

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pertumbuhan pesat industri *e-commerce* secara global telah mengubah secara fundamental cara konsumen bertransaksi dan mengevaluasi produk sebelum melakukan pembelian [1]. Secara global, nilai penjualan *e-commerce* terus menunjukkan tren ekspansi yang sangat besar, di mana estimasi nilai transaksi pada tahun 2024 telah mencapai lebih dari USD 6,33 triliun atau meningkat sebesar 8,81% dibandingkan tahun sebelumnya [2]. Di skala regional dan nasional, jumlah pengguna aktif terus meroket hingga diproyeksikan menyentuh angka 99,1 juta pengguna dengan pertumbuhan mencapai 51,03% dalam lima tahun terakhir, khususnya pada sektor ritel *fashion* yang menyumbang nilai belanja hingga USD 770 juta [3]. Dalam ekosistem yang serba virtual ini di mana pembeli tidak dapat menginspeksi produk secara langsung, fitur ulasan pelanggan menjadi pilar kepercayaan paling krusial untuk memvalidasi kualitas produk sebelum transaksi dilakukan [4].

Meskipun ulasan pelanggan menyimpan wawasan yang sangat berharga, tingginya volume data yang dihasilkan setiap hari membuat proses membaca dan mengevaluasi informasi tersebut secara manual menjadi sangat tidak efisien dan memakan waktu [5]. Untuk mengatasi kendala operasional tersebut, diperlukan sistem otomatisasi analitik untuk mengklasifikasikan opini konsumen secara instan. Perkembangan teknologi kecerdasan buatan untuk analisis sentimen saat ini telah bergeser dari model tradisional berbasis teks tunggal menuju pendekatan multimodal [6]. Pendekatan mutakhir ini mengintegrasikan *Natural Language Processing* dan *Computer Vision* untuk memproses teks keluhan yang disandingkan secara paralel dengan bukti visual aktual yang diunggah oleh pembeli, sehingga evaluasi kepuasan pelanggan dapat dilakukan secara jauh lebih holistik [7].

Mengeksplorasi perkembangan tersebut, berbagai penelitian terdahulu telah berupaya mengembangkan arsitektur analisis sentimen dengan kelebihan dan celah

kelemahannya masing-masing. Sebagai permulaan, penelitian sebelumnya [8] berhasil mengklasifikasikan sentimen ulasan dengan akurasi mencapai 84,20%, namun model ini murni bersifat unimodal teks sehingga buta terhadap bukti fisik berupa gambar.

Di sisi lain, khusus dalam hal penanganan code-mixed yang tidak terstruktur, terdapat artikel penelitian terdahulu [9] yang sukses merancang model bahasa khusus untuk memproses sampel teks *code-mixed* Bengali-Inggris (Banglish) berskala lokal dengan pencapaian *F1-score* sebesar 85,50%. Sayangnya, sistem tersebut mengalami kemunduran fungsional karena tidak dilengkapi dengan kapabilitas ekstraksi gambar ulasan. Kelemahan pemrosesan bahasa campuran lintas bahasa ini juga coba ditangani oleh penelitian lain [10] yang menggunakan arsitektur multibahasa regional berbasis MuRIL yang mencetak *F1-score* teks sebesar 86,10%, namun efektivitas representasi vektornya masih terbatas pada ranah tekstunggal (*unimodal*) dan belum diuji coba untuk mengenali anomali visual pada objek barang.

Pada ranah ekstraksi fitur gambar *User-Generated Content* (UGC) yang dinamis, pemanfaatan model *Convolutional Neural Network* (CNN) klasik seperti ResNet-18 [11] mampu memperoleh akurasi visual sebesar 74,50%. Sementara itu, rumpun *Vision Transformer* (ViT) seperti DeiT-Tiny [12] mencetak skor akurasi sebesar 76,80%. Namun, arsitektur DeiT-Tiny terbukti membutuhkan volume data raksasa untuk mencapai konvergensi yang *robust*, sedangkan CNN konvensional seperti ResNet-18 sering kali gagal menangkap fitur lokal beresolusi tinggi pada foto bukti kerusakan produk yang penuh dengan derau (*noise*) [12]. Kelemahan-kelemahan dari serangkaian penelitian terdahulu tersebut menjadi sangat terasa ketika dihadapkan pada ekosistem platform *e-commerce* berskala besar seperti Daraz, yang tecermin secara langsung dari repositori dataset BanglishRev [13]. Jutaan ulasan pelanggan di dalamnya didominasi oleh teks *code-mixed* yang tidak terstruktur, penggunaan ejaan informal yang berantakan, serta kualitas unggahan foto bukti fisik barang cacat yang sangat bervariasi.

Guna mengatasi keterbatasan performa dari model-model pembandingan tersebut, penelitian terdahulu mulai melirik teknik optimasi arsitektur dan

transparansi model. Penggunaan kerangka kerja pencarian parameter otomatis seperti Optuna dengan menerapkan algoritma *Tree-structured Parzen Estimator* (TPE) pada studi [14] terbukti mampu melompati batas efisiensi penalaan parameter tradisional. Selain itu, untuk mengurai sifat *black-box* dari model berskala besar, implementasi *Explainable Artificial Intelligence* (XAI) melalui pendekatan algoritma *Integrated Gradients* pada artikel [15] mulai digunakan untuk memetakan kontribusi fitur input secara matematis. Namun, integrasi menyeluruh yang menggabungkan penalaan otomatis TPE Optuna dan evaluasi transparansi XAI pada model klasifikasi hibrida multimodal masih sangat jarang dieksplorasi.

Untuk menjawab tantangan ekstrem pada data ulasan tersebut, penelitian ini mengusulkan pengembangan dan komparasi model analisis sentimen multimodal hibrida tingkat lanjut sebagai solusi utama. Kontribusi utama dari penelitian ini berfokus pada integrasi ketangguhan model bahasa lintas-bahasa (*cross-lingual*) XLM-RoBERTa dalam membedah teks *code-mixed* Banglish, yang dikombinasikan secara paralel dengan ketajaman arsitektur modern ConvNeXt-Tiny dalam mendeteksi anomali visual produk secara lokal. Eksperimen komparatif dilakukan secara menyeluruh dengan membandingkan performa model yang diusulkan terhadap arsitektur representasi bahasa MuRIL serta model visi komputer ResNet-18 dan DeiT-Tiny. Selanjutnya, agar kombinasi arsitektur *feature-level fusion* berbasis konkatenasi ini dapat beroperasi pada performa terbaik tanpa terjebak pada proses *trial-and-error* manual, kerangka kerja Optuna dengan algoritma *Tree-structured Parzen Estimator* (TPE) diimplementasikan untuk mengotomatisasi pencarian konfigurasi *hyperparameter* secara otomatis, sementara ketidakseimbangan distribusi kelas ditangani menggunakan teknik *quality-based undersampling*. Sinergi teknologi ini dirancang untuk merumuskan sistem deteksi sentimen multimodal yang mampu saling melengkapi dalam menutupi ambiguitas informasi antar-modalitas, sekaligus menjamin transparansi akuntabilitas keputusan model melalui visualisasi peta atribusi *Explainable Artificial Intelligence* (XAI) berbasis *Integrated Gradients*.

1.2. Rumusan Masalah

Analisis sentimen berbasis visual pada *e-commerce* menghadirkan tantangan tersendiri akibat tingginya variasi kualitas gambar yang diunggah oleh pengguna. Kebutuhan akan pemrosesan visual yang mumpuni mendorong perlunya eksplorasi terhadap berbagai arsitektur *Deep Learning* untuk menemukan model dengan tingkat akurasi yang paling optimal. Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model berbasis teks (XLM-RoBERTa dan MuRIL) dalam melakukan analisis sentimen pada ulasan produk *e-commerce* yang mengandung bahasa *code-mixed*?
2. Bagaimana performa model berbasis gambar (ResNet-18, DeiT-Tiny, dan ConvNeXt-Tiny) dalam melakukan analisis sentimen berdasarkan informasi visual pada ulasan produk *e-commerce*?
3. Apakah penggabungan informasi tekstual dan visual melalui pendekatan multimodal mampu menghasilkan performa analisis sentimen yang lebih baik dibandingkan pendekatan unimodal?
4. Bagaimana pengaruh optimasi *hyperparameter* menggunakan Optuna terhadap performa model *deep learning* yang digunakan dalam analisis sentimen multimodal?
5. Bagaimana tingkat akurasi model hybrid XLM-RoBERTa dan ConvNeXt-Tiny dalam mengklasifikasikan sentimen ulasan produk *e-commerce* dan bagaimana *Explainable AI* mampu menginterpretasikan hasil prediksi tersebut?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis dan membandingkan performa model berbasis teks XLM-RoBERTa dan MuRIL dalam mengklasifikasikan sentimen ulasan produk *e-commerce* yang mengandung bahasa *code-mixed*.

2. Menganalisis dan membandingkan performa model berbasis gambar ResNet-18, DeiT-*Tiny*, dan ConvNeXt-*Tiny* dalam mengklasifikasikan sentimen berdasarkan informasi visual yang terdapat pada ulasan produk *e-commerce*.
3. Mengembangkan model analisis sentimen multimodal yang menggabungkan representasi teks dan gambar menggunakan pendekatan *feature-level fusion*, serta membandingkan performanya terhadap pendekatan unimodal.
4. Mengoptimalkan *hyperparameter* seluruh model menggunakan framework Optuna dengan algoritma *Tree-structured Parzen Estimator* (TPE) untuk memperoleh konfigurasi yang menghasilkan performa klasifikasi sentimen terbaik.
5. Mengidentifikasi kontribusi modalitas tekstual dan visual terhadap proses pengambilan keputusan model hybrid melalui pendekatan *Explainable Artificial Intelligence* (XAI) berbasis Integrated Gradients.

1.4. Urgensi Penelitian

Urgensi dari penelitian ini dapat diidentifikasi secara sistematis dengan mempertimbangkan beberapa aspek berikut:

1. Urgensi Industri

Besarnya volume ulasan *e-commerce* yang terus bertumbuh secara eksponensial menjadikan proses evaluasi ulasan secara manual sebagai pekerjaan yang mustahil dilakukan. Platform *e-commerce* membutuhkan sistem analitik otomatis berskala besar yang tidak hanya mampu mengolah satu jenis data tunggal, melainkan mampu memproses entitas data jamak yang diproduksi secara massal oleh konsumen setiap harinya, seperti teks ulasan multi-bahasa yang bersandingan secara paralel dengan unggahan bukti foto produk, guna mendukung monitoring kepuasan pelanggan secara *real-time*.

2. Urgensi Linguistik

Karakteristik ulasan pada platform *e-commerce* multibahasa didominasi oleh teks *code-mixed* (seperti campuran Bengali-Inggris atau Banglish) yang mengandung variasi ejaan informal tingkat tinggi, penggunaan abjad Latin untuk transliterasi bahasa lokal, serta kosakata *out-of-vocabulary*. Akibatnya,

model analitik berbasis *unimodal* maupun pendekatan statistik standar tidak memiliki kapasitas representasi yang memadai untuk menangkap konteks sentimen asli secara akurat.

3. Urgensi Akademik

literatur *multimodal sentiment analysis* yang berkembang saat ini sebagian besar masih terbatas pada pemrosesan bahasa tunggal berskala kecil atau korpus visual yang homogen. Terdapat kesenjangan teoretis (*theoretical gap*) mengenai bagaimana model *deep learning* mampu mengekstraksi dan mengintegrasikan hubungan semantik dari data operasional dunia nyata (*User-Generated Content*) yang memiliki kompleksitas linguistik tinggi seperti teks *code-mixed* Bengali-Inggris (Banglish) yang ditulis berbarengan dengan unggahan gambar autentik yang penuh *noise*.

4. Urgensi Teknologi

Integrasi pendekatan *deep learning* multimodal berbasis model teks dan gambar menawarkan lompatan teknologi yang krusial. Kombinasi antarmodel diharapkan dapat menyatukan kekuatan pemahaman semantik teks *code-mixed* lintas bahasa dengan pengekstraksian fitur spasial dari banyak objek gambar autentik secara bersamaan. Didukung oleh optimasi otomatis berbasis Optuna (TPE) dan interpretabilitas algoritma XAI, pendekatan hibrida ini merumuskan arsitektur yang tangguh terhadap *class imbalance* dan dapat diterapkan secara andal pada lingkungan operasional industri digital modern.

1.5. Luaran Penelitian

Luaran utama yang ditargetkan dari penelitian ini adalah sebuah artikel ilmiah berkualitas tinggi yang siap dipublikasikan pada *Applied Computer Science* (ACS), yaitu sebuah jurnal bereputasi internasional yang terindeks Scopus dengan peringkat Q2. Melalui publikasi tersebut, hasil penelitian ini diharapkan mampu memberikan kontribusi nyata dalam beberapa aspek berikut:

1. Menyebarluaskan hasil temuan kepada kalangan akademisi serta praktisi, khususnya yang bergerak di bidang *artificial intelligence* dan analisis data hibrida, sehingga dapat menjadi referensi valid bagi penelitian lanjutan maupun pengembangan aplikasi analitik di industri digital.

2. Memberikan kontribusi ilmiah yang signifikan terkait perancangan model analisis sentimen berbasis pendekatan multimodal *deep learning* yang mengintegrasikan kombinasi data teks ulasan dan foto produk aktual sebagai basis pengetahuan utama.
3. Menjadi parameter keberhasilan yang konkret dari implementasi program PROSTEP Penelitian dalam mencetak luaran akademik yang diakui secara global pada jurnal internasional bereputasi.

1.6. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang signifikan dan terukur, baik secara teoritis bagi perkembangan ilmu pengetahuan maupun secara praktis bagi penyelesaian masalah di industri. Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Penelitian Teoritis

Kontribusi akademis dari studi ini berkaitan dengan sumbangannya dalam memperluas pengetahuan terkait analisis sentimen berbasis gambar, khususnya melalui evaluasi berbagai arsitektur *Deep Learning* dan penanganan ketidakseimbangan data pada domain ulasan produk *e-commerce*. Rincian manfaat teoritis tersebut adalah:

- a. Memperkuat kajian ilmiah mengenai multimodal sentiment analysis dengan membandingkan perpaduan arsitektur *Deep Learning*, yaitu kombinasi model bahasa MuRIL dengan model *Computer Vision* (ResNet-18, DeiT, dan ConvNeXt) dalam satu kerangka evaluasi komparatif untuk tugas klasifikasi sentimen ulasan produk.
- b. Memperluas pemahaman mengenai efektivitas arsitektur hibrida dalam menangani karakteristik data ulasan *e-commerce* yang kompleks, seperti teks berbahasa campuran (*code-mixed*) yang disandingkan dengan gambar yang bervariasi dan *noisy*, sehingga memberikan kontribusi terhadap pemetaan kekuatan fusi antar-modalitas.
- c. Menghasilkan *knowledge* baru terkait efektivitas integrasi model bahasa lintas bahasa XLM-RoBERTa dan model visi komputer ConvNeXt-Tiny

dalam meningkatkan performa klasifikasi analisis sentimen multimodal, sehingga dapat menjadi rujukan ilmiah bagi penelitian lanjutan di bidang kecerdasan buatan.

- d. Memberikan kontribusi empiris pada literatur sistem informasi mengenai penerapan kerangka kerja *Knowledge Discovery in Databases* (KDD) yang dipadukan dengan optimasi *hyperparameter* Optuna (algoritma TPE) untuk menyelesaikan masalah klasifikasi pada ulasan berformat teks *code-mixed* dan gambar tak terstruktur.

2. Manfaat Penelitian Praktis

Manfaat praktis dari penelitian ini terkait dengan penerapan langsung hasil analisis terhadap kebutuhan industri, khususnya sektor *e-commerce*. Dengan menggabungkan pendekatan *multimodal sentiment analysis*, penelitian ini diharapkan mampu memberikan solusi nyata yang dapat meningkatkan kualitas layanan dan mendukung pengambilan keputusan bisnis. Adapun manfaat praktis penelitian ini adalah sebagai berikut:

- a. Membantu pemilik toko dan pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) di *e-commerce* dalam memahami sentimen konsumen dengan lebih komprehensif melalui pemrosesan keluhan teks dan bukti gambar secara otomatis, sehingga dapat mempercepat evaluasi kualitas produk dan meningkatkan strategi pelayanan.
- b. Memberikan kerangka kerja analisis yang dapat diadopsi oleh manajemen platform *e-commerce* untuk mendukung *decision-making* berbasis data, khususnya sebagai landasan objektif untuk menyaring ulasan tidak relevan dan memantau tingkat kepuasan pelanggan secara masif dan efisien..