

## **BAB 1. PENDAHULUAN**

### **1.1 Latar Belakang**

Pasar komputer dan laptop terus menunjukkan dinamika yang stabil setelah era pandemi. Lembaga riset Gartner mencatat bahwa total pengiriman PC pada tahun 2024 mencapai 245,3 juta unit secara global, hal ini menandai fase pemulihan pasar dengan pertumbuhan 1,3% dibandingkan tahun sebelumnya (STAMFORD, 2025). Pertumbuhan ini juga berlaku pada pasar dalam negeri, laporan *International Data Corporation* (IDC) yang dikutip dari laman (communications today, 2025) menunjukkan pengiriman PC di Indonesia menembus 4,1 juta unit sepanjang tahun 2024, di mana pertumbuhannya didorong secara masif oleh permintaan segmen konsumen sebesar 16,7% (*Year-on-Year*). Tingginya tingkat adopsi teknologi ini sejalan dengan data (BPS, 2024) yang mencatat 18,06% rumah tangga di Indonesia telah memiliki komputer atau laptop, mempertegas posisi perangkat tersebut sebagai kebutuhan krusial dalam mengakses ekosistem digital. Kondisi ini turut mendorong persaingan antarmerek yang semakin ketat. Lenovo mempertahankan posisi teratas dengan pertumbuhan 7,8% pada kuartal pertama 2023/2024, sementara HP, Dell, Apple, dan Acer terus berlomba memperebutkan pangsa pasar (IDC, 2024).

Persaingan yang ketat ini membuat citra merek menjadi aset yang sangat rentan. Ulasan negatif dari konsumen terbukti menjadi salah satu faktor yang secara langsung mengancam posisi merek di pasar. Penelitian mengenai perpindahan merek pada pengguna laptop Asus menemukan bahwa ketidakpuasan yang dipicu oleh masalah performa dan baterai, yang banyak diekspresikan melalui ulasan daring dan keluhan di media sosial, mendorong 36,67% responden untuk beralih ke merek kompetitor (Meidea dkk., 2024). Ulasan konsumen dengan demikian bukan sekadar informasi pasif, melainkan faktor aktif yang mampu mempercepat pengambilan keputusan dan menggeser loyalitas terhadap suatu merek laptop.

Fenomena ini semakin relevan seiring dengan meningkatnya transaksi jual beli barang melalui platform *e-commerce*. Elektronik terbukti menjadi salah satu kategori

produk yang paling banyak dibeli secara daring oleh konsumen Indonesia dengan persentase mencapai 35%. Posisi Tokopedia sebagai salah satu platform *e-commerce* yang paling banyak digunakan di Indonesia secara otomatis menghasilkan volume transaksi dan ulasan konsumen dalam jumlah yang sangat masif. Besarnya volume ulasan ini menjadi sangat krusial mengingat 52% preferensi belanja daring konsumen didorong oleh ketersediaan testimoni pelanggan. Ulasan positif yang kredibel beserta reputasi toko telah terbukti secara signifikan memengaruhi niat beli dan membantu konsumen membuat keputusan pembelian yang lebih bijaksana untuk menghindari penyesalan (Satria & Suryandari, 2025). Data ulasan dalam volume besar di Tokopedia pada akhirnya menjadi representasi empiris yang nyata dari dinamika perilaku konsumen dalam kategori produk laptop.

Besarnya volume data ulasan yang tidak terstruktur pada *platform* seperti Tokopedia justru menghadirkan tantangan tersendiri. Analisis manual terhadap ribuan ulasan tidak lagi efisien, sehingga dibutuhkan pendekatan komputasi yang mampu mengekstrak, mengklasifikasikan, dan meringkas opini secara otomatis. Analisis sentimen berbasis *machine learning* terbukti efektif untuk membantu organisasi dan bisnis dalam memahami opini publik secara otomatis dan efisien, memantau reputasi merek, serta mengidentifikasi masalah lebih awal sebelum berdampak lebih luas (Asemi dkk., 2023; Baloch dkk., 2023; Lee dkk., 2023; Hssane dkk., 2021). Beberapa dampak positif dari analisis sentimen berbasis *machine learning*, tentu berhubungan langsung dengan penggunaan data yang akan dikelola. Data dapat berupa ulasan yang sering kali mengandung kompleksitas bahasa seperti ambiguitas, sarkasme, dan ekspresi emosional yang sulit diidentifikasi tanpa teknik komputasi yang memadai (Ranjan dkk., 2024).

Salah satu algoritma yang terbukti sangat efektif untuk klasifikasi sentimen adalah XGBoost. Algoritma ini mampu mengelompokkan data teks ke dalam kategori positif, negatif, dan netral, serta meningkatkan akurasi, *recall*, dan *F1-score* dalam berbagai domain, baik pada data teks sederhana seperti ulasan produk maupun data multimodal (Chandrasekaran, S., dkk., 2025; Samih dkk., 2023; Sarathan dkk., 2021).

Dibandingkan dengan algoritma lain seperti SVM, Naïve Bayes, atau model LLM murni, XGBoost menunjukkan performa yang lebih unggul ketika diintegrasikan dengan teknik *word embedding* seperti Sentiment2Vec atau digabungkan dengan model *deep learning* seperti LSTM (Kumar dkk., 2023; Wang, 2025). Potensi XGBoost meski demikian belum sepenuhnya dioptimalkan, terutama dalam konteks ulasan produk laptop berbahasa Indonesia yang memiliki karakteristik linguistik unik.

Performa model klasifikasi sangat dipengaruhi oleh distribusi data pelatihan, di mana kondisi kelas yang tidak seimbang (*imbalanced*) dapat menurunkan kemampuan model mengenali kelas minoritas secara akurat (Wongvorachan dkk., 2023). *Random undersampling* dapat mengurangi sampel kelas mayoritas untuk menyeimbangkan data, tetapi berisiko membuang informasi yang masih relevan bagi model (Komang Dharmendra dkk., 2024; Wongvorachan dkk., 2023). SMOTE mengatasi kelemahan tersebut dengan membuat sampel sintesis baru melalui kombinasi nilai antar data minoritas yang berdekatan, menghasilkan distribusi yang lebih seimbang tanpa duplikasi yang redundan (Budaya & Suniantara, 2024; Komang Dharmendra dkk., 2024). Penggunaan SMOTE terbukti secara konsisten meningkatkan performa *recall* dan *F1-score* pada kelas minoritas, sebagaimana ditunjukkan dalam studi klasifikasi sentimen pada ulasan *e-commerce* (Sediatmoko dkk., 2024) maupun ulasan layanan kesehatan (Budaya & Suniantara, 2024).

Berdasarkan hal tersebut, penelitian ini berfokus pada peningkatan performa model klasifikasi XGBoost melalui beberapa pendekatan sistematis, meliputi optimasi model *Hyperparameter Tuning*, penanganan ketidakseimbangan data, serta rekayasa fitur. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam membantu konsumen membuat keputusan pembelian laptop di *platform* Tokopedia dengan lebih terstruktur dan objektif, sekaligus meminimalisir bias yang sering muncul dalam penilaian manual terhadap ulasan produk.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, rumusan masalah pada penelitian ini adalah bagaimana performa metode klasifikasi XGBoost dengan optimasi *Grid Search*, penanganan *Imbalanced Data* (SMOTE), dan seleksi fitur *Chi-Square* dalam analisis sentimen ulasan produk laptop pada *platform* Tokopedia?

## 1.3 Tujuan

Berdasarkan rumusan masalah tersebut, tujuan penelitian ini adalah menemukan kombinasi hiperparameter terbaik pada model XGBoost untuk mengklasifikasikan sentimen ulasan produk laptop di Tokopedia. Optimasi model ini dilakukan dengan mengintegrasikan seleksi fitur *Chi-Square* dan penanganan ketidakseimbangan data menggunakan SMOTE, guna meningkatkan akurasi dan representativitas klasifikasi bagi calon pembeli.

## 1.4 Batasan Masalah

Batasan-batasan masalah pada penelitian ini adalah sebagai berikut.

- a. Penelitian ini berfokus pada manfaat bagi calon pembeli laptop di Tokopedia.
- b. Data ulasan hanya mencakup bahasa Indonesia.
- c. Tidak mencakup klasifikasi gambar/video dalam ulasan.
- d. Jumlah data ulasan yang dikumpulkan adalah minimal 300 - 1500 per kelas sentimen, memastikan data sampel cukup representatif dan menghindari *underfitting* (gagal mengenali pola).
- e. Batas maksimal input untuk satu kali sesi analisis sentimen adalah empat tautan produk laptop.
- f. Durasi *Scraping* data pada sistem, berkisar antara 16-20 detik/tautan, bergantung pada jumlah ulasan yang tersedia.
- g. Parameter evaluasi produk dibatasi pada beberapa aspek produk seperti, harga jual, spesifikasi layar, kapasitas baterai, dan kinerja sistem.

- h. Analisis perbandingan tidak melibatkan perangkat elektronik maupun periferal tambahan, seperti ponsel cerdas, papan ketik, tetikus, dll.

## **1.5 Manfaat**

### **1.5.1 Manfaat Pengguna**

- a. Memudahkan pengambilan keputusan melalui klasifikasi sentimen terstruktur (positif, negatif, dan netral).
- b. Menghemat waktu penilaian produk dengan sajian rangkuman rekomendasi produk laptop.
- c. Memberikan wawasan objektif tentang produk melalui pendekatan berbasis aspek produk (layar, performa, baterai, dan harga).
- d. Membantu menghindari produk tidak sesuai melalui identifikasi kekurangan produk, berbasis kecenderungan sentimen negatif pada aspek tertentu.

### **1.5.2 Manfaat Institusi Pendidikan**

- a. Kontribusi pengembangan NLP (*Natural Language Processing*) dan analisis sentimen bahasa Indonesia.
- b. Peningkatan reputasi akademik melalui penelitian inovatif.
- c. Potensi publikasi di jurnal/konferensi ilmiah.
- d. Peningkatan kompetensi lulusan di bidang analisis data.

### **1.5.3 Manfaat Penulis**

- a. Penguasaan teknik analisis sentimen dengan metode XGBoost.
- b. Referensi bagi penelitian sejenis di *e-commerce* Indonesia.
- c. Pemahaman dampak sentimen ulasan terhadap perilaku konsumen.