Pemanfaatan *sentiment analysis* terhadap data pelanggan juga memungkinkan organisasi memperoleh informasi mengenai tingkat kepuasan, ketidakpuasan, serta aspek produk atau layanan yang sering dibicarakan oleh konsumen [3].

Salah satu sumber data digital yang menarik untuk analisis opini adalah YouTube. Selain berfungsi sebagai platform berbagi video, YouTube menyediakan ruang bagi pengguna untuk memberikan komentar secara langsung terhadap konten yang ditonton. Komentar tersebut dapat mengandung tanggapan, penilaian, pengalaman, maupun opini pengguna terhadap topik atau produk yang dibahas dalam video. Penelitian terdahulu menunjukkan bahwa komentar YouTube dapat dikumpulkan dan dianalisis menggunakan pendekatan *sentiment analysis* untuk memperoleh gambaran mengenai respons pengguna [4]. Penelitian lain juga menunjukkan bahwa komentar pada video YouTube dapat diproses melalui tahapan *preprocessing* dan klasifikasi untuk mengidentifikasi distribusi sentimen pengguna [3].

Pemanfaatan komentar YouTube menjadi semakin relevan ketika video yang dianalisis membahas produk tertentu. Dalam konteks produk teknologi, komentar pengguna dapat memberikan informasi tambahan mengenai bagaimana suatu produk dipersepsikan oleh konsumen. Penelitian terhadap komentar pada video ulasan produk teknologi menunjukkan bahwa data YouTube dapat digunakan untuk mengidentifikasi pola tanggapan pengguna dan menganalisis sentimen yang muncul dalam diskusi mengenai produk [5]. Dengan demikian, komentar pada video ulasan laptop dapat dipandang sebagai sumber *user-generated content* yang berpotensi memberikan informasi mengenai persepsi pengguna terhadap produk.

Namun, komentar YouTube memiliki karakteristik bahasa yang berbeda dari teks formal. Pengguna dapat menggunakan singkatan, *slang*, variasi ejaan, kata tidak baku, serta bentuk bahasa informal lainnya. Karakteristik tersebut dapat menyulitkan proses klasifikasi karena teks yang digunakan pengguna tidak selalu mengikuti struktur bahasa yang baku. Penelitian mengenai analisis bahasa pada media sosial menunjukkan bahwa teks media sosial cenderung tidak mengikuti standar tata bahasa dan perlu melalui proses pembersihan serta penataan sebelum digunakan dalam *opinion mining* [6]. Selain itu, penelitian mengenai *preprocessing* pada *sentiment analysis* menunjukkan bahwa pilihan tahapan preprocessing dapat memengaruhi kualitas hasil klasifikasi, khususnya ketika data berasal dari media sosial yang memiliki karakteristik bahasa informal . Oleh karena itu, diperlukan tahapan *preprocessing* yang sesuai dengan karakteristik data sebelum komentar digunakan dalam proses klasifikasi.

Dalam industri laptop, persepsi konsumen menjadi penting karena produk laptop memiliki berbagai atribut yang dapat menjadi pertimbangan dalam keputusan pembelian. Konsumen tidak hanya mempertimbangkan aspek harga, tetapi juga dapat memperhatikan kualitas produk, performa, spesifikasi, desain, dan pengalaman penggunaan. Dari perspektif merek, penelitian pada konsumen produk laptop menunjukkan bahwa dimensi *brand awareness*, *perceived quality*, *brand association*, dan *brand loyalty* merupakan bagian penting dalam pembentukan

customer-based brand equity [7]. Penelitian lain juga menunjukkan bahwa *brand awareness*, *brand association*, *perceived quality*, dan *brand loyalty* dapat dipandang sebagai dimensi dalam pengukuran *customer-based brand equity*.

Berdasarkan konteks tersebut, Advan dan Axioo dipilih sebagai objek penelitian karena kedua merek tersebut ditemukan dalam percakapan pengguna pada video YouTube yang menjadi sumber data penelitian. Komentar mengenai kedua merek tersebut kemudian digunakan untuk memperoleh gambaran mengenai kecenderungan sentimen pengguna. Pemilihan Advan dan Axioo dalam penelitian ini tidak dimaksudkan untuk menyatakan bahwa salah satu merek secara umum lebih kuat daripada merek lainnya, tetapi untuk membandingkan kecenderungan opini yang terdapat pada dataset komentar yang dikumpulkan.

Konsep *brand strength* dalam penelitian ini juga perlu dibatasi agar tidak disamakan dengan pengukuran *brand equity* secara menyeluruh. Literatur mengenai *customer-based brand equity* menunjukkan bahwa kekuatan suatu merek dapat dikaji melalui beberapa dimensi, termasuk *brand awareness*, *brand association*, *perceived quality*, dan *brand loyalty*. Penelitian ini tidak mengukur seluruh dimensi tersebut. Sebaliknya, *brand strength* dioperasionalkan secara khusus sebagai indeks berbasis sentimen, yaitu selisih antara jumlah komentar positif dan komentar negatif yang dinormalisasi terhadap jumlah komentar masing-masing merek. Dengan demikian, indeks tersebut merupakan indikator komparatif yang dibangun berdasarkan karakteristik dataset penelitian dan bukan merupakan ukuran universal atau pengganti pengukuran *brand equity*.

Untuk mengolah komentar dalam jumlah besar, diperlukan pendekatan *text mining* yang dapat mengubah data teks menjadi representasi yang dapat diproses oleh algoritma pembelajaran mesin. Tahapan *preprocessing* merupakan bagian penting karena data teks dari media sosial dapat mengandung karakter atau informasi yang tidak relevan seperti URL, tanda baca, kata tidak baku, singkatan, dan kesalahan penulisan. Proses pembersihan dan normalisasi diperlukan agar teks menjadi lebih sesuai untuk tahap ekstraksi fitur dan klasifikasi [8]. Berdasarkan karakteristik tersebut, penelitian ini menerapkan beberapa tahapan *preprocessing*, yaitu *case folding*, pembersihan karakter, normalisasi kata, *stopword removal*, penghapusan kata pendek, dan *stemming*.

Setelah proses *preprocessing*, teks perlu diubah menjadi representasi numerik agar dapat digunakan oleh algoritma klasifikasi. Salah satu metode yang umum digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF memberikan bobot terhadap kata berdasarkan kemunculannya dalam suatu dokumen sekaligus mempertimbangkan seberapa umum kata tersebut muncul dalam keseluruhan kumpulan dokumen. Representasi TF-IDF telah digunakan bersama algoritma *Naive Bayes* dalam berbagai penelitian *sentiment analysis* berbahasa Indonesia [9]. Penelitian lain juga menunjukkan bahwa TF-IDF dapat digunakan sebagai representasi fitur untuk berbagai varian *Naive Bayes* pada data sentimen berbahasa Indonesia [10]

Dalam penelitian ini, Multinomial Naive Bayes digunakan sebagai algoritma klasifikasi sentimen. Algoritma tersebut sesuai untuk tugas klasifikasi teks karena bekerja berdasarkan distribusi fitur kata dalam dokumen. Penggunaan kombinasi TF-IDF dan Multinomial Naive Bayes juga telah diterapkan pada data komentar YouTube berbahasa Indonesia [11], misalnya, menggunakan TF-IDF dan Multinomial Naive Bayes untuk mengklasifikasikan sentimen pada komentar YouTube dan membandingkannya dengan Support Vector Machine [11]. Penelitian tersebut menunjukkan bahwa metode pembelajaran mesin klasik masih dapat digunakan untuk melakukan klasifikasi sentimen pada data

komentar YouTube, meskipun performanya dapat berbeda bergantung pada karakteristik dataset dan model yang digunakan.

Permasalahan lain yang perlu diperhatikan adalah ketidakseimbangan kelas. Ketika jumlah data pada satu kelas jauh lebih besar dibandingkan kelas lainnya, algoritma klasifikasi dapat lebih cenderung memprediksi kelas mayoritas sehingga kemampuan model dalam mengenali kelas minoritas dapat menurun [12].Kondisi tersebut relevan dalam penelitian ini karena distribusi label awal menunjukkan bahwa komentar netral lebih banyak dibandingkan komentar positif dan negatif. Untuk mengurangi ketimpangan distribusi pada data latih, penelitian ini menggunakan Random Oversampling, yaitu menambahkan sampel dari kelas minoritas melalui proses pengambilan ulang sehingga distribusi kelas pada data latih menjadi lebih seimbang[12].

Penerapan *oversampling* juga harus memperhatikan kemungkinan terjadinya *data leakage*. Jika data diperbanyak sebelum proses pembagian *training* dan *testing*, sampel hasil duplikasi dapat memiliki representasi yang sama atau sangat mirip pada kedua bagian data sehingga evaluasi model menjadi terlalu optimistis. Oleh karena itu, penelitian ini melakukan train-test split terlebih dahulu, kemudian Random Oversampling hanya diterapkan pada data latih. Pendekatan yang memisahkan proses penanganan ketidakseimbangan dari data pengujian juga digunakan dalam penelitian *machine learning* yang menerapkan strategi *leakage-free* dengan melakukan *oversampling* hanya pada *training set* [13].

Selain klasifikasi sentimen, penelitian ini memanfaatkan semantic retrieval untuk menemukan komentar yang relevan terhadap suatu *query*. Berbeda dengan pencarian berbasis kata kunci yang terutama bergantung pada kesamaan istilah, *semantic retrieval* menggunakan representasi *embedding* untuk menangkap kedekatan makna antara *query* dan dokumen atau komentar. Salah satu pendekatan yang dapat digunakan adalah mengubah teks menjadi *sentence embedding* dan menghitung kemiripan antarvektor menggunakan *cosine similarity*. Penelitian menunjukkan bahwa *sentence embedding* berbasis transformer dapat digunakan untuk merepresentasikan teks sebagai vektor dan *cosine similarity* dapat digunakan untuk mengambil teks yang memiliki kemiripan semantik.

Pendekatan berbasis embedding juga telah digunakan dalam sistem *semantic search*, yaitu dengan mengubah *query* dan dokumen menjadi representasi vektor kemudian melakukan pemeringkatan berdasarkan *cosine similarity*[14].

Dalam penelitian ini, hasil *semantic retrieval* digunakan untuk menampilkan beberapa komentar yang memiliki kedekatan semantik dengan pertanyaan atau *query* pengguna. Sistem tersebut hanya melakukan proses representasi, pencarian, dan pemeringkatan komentar sehingga tidak menghasilkan jawaban baru melalui model generatif.

Atas dasar tersebut, sistem yang dikembangkan dalam penelitian ini tidak dikategorikan sebagai Retrieval-Augmented Generation (RAG). RAG pada dasarnya menggabungkan mekanisme *retrieval* dengan model generatif sehingga informasi yang diperoleh dari sumber eksternal digunakan sebagai konteks untuk menghasilkan respons [15]. Sementara itu, sistem dalam penelitian ini berhenti pada tahap *retrieval*, yaitu menemukan dan menampilkan komentar yang paling relevan berdasarkan kedekatan semantik. Pembatasan tersebut penting agar istilah dan kontribusi sistem yang dikembangkan sesuai dengan kemampuan aktual sistem.

Berdasarkan uraian tersebut, penelitian ini memiliki tiga tujuan utama. Pertama, mengidentifikasi distribusi sentimen pengguna terhadap laptop Advan dan Axioo berdasarkan komentar pada video YouTube. Kedua, membangun model klasifikasi sentimen menggunakan TF-IDF dan Multinomial Naive Bayes dengan

penanganan ketidakseimbangan kelas melalui Random Oversampling yang hanya diterapkan pada data latih. Ketiga, menghitung indeks *brand strength* berbasis sentimen sebagai indikator komparatif berdasarkan dataset serta mengimplementasikan *semantic retrieval* untuk menemukan komentar yang relevan terhadap suatu *query*. Dengan pendekatan tersebut, penelitian diharapkan dapat memberikan gambaran kuantitatif mengenai kecenderungan opini pengguna terhadap kedua merek sekaligus menunjukkan bagaimana metode *text mining*, klasifikasi sentimen, dan *semantic retrieval* dapat diterapkan pada komentar YouTube berbahasa Indonesia.

METODE

Penelitian ini menggunakan kerangka ***Cross-Industry Standard Process for Data Mining*** **(CRISP-DM)** sebagai alur umum dalam pelaksanaan penelitian. CRISP-DM menyediakan proses terstruktur yang terdiri atas enam tahapan utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Kerangka tersebut digunakan dalam penelitian ini karena dapat memberikan alur sistematis mulai dari perumusan kebutuhan analisis, pemahaman dan persiapan data, pembangunan model, evaluasi hasil, hingga penyajian atau pemanfaatan hasil analisis. Dalam penelitian ini, keenam tahapan CRISP-DM diadaptasi sesuai dengan kebutuhan analisis sentimen komentar YouTube. Tahap *business understanding* digunakan untuk merumuskan kebutuhan analisis sentimen dan perbandingan kecenderungan sentimen merek Advan dan Axioo. Tahap *data understanding* mencakup pengumpulan dan pemeriksaan data komentar YouTube. Tahap *data preparation* meliputi pembersihan data, *preprocessing*, pelabelan sentimen, dan pembagian data. Tahap *modeling* mencakup pembobotan TF-IDF, penyeimbangan data latih menggunakan Random Oversampling, dan klasifikasi menggunakan Multinomial Naive Bayes. Tahap *evaluation* digunakan untuk mengukur kinerja model menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Selanjutnya, hasil klasifikasi digunakan untuk menghitung indeks *brand strength* berbasis sentimen dan melakukan *semantic retrieval* sebagai bagian dari pemanfaatan hasil analisis.

Penggunaan CRISP-DM pada penelitian berbasis komentar YouTube juga telah diterapkan dalam penelitian sebelumnya, antara lain untuk analisis jejaring sosial dan sentimen pada konten vlog, yang menunjukkan bahwa kerangka tersebut dapat digunakan untuk mengorganisasi tahapan pengolahan data hingga evaluasi hasil analisis [16].

Business Understanding

Permasalahan penelitian berangkat dari kebutuhan untuk memahami persepsi pengguna terhadap laptop Advan dan Axioo berdasarkan data opini digital. Komentar YouTube dipilih karena dapat memuat pengalaman dan penilaian pengguna secara langsung. Informasi yang ingin diperoleh meliputi kecenderungan sentimen, performa model klasifikasi, perbandingan indeks brand strength berbasis sentimen, serta contoh komentar yang relevan terhadap query.

Pengumpulan Data

Data dikumpulkan dari komentar YouTube pada video yang membahas atau membandingkan laptop Advan dan Axioo. Pengambilan data menggunakan library YoutubeCommentDownloader. Pada proses pemeriksaan URL, video yang memiliki video ID sama tidak dihitung sebagai video berbeda meskipun parameter URL berbeda. Setelah penyaringan tersebut diperoleh 21 video unik. Scraping menghasilkan 3.519 komentar. Seluruh komentar kemudian diperiksa untuk menghapus duplikasi sehingga diperoleh 3.417 komentar unik sebagai dataset akhir.

Preprocessing Data

Preprocessing dilakukan karena komentar YouTube memiliki bentuk teks yang tidak seragam. Tahap pertama adalah case folding dengan mengubah seluruh karakter menjadi huruf kecil. Tahap kedua adalah menghapus URL, karena alamat web tidak diperlukan untuk klasifikasi sentimen. Tahap ketiga adalah menghapus karakter yang tidak relevan dan merapikan spasi. Tahap keempat adalah normalisasi kata informal, misalnya "gk", "ga", "nggak", dan "tdk" dinormalisasi menjadi "tidak". Tahap kelima adalah tokenisasi untuk memisahkan teks menjadi kata. Tahap keenam adalah stopword removal. Namun, kata negasi seperti "tidak", "bukan", "jangan", "belum", "tanpa", dan "tak" dipertahankan karena dapat mengubah makna sentimen. Tahap berikutnya adalah penghapusan kata yang terlalu pendek, kemudian dilakukan *stemming* menggunakan Sastrawi untuk mengubah kata berimbuhan menjadi bentuk dasarnya. Seluruh tahapan *preprocessing* diterapkan secara konsisten karena pilihan teknik dan urutan *preprocessing* dapat memengaruhi kinerja *sentiment analysis*. Penelitian terdahulu menunjukkan bahwa perubahan teknik dan urutan *preprocessing* dapat menghasilkan perbedaan performa klasifikasi, sehingga urutan pemrosesan perlu ditetapkan secara konsisten dalam suatu eksperimen[17].

Pelabelan Sentimen

Pelabelan awal dilakukan secara otomatis menggunakan TextBlob berdasarkan nilai *polarity*. Komentar dengan *polarity* > 0,1 diberi label positif, *polarity* < −0,1 diberi label negatif, sedangkan nilai antara −0,1 dan 0,1 diberi label netral. Label tersebut digunakan sebagai *weak label*, bukan *ground truth*, karena TextBlob memiliki keterbatasan dalam menangani karakteristik bahasa Indonesia [18]. Keterbatasan kualitas label ini perlu diperhatikan karena dapat memengaruhi hasil klasifikasi.

Identifikasi Merek

Identifikasi merek dilakukan berdasarkan kemunculan kata "Advan" dan "Axioo" pada komentar. Komentar yang hanya menyebut Advan atau Axioo dikelompokkan sesuai mereknya, sedangkan komentar yang menyebut kedua merek dimasukkan ke kategori Keduanya. Sebanyak 315 komentar termasuk kategori Keduanya dan tidak digunakan dalam perhitungan indeks *brand strength* untuk mencegah satu komentar berkontribusi pada dua merek sekaligus.

Pembagian Data

Dataset akhir sebanyak 3.417 komentar dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratified split. Dengan prosedur tersebut diperoleh 2.733 data latih dan 684 data uji. Stratification digunakan agar ketiga kelas tetap terwakili pada kedua bagian data. Data uji disimpan dalam kondisi asli dan tidak digunakan untuk menentukan vocabulary, nilai IDF, maupun proses oversampling.

Pembobotan TF-IDF

TF-IDF digunakan untuk mengubah komentar menjadi matriks numerik. Pada tahap training, vectorizer di-fit hanya pada data latih sehingga vocabulary dan nilai IDF tidak memperoleh informasi dari data uji. Data uji kemudian ditransformasi menggunakan vectorizer yang sama. Pembatasan ini penting untuk mencegah data leakage. Secara umum, bobot TF-IDF dapat dipahami sebagai kombinasi antara frekuensi suatu kata dalam dokumen dan tingkat kelangkaan kata tersebut pada seluruh dokumen[9].

Random Oversampling

Data latih memiliki distribusi kelas yang tidak seimbang sehingga Random Oversampling diterapkan hanya pada data latih untuk mengurangi dominasi kelas mayoritas. Oversampling dilakukan setelah pembagian data, sedangkan data uji tetap dalam kondisi asli. Setelah oversampling,

ketiga kelas pada data latih menjadi seimbang. Prosedur ini diterapkan untuk mengurangi risiko *data leakage* dan memperoleh evaluasi model yang lebih objektif [12].

Klasifikasi Multinomial Naive Bayes

Multinomial Naive Bayes digunakan untuk memprediksi tiga kelas sentimen. Model dilatih menggunakan data latih yang telah diseimbangkan. Algoritma ini mengasumsikan independensi bersyarat antarfitur berdasarkan kelas. Meskipun asumsi tersebut sederhana, Naive Bayes sering digunakan sebagai baseline klasifikasi teks karena efisien dan mudah diimplementasikan[19].

Evaluasi Model

Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. *Accuracy* mengukur proporsi prediksi yang benar, *precision* mengukur ketepatan prediksi suatu kelas, sedangkan *recall* mengukur kemampuan model dalam mengenali anggota suatu kelas. *F1-score* merupakan *harmonic mean* dari *precision* dan *recall*. Karena distribusi kelas tidak seimbang, penelitian ini juga menggunakan *macro average* selain *weighted average*. *Macro average* memberikan bobot yang sama pada setiap kelas sehingga performa kelas minoritas tetap diperhitungkan [20].

Perhitungan *Brand Strength*

Indeks *brand strength* berbasis sentimen dihitung berdasarkan selisih jumlah komentar positif dan negatif yang dinormalisasi terhadap total komentar pada masing-masing merek. Rumus yang digunakan adalah Indeks Brand Strength = (Jumlah Positif − Jumlah Negatif) / Total Komentar Merek. Indeks ini bukan merupakan ukuran standar *brand equity*, melainkan indikator komparatif untuk menggambarkan kecenderungan sentimen pada dataset penelitian. Komentar kategori Keduanya dikeluarkan dari perhitungan sehingga jumlah komentar yang digunakan untuk masing-masing merek berbeda dari total keseluruhan dataset.

Semantic Retrieval

Semantic retrieval menggunakan model *Sentence Transformer* multilingual untuk menghasilkan *embedding* komentar dan *query*. Representasi tersebut kemudian dibandingkan menggunakan *cosine similarity* untuk mengukur kedekatan semantik[14].Lima komentar dengan nilai *similarity* tertinggi dipilih sebagai hasil *retrieval*. Pendekatan ini tidak menghasilkan jawaban generatif sehingga istilah yang digunakan adalah *semantic retrieval*, bukan RAG.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan dan Pembersihan Data

Tabel 1. Pengumpulan Dataset

Tahapan	Jumlah	Keterangan
Data hasil scraping	3.519	Komentar dari 21 video unik
Data setelah penghapusan duplikat	3.417	Dataset akhir
Komentar yang menyebut kedua merek	315	Tidak digunakan dalam indeks brand utama

Hasil scraping menghasilkan 3.519 komentar. Setelah dilakukan pemeriksaan duplikasi, sebanyak 102 komentar dihapus sehingga tersisa 3.417 komentar unik. Penghapusan duplikasi diperlukan agar komentar yang sama tidak memberikan bobot berulang dalam distribusi sentimen maupun proses klasifikasi. Selain itu, 315 komentar teridentifikasi menyebut Advan dan Axioo secara bersamaan. Komentar tersebut dipertahankan dalam dataset klasifikasi, tetapi tidak dimasukkan ke perhitungan indeks brand strength agar atribusi merek tidak tumpang tindih.

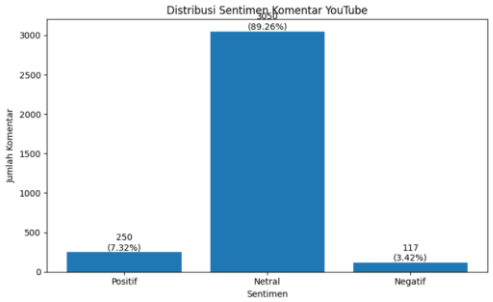
Hasil Distribusi Sentimen

Setelah proses *preprocessing* dan pelabelan, diperoleh tiga kategori sentimen.

Tabel 2. Distribusi Sentimen

Sentimen	Jumlah	Persentase
Positif	250	7,32%
Netral	3.050	89,26%
Negatif	117	3,42%
Total	3.417	100%

Distribusi menunjukkan dominasi yang sangat kuat pada kelas netral, yaitu 3.050 komentar atau 89,26%. Komentar positif berjumlah 250 atau 7,32%, sedangkan komentar negatif berjumlah 117 atau 3,42%. Dominasi kelas netral menjadi salah satu karakteristik paling penting dalam penelitian karena memengaruhi interpretasi performa model. Accuracy yang tinggi pada dataset seperti ini dapat terjadi jika model banyak memprediksi kelas mayoritas. Oleh sebab itu, hasil klasifikasi tidak cukup dinilai hanya melalui accuracy, tetapi juga melalui macro F1 dan metrik per kelas.



Gambar 1. Distribusi Sentimen Komentar YouTube

Sumber: Hasil pengolahan data penelitian.

Dominasi netral juga perlu dipahami sebagai hasil dari mekanisme pelabelan TextBlob. Label netral berarti nilai polarity berada pada rentang yang telah ditentukan, bukan berarti komentar tidak memiliki informasi mengenai produk. Komentar seperti pertanyaan spesifikasi, penyebutan harga, atau pernyataan faktual dapat masuk kategori netral walaupun tetap berguna untuk memahami persepsi konsumen.

Distribusi Sentimen Berdasarkan Merek

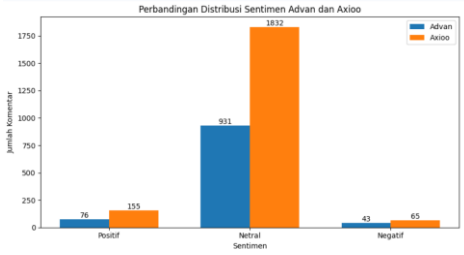
Distribusi sentimen kemudian dipisahkan berdasarkan merek.

Tabel 3. Distribusi Sentimen Dan Indeks Brand Strength

Brand	Negatif	Netral	Positif	Total	Indeks Brand Strength
Advan	43	931	76	1.050	0,031429
Axioo	65	1.832	155	2.052	0,043860

Pada kelompok Advan terdapat 43 komentar negatif, 931 netral, dan 76 positif. Pada kelompok Axioo terdapat 65 komentar negatif, 1.832 netral, dan 155 positif. Jumlah komentar Axioo lebih besar sehingga perbandingan jumlah absolut tidak dapat digunakan secara langsung. Oleh karena itu, indeks dihitung dengan menormalisasi selisih komentar positif dan negatif terhadap jumlah komentar merek.

Indeks Advan dihitung sebagai (76 − 43) / 1.050 = 0,031429. Indeks Axioo dihitung sebagai (155 − 65) / 2.052 = 0,043860. Selisih kedua indeks sebesar 0,012431. Berdasarkan indikator penelitian, Axioo memiliki kecenderungan indeks brand strength berbasis sentimen yang lebih tinggi. Namun, perbedaan ini harus dibaca secara terbatas karena indeks tidak mencakup seluruh dimensi kekuatan merek dan hanya menggunakan komentar YouTube yang tersedia.



Gambar 2. Perbandingan Distribusi Sentimen Komentar Advan dan Axioo

Sumber: Hasil pengolahan data penelitian.

Hasil Pembagian Data dan Oversampling

Tabel 4. Distribusi Kelas Data Latih dan Data Uji Sebelum dan Sesudah Oversampling

Tahap	Negatif	Netral	Positif	Total
Data latih sebelum oversampling	94	2.439	200	2.733
Data latih setelah oversampling	2.439	2.439	2.439	7.317
Data uji	23	611	50	684

Data dibagi menjadi 2.733 data latih dan 684 data uji. Data latih awal sangat didominasi kelas netral. Random Oversampling kemudian meningkatkan kelas minoritas hingga masing-masing kelas pada data latih berjumlah 2.439. Data uji tidak mengalami oversampling. Dengan prosedur ini, model memperoleh data latih yang seimbang tetapi tetap dievaluasi pada data uji yang merepresentasikan distribusi asli. Pendekatan tersebut menghindari *data leakage* dan menjaga data uji tetap merepresentasikan distribusi asli.

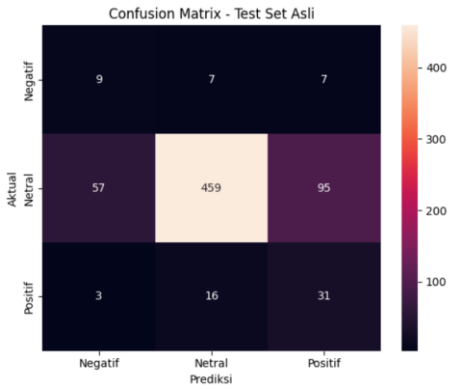
Hasil *Classification Report*

Tabel 5. Hasil Classification Report

Kelas	Precision	Recall	F1-score	Support
Negatif	0,130	0,391	0,196	23
Netral	0,952	0,751	0,840	611
Positif	0,233	0,620	0,339	50
Macro Average	0,439	0,588	0,458	684
Weighted Average	0,872	0,730	0,782	684

Hasil menunjukkan bahwa kelas netral memperoleh performa paling baik dengan precision 0,952, recall 0,751, dan F1-score 0,840. Kondisi ini berkaitan dengan jumlah kelas netral yang jauh lebih besar pada data uji. Kelas negatif memperoleh precision 0,130, recall 0,391, dan F1-score 0,196. Kelas positif memperoleh precision 0,233, recall 0,620, dan F1-score 0,339. Rendahnya F1 pada dua kelas minoritas menunjukkan bahwa model masih kesulitan membedakan komentar yang memiliki polaritas positif atau negatif.

Perbedaan antara weighted F1 sebesar 0,782 dan macro F1 sebesar 0,458 sangat penting. Weighted F1 memberi kontribusi lebih besar kepada kelas netral karena jumlahnya paling banyak. Sebaliknya, macro F1 menghitung rata-rata setiap kelas secara setara. Oleh karena itu, macro F1 lebih tepat digunakan untuk menunjukkan bahwa performa model belum merata.



Gambar 3. Confusion Matrix Multinomial Naive Bayes

Sumber: Hasil pengolahan data penelitian.

Evaluasi Model Secara Keseluruhan

Tabel 6. Evaluasi Model

Metrik	Nilai	Interpretasi
Accuracy	0,730	73,0% prediksi pada data uji benar
Precision weighted	0,872	Ketepatan prediksi berbobot kelas
Recall weighted	0,730	Kemampuan menemukan kelas secara berbobot
F1-score weighted	0,782	Kinerja agregat berbobot
Precision macro	0,439	Rata-rata precision antar kelas
Recall macro	0,588	Rata-rata recall antar kelas
F1-score macro	0,458	Rata-rata F1 antar kelas

Accuracy sebesar 0,730 menunjukkan bahwa 73,0% prediksi pada data uji sesuai dengan label yang tersedia. Namun, angka ini tidak boleh ditafsirkan sebagai bukti bahwa model sangat baik dalam memahami sentimen karena distribusi kelas tidak seimbang. Weighted F1 sebesar 0,782 juga banyak dipengaruhi oleh kelas netral. Macro F1 sebesar 0,458 memberikan gambaran yang lebih konservatif dan menunjukkan adanya kesenjangan performa antar kelas.

Secara praktis, model cukup baik untuk mengenali komentar netral, tetapi belum memadai jika tujuan utama sistem adalah mengidentifikasi komentar positif dan negatif secara seimbang. Hasil penelitian pada data berbahasa Indonesia juga menunjukkan bahwa model dan

representasi bahasa dapat memberikan performa yang berbeda secara signifikan.

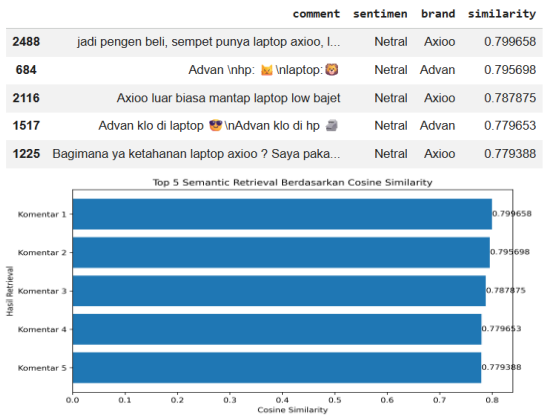
Hasil Semantic Retrieval

Query pengujian yang digunakan adalah "kelebihan dan kekurangan laptop Advan dan Axioo". Sistem mengubah query menjadi embedding kemudian menghitung cosine similarity terhadap seluruh komentar. Lima komentar dengan similarity tertinggi digunakan sebagai hasil retrieval. Pada notebook revisi, nilai similarity lima hasil teratas berada pada kisaran sekitar 0,779–0,800.

Tabel 7. Hasil Semantic Retrieval

Peringkat	Similarity	Keterangan
1	0,799658	Komentar dengan kedekatan semantik tertinggi
2	0,795698	Komentar relevan berikutnya
3	0,787875	Komentar relevan berikutnya
4	0,779653	Komentar relevan berikutnya
5	0,779388	Komentar relevan kelima

Nilai similarity menunjukkan tingkat kedekatan representasi vektor, bukan probabilitas bahwa komentar tersebut benar atau merupakan jawaban terbaik. Karena hanya lima komentar yang ditampilkan, hasil retrieval tidak dapat digunakan untuk menghitung frekuensi topik seluruh dataset. Retrieval lebih tepat diposisikan sebagai alat eksplorasi untuk memberikan contoh konteks opini yang relevan terhadap pertanyaan pengguna.



	comment	sentimen	brand	similarity
2488	jadi pengen beli, sempet punya laptop axioo, l...	Netral	Axioo	0.799658
684	Advan \nhp: \nlaptop:	Netral	Advan	0.795698
2116	Axioo luar biasa mantap laptop low bajet	Netral	Axioo	0.787875
1517	Advan klo di laptop \nAdvan klo di hp	Netral	Advan	0.779653
1225	Bagimana ya ketahanan laptop axioo ? Saya paka...	Netral	Axioo	0.779388

Gambar 4. Hasil Semantic Retrieval Berdasarkan Cosine Similarity
Sumber: Hasil pengolahan data penelitian.

Pembahasan

Temuan pertama adalah dominasi sentimen netral. Sebanyak 89,26% komentar masuk kategori netral berdasarkan aturan TextBlob. Kondisi ini dapat menunjukkan bahwa banyak komentar bersifat informatif, deskriptif, atau tidak memiliki polarity yang cukup kuat. Namun, karena label dibuat secara otomatis, hasil ini tidak dapat langsung digunakan untuk menyatakan bahwa mayoritas konsumen memiliki sikap netral terhadap kedua merek.

Temuan kedua adalah adanya perbedaan dalam indeks brand strength berbasis sentimen. Axioo memperoleh nilai 0,043860, sedangkan Advan memperoleh 0,031429. Nilai positif kedua merek menunjukkan bahwa jumlah komentar positif lebih banyak daripada negatif pada kelompok komentar yang hanya menyebut satu merek. Axioo memiliki indeks lebih tinggi setelah dinormalisasi. Walaupun demikian, indeks ini tidak memasukkan awareness, perceived quality, loyalty, atau asosiasi merek secara langsung.

Temuan ketiga adalah performa model yang tidak seimbang. Model menghasilkan F1 0,840 pada kelas netral, tetapi hanya 0,196 pada kelas negatif dan 0,339 pada kelas positif. Hal tersebut menunjukkan bahwa oversampling memberikan representasi kelas minoritas pada data latih, tetapi belum otomatis membuat model mampu mengenali karakteristik bahasa minoritas secara akurat.

Temuan keempat berkaitan dengan validitas evaluasi. Prosedur pada versi penelitian ini lebih baik karena train-test split dilakukan sebelum oversampling dan TF-IDF hanya di-fit pada data latih. Perbaikan ini penting agar informasi dari data uji tidak digunakan dalam proses pembelajaran, sehingga evaluasi model menjadi lebih ketat dan lebih realistis.

Temuan kelima berkaitan dengan penggunaan TextBlob. Karena TextBlob digunakan untuk menghasilkan label yang

kemudian dipelajari oleh Naive Bayes, hasil evaluasi terutama menunjukkan kemampuan model untuk mempelajari pola label otomatis tersebut. Validasi manusia belum dilakukan. Oleh karena itu, penelitian selanjutnya sebaiknya membuat sebagian dataset berlabel manual dan membandingkan label manual dengan label TextBlob.

Temuan keenam adalah fungsi semantic retrieval. Retrieval dapat membantu pengguna menemukan komentar yang relevan tanpa harus membaca ribuan komentar satu per satu. Namun, retrieval tidak boleh dianggap sebagai bukti statistik mengenai topik yang paling dominan. Jika ingin menghasilkan sistem RAG yang sesungguhnya, penelitian perlu menambahkan model generatif yang menyusun jawaban berdasarkan komentar hasil retrieval dan mengevaluasi kualitas jawaban tersebut.

Secara keseluruhan, penelitian menunjukkan bahwa kombinasi text mining, sentiment classification, indeks sentimen, dan semantic retrieval dapat digunakan untuk mengeksplorasi persepsi pengguna terhadap produk laptop. Akan tetapi, hasilnya sangat bergantung pada kualitas data, metode pelabelan, ketidakseimbangan kelas, dan karakteristik bahasa komentar. Oleh karena itu, hasil penelitian lebih tepat digunakan sebagai gambaran persepsi pada dataset YouTube yang dikumpulkan daripada sebagai representasi seluruh populasi konsumen.

KESIMPULAN

Penelitian berhasil menganalisis 3.417 komentar YouTube unik mengenai laptop Advan dan Axioo menggunakan text mining, analisis sentimen, Multinomial Naive Bayes, indeks brand strength berbasis sentimen, dan semantic retrieval.

Distribusi label terdiri atas 250 komentar positif, 3.050 netral, dan 117 negatif. Kelas netral menjadi kategori dominan dengan 89,26% dari dataset.

Setelah data dibagi dengan stratified split dan Random Oversampling hanya diterapkan pada data latih, Multinomial Naive Bayes menghasilkan accuracy 0,730, weighted F1-score 0,782, dan macro F1-score 0,458. Model paling baik mengenali kelas netral, sedangkan kelas negatif dan positif masih memiliki F1 yang rendah.

Indeks brand strength berbasis sentimen menghasilkan 0,031429 untuk Advan dan 0,043860 untuk Axioo. Berdasarkan indikator tersebut Axioo memperoleh nilai lebih tinggi, tetapi hasil tersebut tidak dapat dianggap sebagai ukuran keseluruhan brand equity.

Semantic retrieval berhasil memperoleh komentar yang relevan terhadap query menggunakan embedding dan cosine similarity. Implementasi penelitian merupakan semantic retrieval berbasis embedding dan belum termasuk RAG karena tidak memiliki komponen generatif.

DAFTAR PUSTAKA

[1] O. Grljević dan Z. Bošnjak, "Sentiment analysis of customer data," *Strategic Management*, vol. 23, no. 3, hlm. 38–49, 2018, doi: 10.5937/StraMan1803038G.

[2] L. Xinwei, Y. K. Tse, dan F. Fastoso, "Unleashing the power of social media data in business decision making: an exploratory study," *Enterprise Information Systems*, vol. 18, no. 1, hlm. 2243603, Jan 2024, doi: 10.1080/17517575.2023.2243603.

[3] Institute of Economy and Informatics, Faculty of Organization and Management of Silesian University of Technology dan M. Wyskwarski, "SENTIMENT ANALYSIS OF COMMENTS POSTED ON YOUTUBE VIDEOS RELATED TO PHOTOVOLTAICS," *SPSUTOM*, vol. 2023, no. 185, hlm. 617–637, 2023, doi: 10.29119/1641-3466.2023.185.36.

[4] S. Pichad, "Analysing Sentiments for YouTube Comments using Machine

Learning," *IJRASET*, vol. 11, no. 5, hlm. 1934–1938, Mei 2023, doi: 10.22214/ijraset.2023.51973.

[5] H. Jin, "From Information to Affirmation: An Investigation on the Echo Chamber Effect from YouTube Comments under Technology Product Reviews," 2022, doi: 10.17605/OSF.IO/J2WQM.

[6] K. Victor Rajan dan F. Frejus, "Machine Learning Techniques for Opinion Mining from Social Media," *Computer*, Mei 2022, doi: 10.54647/computer52271.

[7] Former Dean and Assistant Marketing Professor at Graduate School of Management at Sharif University of Technology, Iran, Dr. A. Miremadi, O. Ghanadiof, dan Graduate of Industrial Management, School of Technology, University of Central Missouri, USA, "The Ultimate Influences of Brand Equity Dimensions on Consumer Decision in Hi-Tech Market," *AJRSP*, vol. 3, no. 27, hlm. 27–46, Jul 2021, doi: 10.52132/Ajrsp.e.2021.272.

[8] M. Ranjan, S. Tiwari, A. M. Sattar, dan N. S. Tatkar, "A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing," dalam *RAiSE-2023*, MDPI, Jan 2024, hlm. 181. doi: 10.3390/engproc2023059181.

[9] R. Wati, S. Ernawati, dan H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *JAMIKA*, vol. 13, no. 1, hlm. 84–93, Apr 2023, doi: 10.34010/jamika.v13i1.9424.

[10] N. Umar dan M. Adnan Nur, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no. 4, hlm. 585–590, Agu 2022, doi: 10.29207/resti.v6i4.4179.

[11] A. S. Iedwan, N. Mauliza, Y. Pristyanto, A. D. Hartanto, dan A. N. Rohman, "Comparative Performance of SVM and Multinomial Naïve Bayes in Sentiment Analysis of the Film 'Dirty Vote,'" *sji*, vol. 11, no. 3, hlm. 839–848, Nov 2024, doi: 10.15294/sji.v11i3.10290.

[12] P. Wibowo dan C. Fatichah, "An in-depth performance analysis of the oversampling techniques for high-class imbalanced dataset," *regist. j. ilm. teknol. sist. inf.*, vol. 7, no. 1, hlm. 63, Feb 2021, doi: 10.26594/register.v7i1.2206.

[13] S. F. N. Permatasari dan E. Ermatita, "A Two-Stage Hybrid Oversampling and Ensemble Learning Framework for Improved Type 2 Diabetes Mellitus Classification," *ijeeemi*, vol. 8, no. 2, hlm. 202–217, Mei 2026, doi: 10.35882/ijeeemi.v8i2.308.

[14] D. Katta, "Multilingual Text Document Clustering and Classification," *IJRASET*, vol. 13, no. 7, hlm. 1009–1014, Jul 2025, doi: 10.22214/ijraset.2025.73119.

[15] S. R. El-Beltagy dan M. A. Abdallah, "Exploring Retrieval Augmented Generation in Arabic," 2024, *arXiv*. doi: 10.48550/ARXIV.2408.07425.

[16] Y. A. Singgalen, "Implementation of CRISP-DM for Social Network Analysis (SNA) of Tourism and Travel Vlog Content Reviews," *mib*, vol. 8, no. 1, hlm. 572, Jan 2024, doi: 10.30865/mib.v8i1.7323.

[17] S. Magsarjav, M. Humphries, J. Tuke, dan L. Mitchell, "Best Preprocessing Techniques for Sentiment Analysis," 2026, *arXiv*. doi: 10.48550/ARXIV.2606.24055.

[18] C. L. H. S. Nivak, "Sentiment Analysis of Ten Tourism Destinations in Indonesia," *J. Ind Tour Dev Std*, vol. 9, no. 3, hlm. 162–168, Sep 2021, doi: 10.21776/ub.jitode.2021.009.03.03.

[19] N. Umar dan M. Adnan Nur, "Application of Naïve Bayes

Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no. 4, hlm. 585–590, Agu 2022, doi: 10.29207/resti.v6i4.4179.

[20] S. Riyanto, I. S. Sitanggang, T. Djatna, dan T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *IJACSA*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406116.