ANALISIS SENTIMEN DAN IDENTIFIKASI PRODUK TERLARIS PADA E-COMMERCE PAKAIAN ANAK MENGGUNAKAN K-NEAREST NEIGHBOR

Siti Jatsiah[1)], Ridha Adjie Eryadi[2)], Mohammad Nur Fitriyadi[3)]

1. Informatika, Universitas Teknologi Digital
   email: siti20122004@digitechuniversity.ac.id
2. Informatika, Universitas Teknologi Digital
   email: ridhaadjie@digitechuniversity.ac.id
3. Informatika, Universitas Teknologi Digital
   email: mohammadnur@digitechuniversity.ac.id

Abstract

The rapid growth of the children's clothing E-Commerce industry requires businesses to understand consumer preferences through digital review data. This study aims to perform Sentiment Analysis of customer reviews and identify best-selling products at the Baju Anak Kece Online Store by utilizing Text Mining techniques. The primary challenge addressed is the volume of unstructured customer reviews, making it difficult for store owners to accurately determine satisfaction levels and purchasing patterns. The method employed in this research is the K-Nearest Neighbor (KNN) algorithm to classify customer reviews into three sentiment categories: Positive, Neutral, and Negative. Text preprocessing stages include case folding, cleaning, tokenizing, Stopword removal, and Stemming using the Sastrawi Library, followed by word weighting using the TF-IDF method. To illustrate the scale of the experiment, this study utilized a dataset of 326 customer reviews, which were validated and divided into an 80% training set and a 20% testing set. The results indicate that the KNN algorithm is capable of classifying customer sentiment with an Accuracy rate of 69.23%. These findings are integrated to identify best-selling products based on the dominance of positive sentiment volume, providing actionable marketing strategy recommendations for the store to optimize stock and enhance competitiveness in the digital market.

Kata Kunci : *Text Mining, K-Nearest Neighbor, Sentiment Analysis, Children Clothing, TF-IDF*

A. PENDAHULUAN

Pesatnya pertumbuhan E-Commerce di Indonesia, khususnya pada segmen pakaian anak, mendorong persaingan yang semakin ketat di *Marketplace*. Dalam persaingan retail digital ini, ulasan pelanggan menjadi salah satu aset data tekstual terpenting untuk mengetahui tingkat kepuasan, keluhan, serta preferensi nyata konsumen terhadap suatu produk secara berkala (Irwannia & Lubis, 2025). Berbagai literatur ilmiah menyebutkan bahwa identifikasi produk terlaris (*best-seller*) merupakan salah satu kunci utama dalam pengambilan keputusan operasional perusahaan, termasuk estimasi pengadaan stok barang baru, meminimalkan kerugian akibat *deadstock*, dan penyusunan strategi promosi yang lebih tertarget. Dengan menyasar lini bisnis baju anak, analisis sentimen dari ulasan orang tua memiliki karakteristik yang sangat unik karena faktor kenyamanan bahan, ketepatan ukuran, dan kesesuaian warna sangat dominan dalam memengaruhi loyalitas dan keputusan pembelian kembali (Nabiilah et al., 2025).

Namun, permasalahan utama yang sering dihadapi oleh pemilik Toko *Online* Baju Anak Kece adalah ulasan pelanggan yang masuk setiap harinya bersifat tidak terstruktur, memiliki volume yang sangat besar, serta kerap menggunakan bahasa sehari-hari atau slang yang menyulitkan pengelompokkan manual. Keterbatasan waktu membuat evaluasi kualitas produk menjadi tidak optimal. Oleh karena itu, diperlukan sebuah pendekatan berbasis teknologi komputeryakni teknik *Text Mining*. *Text Mining* secara khusus bekerja untuk mengekstraksi informasi berharga dan pola tersembunyi dari sekumpulan data teks tidak terstruktur melalui serangkaian tahapan komputasi yang meliputi text *preprocessing*, transformasi fitur numerik, hingga pemodelan klasifikasi menggunakan *Machine Learning* (Feldman & Sanger, 2007).

Salah satu algoritma *Machine Learning* yang dikenal sangat efektif, efisien, dan memiliki basis matematis yang kuat untuk tugas klasifikasi teks berskala kecil hingga menengah adalah K-Nearest Neighbor (KNN) (Saifurridho et al., 2024). Algoritma KNN bekerja dengan cara mengklasifikasikan data uji baru berdasarkan tingkat kedekatan jarak koordinat (similiaritas) terhadap sejumlah data latih terdekat yang sudah memiliki label. Penerapan KNN pada *text mining* ulasan toko *online* sangat cocok digunakan untuk mengelompokkan ragam respons emosional pelanggan ke dalam kategori sentimen tertentu secara presisi (Apip & Kurniawan, 2026). Berdasarkan urgensi tersebut, penelitian ini bertujuan untuk mengonstruksi sebuah sistem analisis sentimen otomatis yang terintegrasi dengan modul identifikasi produk terlaris. Penelitian ini diharapkan mampu memberikan kontribusi nyata secara teoretis dalam pengembangan ilmu pengolahan bahasa alami serta kontribusi praktis dalam memberikan rekomendasi manajemen bisnis berbasis data retail digital yang adaptif (Indrayuni, 2017).

B. METODE PENELITIAN

Metode penelitian ini disusun secara sistematis menggunakan pendekatan kuantitatif berbasis eksperimen laboratorium komputer untuk menjamin validitas hasil pengolahan data. Objek penelitian yang dipilih adalah Toko Online Baju Anak Kece yang beroperasi aktif di platform E-Commerce Shopee (Octavia & Berlilana, 2025). Alur pengerjaan penelitian secara garis besar terbagi ke dalam empat tahapan utama: (1) Pengumpulan Data (*Scraping*), (2) *Preprocessing* Teks, (3) Pembobotan Fitur dengan TF-IDF, dan (4) Pemodelan Klasifikasi menggunakan Algoritma K-Nearest Neighbor (KNN). Penjelasan mendetail mengenai masing-masing tahapan diuraikan sebagai berikut:

1. Pengumpulan Data (*Data Scraping*)

Proses pengumpulan data ulasan pelanggan dilakukan dengan teknik *web scraping* menggunakan pustaka Python yang mengekstrak data langsung dari API publik Shopee. Ruang lingkup pengambilan data dibatasi pada transaksi penjualan produk pakaian anak periode November 2025 sampai dengan Februari 2026. Atribut data yang diambil meliputi nama ID pengguna, tanggal ulasan, teks komentar ulasan, rating bintang yang diberikan (1 hingga 5), serta kode identifikasi produk (product ID). Dataset total awal yang berhasil dihimpun melalui aktivitas *web scraping* pada akun Toko *Online* Baju Anak Kece di platform Shopee berjumlah 326 data record ulasan riil konsumen (data mentah). Sebelum dilakukan pemodelan, dataset mentah ini terlebih dahulu melalui tahapan *preprocessing* dan penyaringan data (*cleansing*), di mana ulasan yang bersifat *noise*, duplikat, atau ulasan kosong yang tidak memiliki bobot kata setelah penghapusan karakter *non*-alfabet (seperti emoji dan simbol) dieliminasi dari sistem. Dari proses pembersihan tersebut, diperoleh dataset bersih (*clean* data) yang valid sebanyak 295 data record ulasan.

Dataset valid inilah yang kemudian dibagi secara acak menggunakan metode standardisasi eksperimen *Hold-out Method* dengan proporsi pembagian yang mendekati rasio seimbang 80:20. Melalui penerapan pembagian data tersebut, diperoleh sebaran data berupa 230 data *record* ulasan latih (training set) yang digunakan oleh sistem untuk mempelajari pembobotan kosa kata dan pola karakteristik kalimat sentimen, serta 65 data *record* ulasan uji acak (testing set) yang murni digunakan sebagai instrumen pengujian keandalan prediksi klasifikasi model KNN.

2. Tahapan Preprocessing Teks (Pembersihan Data)

Data teks hasil *scraping* tidak dapat langsung diumpankan ke dalam algoritma klasifikasi karena mengandung derau (*noise*) yang tinggi. Oleh karena itu, tahapan *preprocessing* teks wajib dilakukan untuk menstandardisasi dokumen kalimat menjadi kumpulan kata bersih. Tahapan preprocessing teks dalam penelitian ini terdiri dari lima urutan logis sebagai berikut:

- Case Folding: Mengubah seluruh karakter huruf di dalam teks ulasan menjadi huruf kecil (*lowercase*). Langkah ini memastikan bahwa kata yang sama namun memiliki perbedaan penulisan huruf kapital (misalnya "Baju", "BAJU", dan "baju") akan diidentifikasi sebagai satu token kata yang seragam.
- Cleansing: Proses pembersihan dokumen dari komponen *non*-alfabet yang tidak merepresentasikan nilai sentimen. Pembersihan dilakukan menggunakan skrip Regular Expression (Regex) di Python untuk menghapus angka, simbol matematika, tanda baca (titik, koma, tanda seru), karakter spasi kosong berlebih (*whitespace*), tautan alamat URL, serta karakter emoji visual.
- Tokenizing: Memotong string kalimat ulasan yang panjang menjadi satuan elemen kata tunggal terpisah yang disebut sebagai token. Proses ini memecah kalimat berdasarkan spasi antarkata, sehingga mempermudah penghitungan frekuensi kemunculan setiap kata unik dalam satu dokumen ulasan.
- Filtering (Stopword Removal): Menghilangkan kata-kata umum (stopwords) dalam tata bahasa Indonesia yang memiliki frekuensi kemunculan tinggi namun tidak memiliki bobot informasi penting terhadap penentuan kelas sentimen opini. Contoh kata yang dihapus adalah kata depan dan kata hubung seperti "yang", "di", "dan", "dari",

"untuk", "ke", "saya", "ini". Daftar *stopword* mengacu pada kamus standar yang disesuaikan dengan karakteristik ulasan belanja *online*.

- Stemming: Mengonversi kata-kata yang memiliki imbuhan (awalan, akhiran, atau sisipan) kembali menjadi bentuk kata dasarnya (*root word*). Proses ini sangat krusial dalam pemrosesan bahasa Indonesia yang kaya akan morfologi imbuhan. Pada penelitian ini, proses *stemming* dijalankan secara otomatis dengan mengintegrasikan Library Sastrawi Python, yang bekerja mencocokkan kata berimbuhan terhadap algoritma Nazief-Adriani dan kamus besar bahasa Indonesia (Ma'rifah et al., 2020). Sebagai ilustrasi konkret mengenai jalannya proses transformasi data pada tahap *preprocessing* teks, Tabel 1 menyajikan contoh simulasi pembersihan pada satu sampel ulasan pelanggan tidak baku.

Tabel 1. Ilustrasi Tahapan Preprocessing Teks Ulasan Pelanggan

Tahapan Preprocessing	Hasil Transformasi String Teks
Data Ulasan Mentah (Scraping)	"Baju NYA Bagusss bgt!! Tapi agak tipis yaa.. Pengiriman cepet G ☺"
1. Case Folding	"baju nya bagusss bgt!! tapi agak tipis yaa.. pengiriman cepet G ☺"
2. Cleansing	"baju nya bagusss bgt tapi agak tipis yaa pengiriman cepet"
3. Tokenizing	['baju', 'nya', 'bagusss', 'bgt', 'tapi', 'agak', 'tipis', 'yaa', 'pengiriman', 'cepet']
4. Filtering (Stopword)	['baju', 'bagusss', 'bgt', 'tipis', 'pengiriman', 'cepet']
5. Stemming (Sastrawi)	['baju', 'bagus', 'banget', 'tipis', 'kirim', 'cepat']

3. Pembobotan Kata Term Frequency - Inverse Document Frequency (TF-IDF)

Setelah data teks berhasil dibersihkan menjadi kumpulan kata dasar, data tersebut harus ditransformasikan ke dalam representasi numerik matriks bobot agar dapat diproses oleh perhitungan matematis algoritma KNN. Metode pembobotan kata yang diimplementasikan adalah TF-IDF (Prasetyo, 2012). Metode ini menggabungkan dua konsep pengukuran, yaitu *Term Frequency* (TF) yang menghitung frekuensi kemunculan suatu kata kunci spesifik dalam sebuah dokumen ulasan, dan *Inverse Document Frequency* (IDF) yang mengukur tingkat kelangkaan atau keunikan kata tersebut di seluruh populasi dokumen ulasan yang ada. Rumus matematis kalkulasi pembobotan nilai TF-IDF dinyatakan dalam persamaan berikut:

$$W_{d,t} = TF_{d,t} \times log(N / DF_t) \quad (1)$$

Di mana $W_{d,t}$ merepresentasikan bobot akhir dari term atau kata kunci t di dalam dokumen ulasan d. Variabel $TF_{d,t}$ menunjukkan frekuensi kemunculan kata t dalam dokumen d. Variabel N adalah total keseluruhan jumlah dokumen ulasan yang ada di dalam dataset penelitian, sedangkan DF_t merupakan *Document Frequency*, yaitu jumlah dokumen yang mengandung kata kunci t minimal sebanyak satu kali. Nilai bobot yang tinggi mengindikasikan bahwa kata kunci tersebut memiliki tingkat kepentingan yang signifikan dalam membedakan karakteristik ulasan antarproduk.

4. Klasifikasi K-Nearest Neighbor (KNN)

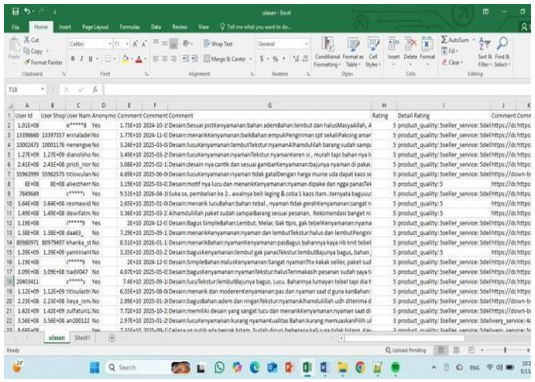
Tahap akhir dari kerangka kerja metodologi adalah pemodelan klasifikasi sentimen menggunakan algoritma K-Nearest Neighbor (KNN) (Halder et al., 2024). KNN merupakan algoritma berbasis pembelajaran terawasi (*supervised learning*) yang bekerja dengan membandingkan kesamaan karakteristik data uji baru terhadap data latih yang sudah ada di dalam basis data ruang fitur (Hidayati & Hermawan, 2021). Tingkat kesamaan antar-vektor ulasan dihitung menggunakan formula kedekatan geometris jarak koordinat ruang, yaitu rumus Euclidean Distance (Dinata et al., 2020). Persamaan matematis untuk menghitung jarak koordinat Euclidean Distance dirumuskan sebagai berikut:

$$d(x, y) = \sqrt{[\sum_{i=1}^{n} (x_i - y_i)^2]} \qquad (2)$$

Di mana d(x, y) merepresentasikan nilai jarak skalar koordinat geometris antara vektor dokumen uji x dan vektor dokumen latih y. Variabel x_i dan y_i mewakili nilai bobot fitur kata ke-i hasil perhitungan TF-IDF, sedangkan n melambangkan total dimensi jumlah kata unik (kosa kata) yang terbentuk di dalam seluruh dataset. Nilai jarak d yang semakin kecil menunjukkan tingkat kemiripan makna kosa kata ulasan yang semakin tinggi antara kedua dokumen tersebut. Langkah - langkah operasional jalannya pengujian algoritma klasifikasi KNN di dalam sistem informasi ini dieksekusi secara berurutan sebagai berikut: (a) Menetapkan parameter konstan jumlah tetangga terdekat, di mana dalam penelitian eksperimen ini ditentukan nilai konstanta ideal sebesar K=3. (b) Menghitung nilai jarak koordinat Euclidean Distance dari satu sampel dokumen data uji terhadap seluruh baris data latih yang ada di *database*. (c) Mengurutkan hasil perhitungan jarak skalar tersebut mulai dari nilai terkecil hingga nilai terbesar (secara *ascending*). (d) Mengambil sejumlah K=3 baris data latih teratas yang memiliki nilai jarak paling minimum (paling dekat). (e) Melakukan *voting* mayoritas terhadap label kelas sentimen (Positif, Netral, atau Negatif) yang melekat pada 3 data latih terdekat tersebut. (f) Menetapkan kategori kelas sentimen mayoritas hasil *voting* sebagai label prediksi final untuk dokumen ulasan uji baru tersebut (Hendriyanto & Sari, 2022).

C. HASIL DAN PEMBAHASAN

Dataset total penelitian yang berhasil dihimpun melalui aktivitas *web scraping* pada akun Toko *Online* Baju Anak Kece di platform Shopee berjumlah 326 data record ulasan riil konsumen. Sebelum dilakukan pemodelan, dataset terlebih dahulu divalidasi dan dibagi secara acak menggunakan metode standardisasi eksperimen *Hold-out Method* dengan proporsi pembagian yang seimbang sebesar 80% dialokasikan sebagai data latih (*training set*) dan 20% dialokasikan sebagai data uji (*testing set*). Melalui penerapan proporsi pembagian data tersebut, diperoleh sebaran data berupa 230 data record ulasan latih yang digunakan oleh sistem untuk mempelajari pembobotan kosa kata dan pola karakteristik kalimat sentimen, serta 65 data record ulasan uji acak yang murni digunakan sebagai instrumen pengujian keandalan prediksi klasifikasi model KNN. Implementasi fungsional dari keseluruhan sistem informasi *text mining* retail ini dirancang secara modular dan terintegrasi, di mana bagan alur arsitektur logika komputasi program disajikan secara komprehensif pada Gambar 1.



Gambar 1. Hasil data scraping

Guna mengukur tingkat performa, keandalan, dan validitas hasil prediksi dari pemodelan algoritma K-Nearest Neighbor (KNN) dengan nilai K=3, dilakukan proses pengujian matematis menggunakan instrumen statistik standar Confusion Matrix. Evaluasi melalui Confusion Matrix bekerja dengan cara memetakan matriks silang antara label kelas sentimen aktual yang ditentukan secara manual oleh pakar bahasa (*ground truth*) terhadap label kelas sentimen hasil keluaran prediksi otomatis program komputer. Rincian data kuantitatif mengenai performa hasil pengujian klasifikasi sentimen pada 65 sampel data uji disajikan secara mendetail pada Tabel 1.

Tabel 2. Matriks Evaluasi Klasifikasi Performa Sentimen Menggunakan Confusion Matrix (K=3)

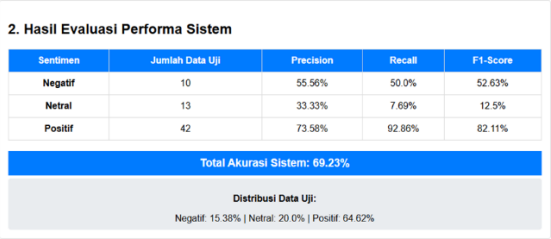
Kelas Aktual \ Prediksi	Negatif	Netral	Positif
Negatif (Actual)	5	2	2
Netral (Actual)	1	1	10
Positif (Actual)	4	1	39

Berdasarkan tabulasi silang nilai pengujian pada Tabel 1, tingkat nilai akurasi (*Accuracy Rate*) menyeluruh dari sistem informasi dihitung secara matematis dengan membagi akumulasi total jumlah data uji yang berhasil diprediksi secara tepat pada diagonal utama matriks (5 data negatif + 1 data netral + 39 data positif) dengan total keseluruhan sampel data uji eksperimen (65 data). Melalui formulasi kalkulasi tersebut, sistem pengklasifikasi sentimen otomatis berbasis KNN ini berhasil membuktikan keandalannya dengan mencatatkan nilai persentase akurasi total sebesar 69.23% (Sokolova & Lapalme, 2009). Hasil kuantitatif ini memperjelas bahwa pemanfaatan *text mining* dengan pemodelan jarak kedekatan koordinat vektor terbukti valid untuk memisahkan persepsi opini konsumen secara otomatis pada ranah *e-commerce* retail pakaian anak (Afif & Nugroho, 2025).

Apabila dilakukan analisis performa lebih mendalam pada masing-masing kategori kelas sentimen secara parsial, dapat ditarik fakta ilmiah yang sangat menarik. Tingkat akurasi performa klasifikasi tertinggi berada pada pendeteksian sentimen kelas Positif, di mana sistem berhasil mencatatkan persentase nilai sensitivitas (*Recall*) yang sangat superior, yaitu mencapai angka 92.86% (39 data terprediksi benar dari total 44 data aktual positif). Tingginya nilai performa Recall kelas positif ini dipengaruhi secara signifikan oleh karakteristik alamiah dari persebaran dataset ulasan *e-commerce* di Indonesia. Di platform Shopee, mayoritas pelanggan Toko *Online* Baju Anak Kece memiliki kecenderungan psikologis yang kuat untuk memberikan rating bintang lima disertai dengan komentar teks berisi apresiasi dan kepuasan jika barang yang diterima sesuai dengan harapan. Dampaknya, volume kosa kata bermakna positif melimpah di dalam database data latih (*training set*), sehingga algoritma KNN memiliki basis referensi vektor koordinat yang sangat kaya dan matang untuk mengenali ulasan positif baru (Pratama & Waluyo, 2024).

Sebaliknya, performa sistem dalam mengenali kategori kelas sentimen Netral menjadi titik kelemahan paling krusial

dalam pemodelan ini, dengan nilai sensitivitas (*Recall*) terendah yaitu hanya sebesar 7.69% (hanya 1 data terprediksi benar dari total 12 data aktual netral, sedangkan 10 data lainnya salah diklasifikasikan masuk ke kelas positif). Berdasarkan analisis kesalahan komputasi (*misclassification error analysis*), kelemahan mendasar ini dipicu oleh dua faktor utama. Pertama, terjadinya fenomena ketidakseimbangan jumlah sebaran dataset (*imbalanced data*), di mana jumlah sampel ulasan netral di dalam data latih sangat minoritas dibandingkan dengan kelas positif. Kedua, adanya ambiguitas semantik yang tinggi pada struktur kalimat ulasan netral di *marketplace* Indonesia. Konsumen seringkali menuliskan komentar ulasan campuran yang menggabungkan unsur ekspresi kepuasan dan keluhan minor dalam satu kalimat tunggal yang pendek, seperti contoh ulasan: "Baju anak nya lumayan bagus, tapi sayang bahan kain agak tipis ya, pengiriman standar". Ketika teks tersebut diekstrak menjadi fitur bobot numerik TF-IDF, nilai koordinat geometrisnya menjadi sangat berhimpitan dan bias dengan ruang vektor kelas sentimen positif atau negatif, sehingga saat proses *voting* mayoritas K=3 dijalankan, data netral tersebut kalah suara dan salah teridentifikasi masuk ke kelas sentimen lain. Distribusi perbandingan visual grafik batang yang menggambarkan kontras persebaran total hasil pengujian sentimen ulasan pelanggan disajikan pada Gambar 2



2. Hasil Evaluasi Performa Sistem

Sentimen	Jumlah Data Uji	Precision	Recall	F1-Score
Negatif	10	55.56%	50.0%	52.63%
Netral	13	33.33%	7.69%	12.5%
Positif	42	73.58%	92.86%	82.11%

Total Akurasi Sistem: 69.23%

Distribusi Data Uji:

Negatif: 15.38% | Netral: 20.0% | Positif: 64.62%

Gambar 2 Hasil Klarifikasi Sentimen

Tahapan puncak yang menjadi nilai tambah utama dari arsitektur sistem informasi ini adalah integrasi otomatis antara *output* klasifikasi sentimen teks dengan modul pelacakan inventaris gudang untuk mengidentifikasi produk terlaris (*best-selling products*). Modul sistem diprogram menggunakan algoritma agregasi yang secara berkala melakukan pemindaian data ulasan, mengelompokkannya berdasarkan ID produk, dan melakukan kalkulasi matematika untuk menghitung produk yang mengumpulkan volume akumulasi sentimen positif tertinggi di setiap periodenya. Peringkat produk terlaris tidak lagi hanya ditentukan secara bias lewat visualisasi total kuantitas penjualan, melainkan disaring berdasarkan kepuasan riil yang terekam pada teks komentar pelanggan.

Melalui hasil ekstraksi teks yang dilakukan oleh sistem informasi, produk-produk pakaian anak yang menempati urutan lima besar terlaris secara konsisten didominasi oleh dokumen ulasan yang memiliki frekuensi kemunculan kata kunci positif yang tinggi, seperti kata dasar: "bahan" (nyaman, adem, lembut, menyerap keringat), "warna" (sesuai, cerah, bagus), "jahit" (rapi, kuat), serta kata "harga" (murah, terjangkau, bersaing). Kosa kata dasar ini membuktikan secara empiris bahwa aspek kenyamanan material kain dan efisiensi harga merupakan variabel penentu paling sensitif yang memengaruhi tingkat kepuasan orang tua dalam membeli pakaian anak secara online. Untuk memberikan kemudahan operasional bagi pengguna, sistem ini dilengkapi dengan tampilan antarmuka berbasis web (*web-based user interface*) yang intuitif yang dibangun menggunakan *framework* Python, di mana visualisasi *dashboard* utama dan visualisasi detail pelacakan sentimen produk disajikan secara lengkap berturut-turut pada Gambar 3.

3. Identifikasi Produk Baju Anak Terlaris

Berdasarkan jumlah sentimen ulasan **Positif** terbanyak:

Peringkat	Nama Produk (Baju Anak)	Jumlah Ulasan Positif
#1	Mickey Mouse Top 1-6 Tahun Kaos Anak Perempuan Korea Style	**25** Ulasan
#2	LEGGING NECY Rib Knite Premium Usia 0-6 Tahun Leging Kriwil Anak Perempuan Gaya Korea Style	**22** Ulasan
#3	NAILA SKIRT KOREA STYLE 1-7 Tahun Rok Span Anak Rajut Kancing	**20** Ulasan
#4	Rok Anak Rajut 1-6 Tahun Rok Rajut Span Anak Perempuan Rok Anak Kekinian	**20** Ulasan
#5	Atasan Crop Top Belah Bahu Rib Premium	**16** Ulasan

Gambar 3. Hasil identifikasi produk terlaris

Dampak dan implikasi manajerial praktis dari implementasi perangkat lunak sistem informasi analisis sentimen ini memberikan keunggulan kompetitif (*competitive advantage*) yang masif bagi manajemen Toko *Online* Baju Anak Kece di tengah ketatnya persaingan industri *e-commerce*. Melalui visualisasi data yang disajikan oleh sistem, pemilik toko dapat secara instan mengubah data teks ulasan yang awalnya bersifat pasif dan menumpuk menjadi pengetahuan bisnis yang aktif dan taktis. Pemilik toko dapat mengambil keputusan strategis berbasis data (*data-driven decision making*) secara akurat dan *real-time*, meliputi: (a) Optimalisasi manajemen inventaris gudang melalui pengadaan persediaan (*restocking*) yang agresif pada model pakaian anak yang terbukti menempati peringkat sentimen positif tertinggi untuk mencegah kehilangan momentum pasar. (b) Mitigasi risiko kerugian finansial akibat penumpukan barang yang tidak laku (*deadstock prevention*) dengan cara menghentikan produksi atau pembelian produk yang mendapatkan akumulasi sentimen negatif tinggi di pasar. (c) Evaluasi kinerja pemasok kain (*supplier quality control*) secara berkala apabila ditemukan lonjakan kata kunci sentimen negatif terkait bahan baju yang kasar atau jahitan yang mudah lepas. (d) Formulasi strategi kampanye pemasaran digital yang jauh lebih efektif dan tertarget dengan cara memanfaatkan kata-kata kunci sentimen positif tertinggi hasil ekstraksi sistem sebagai materi jargon promosi iklan di media sosial, sehingga pesan iklan yang disampaikan selaras dengan persepsi kepuasan riil konsumen di pasar *e-commerce* (Hidayati & Hermawan, 2021).

D. KESIMPULAN DAN SARAN

Kesimpulan berisi rangkuman singkat atas hasil penelitian dan pembahasan. Algoritma K-Nearest Neighbor (KNN) berpadu dengan metode pembobotan TF-IDF dan tahapan *preprocessing* teks berhasil diimplementasikan untuk melakukan klasifikasi sentimen ulasan pelanggan Toko *Online* Baju Anak Kece menjadi tiga kategori dengan akurasi sebesar 69.23%. Fitur sistem mampu mendeteksi serta merekomendasikan produk terlaris berdasarkan volume sentimen positif terbanyak guna mendukung optimalisasi operasional stok barang retail (Putri et al., 2026).

Saran atau *recommendation* diisi hanya jika ada. Mengingat adanya keterbatasan performa pada kelas Netral akibat data yang tidak seimbang, disarankan bagi penelitian selanjutnya untuk menerapkan metode penanganan data *imbalanced* seperti SMOTE (*Synthetic Minority Over-sampling Technique*) serta mengeksplorasi penggunaan algoritma klasifikasi lain atau pendekatan berbasis *Deep Learning* untuk meningkatkan akurasi klasifikasi teks. selanjutnya untuk mengimplementasikan metode penanganan data *imbalanced*, seperti algoritma SMOTE (*Synthetic Minority Over-sampling Technique*) pada tahap *preprocessing*. Selain itu, disarankan pula untuk mengeksplorasi perbandingan performa menggunakan algoritma klasifikasi alternatif lainnya seperti *Support Vector Machine* (SVM), Naive Bayes, atau pendekatan berbasis arsitektur *Deep Learning* (seperti LSTM atau BERT) guna mendongkrak tingkat akurasi klasifikasi teks yang jauh lebih optimal.

E. REFERENSI

[1] Afif, R., & Nugroho, K. (2022). Analisis sentimen dan prediksi ulasan pada aplikasi Info BMKG. Jurnal Informatika, 15(2).

[2] Buntoro, G. A. (2017). Analisis sentimen calon Gubernur DKI Jakarta 2017 di Twitter. Jurnal Antisipasi, 1(1), 32–41.

[3] Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk klasifikasi transportasi bus. ILKOM Jurnal Ilmiah, 12(2), 104–111. https://doi.org/10.33096/ilkom.v12i2.539.104-111

[4] Eryadi, R. A. (2026). Comparative analysis of SVM and XGBoost classifiers with HOG features for concrete crack detection. Manuscript Submitted for Publication.

[5] Feldman, R., & Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

[6] Guntara, R. G., Kashira, F. B., Amri, T. K., Restu, L. B., & Susanto, F. R. (2024). Analisis penjualan handphone di Tokopedia dengan teknik web scraping menggunakan Python pada Google Colab. ULIL ALBAB: Jurnal Ilmiah Multidisiplin, 3(4), 69–75. https://doi.org/10.56799/jim.v3i4.3200

[7] Hidayati, N., & Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. Journal of Engineering and Applied Technology, 2(2), 85–93. https://doi.org/10.21831/jeatech.v2i2.42777

[8] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

[9] Motaz, A. (2013). Pengenalan Pemrograman Web (Vol. 2). Penerbit A.

[10] Prasetyo, E. (2012). Data Mining: Konsep dan Aplikasi Menggunakan MATLAB. Andi Offset.

[11] Pratama, R. A., & Waluyo, A. F. (2023). Implementasi sistem e-commerce berbasis mobile Android menggunakan REST API dan web service pada perusahaan percetakan. Journal of Information Technology and Computer Science.

[12] Putri, T. R. D., Sanjaya, M. R., Firdaus, M. A., & Indah, D. R. (2026). Implementasi algoritma K-Nearest Neighbors dalam analisis sentimen ulasan aplikasi Bank Aladin. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 6(2), 608–618. https://doi.org/10.57152/malcom.v6i2.2633

[13] Saifurridho, M., Martanto, M., & Hayati, U. (2024). Analisis algoritma K-Nearest Neighbor terhadap sentimen pengguna aplikasi Shopee. Jurnal Informatika Terpadu, 10(1), 21–26. https://doi.org/10.54914/jit.v10i1.1054