

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam bidang rekayasa dan pemeliharaan perangkat lunak (*software engineering and maintenance*), fase pemeliharaan setelah aplikasi dirilis merupakan bagian penting dalam siklus hidup pengembangan perangkat lunak. Pada fase ini, pengembang perlu mengidentifikasi kesalahan fungsi, memperbaiki *bug*, menyesuaikan aplikasi terhadap perubahan kebutuhan pengguna, serta meningkatkan fitur agar perangkat lunak tetap dapat digunakan secara optimal. Salah satu sumber informasi yang dapat dimanfaatkan untuk mendukung proses tersebut adalah ulasan pengguna pada *platform* distribusi aplikasi seperti Google Play Store. Ulasan pengguna tidak hanya mencerminkan tingkat kepuasan, tetapi juga dapat memuat laporan gangguan, keluhan penggunaan, saran fitur, dan informasi lain yang relevan bagi proses pemeliharaan perangkat lunak [1].

Pemanfaatan ulasan pengguna menghadapi kendala karena jumlah data terus bertambah dan sebagian besar ditulis dalam bentuk teks tidak terstruktur. Ulasan sering mengandung singkatan, bahasa tidak baku, kesalahan penulisan, kata gaul, dan kalimat pendek, sehingga pembacaan dan pengelompokan secara manual membutuhkan waktu serta berpotensi menghasilkan penilaian yang tidak konsisten. Kondisi tersebut relevan pada aplikasi Picsart yang memiliki beragam fungsi penyuntingan foto, video, dan fitur berbasis kecerdasan buatan. Ulasan pengguna tidak hanya menyampaikan kepuasan atau ketidakpuasan, tetapi juga dapat memuat informasi yang dapat ditindaklanjuti (*actionable feedback*), terutama laporan kegagalan fungsi dan permintaan fitur. Dalam konteks pemeliharaan perangkat lunak, laporan *bug* berkaitan dengan kebutuhan pemeliharaan korektif, sedangkan saran fitur dapat menjadi masukan bagi pemeliharaan perfektif. Oleh karena itu, kebutuhan analisis tidak berhenti pada pemrosesan ulasan dalam jumlah besar, tetapi juga pada kemampuan membedakan polaritas sentimen dan jenis umpan balik yang memiliki nilai tindak lanjut bagi pengembang.

Analisis sentimen pada penelitian terdahulu umumnya mengelompokkan teks ke dalam dua atau tiga kelas, yaitu positif, negatif, dan netral. Pendekatan tersebut berguna untuk mengetahui kecenderungan opini, tetapi belum secara langsung memisahkan informasi operasional seperti laporan *bug* dan saran fitur. Ulasan dengan nada negatif dapat berisi kegagalan fungsi yang membutuhkan tindakan teknis, sedangkan saran fitur dapat

disampaikan tanpa menunjukkan ketidakpuasan. Atas dasar itu, penelitian ini menggunakan taksonomi hibrida yang menggabungkan tiga kelas polaritas sentimen (positif, negatif, dan netral) dengan dua kelas intensi umpan balik (laporan *bug* dan saran fitur) dalam skema *single-label multi-class classification*. Pembentukan lima kelas tersebut menimbulkan tantangan lain berupa ketidakseimbangan distribusi data. Jika kondisi ini tidak ditangani, model berpotensi lebih dominan mempelajari kelas mayoritas sehingga akurasi keseluruhan dapat terlihat tinggi tanpa menjamin keseimbangan kemampuan mengenali setiap kelas. Karena itu, penelitian mengevaluasi *class weighting* sebagai mekanisme pemberian penalti lebih besar terhadap kesalahan pada kelas yang kurang terwakili.

Selain masalah klasifikasi dan ketidakseimbangan kelas, terdapat ketidaksesuaian antara peringkat bintang dan isi tekstual ulasan. Peringkat bintang sering digunakan sebagai indikator ringkas penilaian pengguna, tetapi nilainya belum tentu identik dengan makna yang dinyatakan dalam narasi. Laporan *bug* atau keluhan dapat memperoleh peringkat tinggi, sedangkan ulasan positif dapat memperoleh peringkat rendah. Oleh sebab itu, penyaringan ulasan yang hanya didasarkan pada *rating* berpotensi melewatkan informasi penting. Penelitian ini menggunakan IndoBERT karena model tersebut membentuk representasi kata berdasarkan konteks kalimat berbahasa Indonesia, sehingga relevan untuk membedakan ulasan yang menggunakan kata serupa tetapi memiliki maksud berbeda. Sebagai contoh, frasa “tidak bisa digunakan” dapat menunjukkan laporan *bug* apabila berkaitan dengan kegagalan fungsi, tetapi dapat menjadi keluhan negatif apabila berkaitan dengan pembatasan layanan atau ketidakpuasan pengguna.

Berdasarkan penelitian terdahulu yang ditinjau, Daqiqil ID dkk. [2] dan Samudera dkk. [3] masih berfokus pada klasifikasi dua kelas sentimen, sedangkan Mursidah dkk. [4] serta Supriyadi dan Makatita [5] menggunakan tiga kelas sentimen. William dkk. [6] menekankan pengenalan emosi, Yulianti dan Nissa [7] membahas analisis sentimen berbasis aspek, dan Sinaga dkk. [8] menerapkan klasifikasi sentimen yang lebih terperinci tetapi tetap berorientasi pada polaritas. Dari studi-studi yang ditinjau tersebut, *research gap* penelitian ini bukan sekadar belum diterapkannya IndoBERT pada ulasan Picsart. Celah yang lebih spesifik adalah belum ditemukannya integrasi secara bersamaan antara: (1) klasifikasi yang memadukan polaritas sentimen dengan *actionable feedback* berupa laporan *bug* dan saran fitur; (2) penanganan distribusi kelas tidak seimbang melalui *class weighting* yang dievaluasi terhadap *baseline* dengan konfigurasi yang sama; dan (3) analisis ketidaksesuaian antara makna tekstual ulasan dan peringkat bintang aktual. Ketiga celah tersebut menjadi dasar langsung bagi rancangan eksperimen dan analisis dalam penelitian ini.

Posisi penelitian ini adalah memberikan kontribusi empiris dan metodologis melalui integrasi tiga komponen tersebut, bukan mengusulkan arsitektur *transformer* atau algoritma baru. IndoBERT digunakan sebagai model pra-latih yang di-*fine-tuning* untuk klasifikasi hibrida lima kelas; *class weighting* diuji melalui *controlled comparison* terhadap model *baseline*; sedangkan hubungan antara label tekstual dan peringkat bintang dianalisis menggunakan tabulasi silang dan Uji Kruskal-Wallis. Kebaruan integratif penelitian terletak pada penggabungan klasifikasi hibrida sentimen dan intensi, evaluasi penanganan *class imbalance*, serta analisis *text-rating discrepancy* dalam satu rancangan penelitian pada ulasan Picsart berbahasa Indonesia. Berdasarkan posisi tersebut, penelitian ini diangkat dengan judul “Klasifikasi Hibrida Sentimen dan Intensi pada Ulasan Aplikasi Picsart Menggunakan IndoBERT”.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Bagaimana kinerja model IndoBERT dalam melakukan klasifikasi hibrida sentimen dan intensi pada ulasan aplikasi Picsart berbahasa Indonesia ke dalam lima kelas, yaitu positif, negatif, netral, laporan *bug*, dan saran fitur, pada kondisi distribusi kelas yang tidak seimbang, ditinjau dari *accuracy* serta *precision*, *recall*, dan *F1-score* per kelas?
2. Bagaimana pengaruh penerapan *class weighting* terhadap keseimbangan kinerja model IndoBERT dibandingkan model *baseline* tanpa pembobotan kelas melalui pendekatan uji ablasi, ditinjau dari *macro F1-score*, *recall macro*, *precision macro*, *accuracy*, dan kinerja kelas minoritas?
3. Bagaimana pola ketidaksesuaian antara label kelas aktual (*operational ground truth*) berdasarkan isi tekstual ulasan dan peringkat bintang aktual, serta apakah terdapat perbedaan distribusi peringkat bintang yang signifikan di antara lima kelas ulasan?
4. Bagaimana hasil klasifikasi hibrida sentimen dan intensi dapat dimanfaatkan sebagai alat bantu penyaringan awal untuk mendukung identifikasi laporan *bug* dan saran fitur dalam proses pemeliharaan perangkat lunak?

1.3 Tujuan

Tujuan utama yang ingin dicapai melalui pelaksanaan penelitian ini adalah:

1. Mengevaluasi kinerja model IndoBERT dalam melakukan klasifikasi hibrida sentimen dan intensi pada ulasan aplikasi Picsart berbahasa Indonesia ke dalam lima kelas, yaitu

positif, negatif, netral, laporan *bug*, dan saran fitur, pada kondisi distribusi kelas yang tidak seimbang berdasarkan *accuracy* serta *precision*, *recall*, dan *F1-score* per kelas.

2. Menganalisis pengaruh penerapan *class weighting* terhadap keseimbangan kinerja model IndoBERT melalui perbandingan dengan model *baseline* tanpa pembobotan kelas menggunakan pendekatan uji ablasi, berdasarkan perubahan *macro F1-score*, *recall macro*, *precision macro*, *accuracy*, dan kinerja kelas minoritas.
3. Menganalisis pola ketidaksesuaian antara label kelas aktual (*operational ground truth*) berdasarkan isi tekstual ulasan dan peringkat bintang aktual, serta menguji perbedaan distribusi peringkat bintang di antara lima kelas ulasan menggunakan Uji Kruskal-Wallis H.
4. Menganalisis peran hasil klasifikasi hibrida sentimen dan intensi sebagai alat bantu penyaringan awal untuk mendukung identifikasi laporan *bug* dan saran fitur dalam proses pemeliharaan perangkat lunak.

1.4 Manfaat

Kontribusi penelitian ini bersifat empiris dan metodologis. Penelitian ini mengevaluasi dan mengintegrasikan pendekatan yang telah ada untuk menjawab permasalahan klasifikasi ulasan aplikasi. Kontribusi penelitian dirumuskan sebagai berikut:

1. Menyediakan evaluasi empiris IndoBERT pada tugas *single-label multi-class classification* dengan taksonomi hibrida lima kelas, yaitu positif, negatif, netral, laporan *bug*, dan saran fitur, sehingga evaluasi tidak hanya berfokus pada polaritas sentimen tetapi juga mencakup *actionable feedback* yang relevan bagi pemeliharaan perangkat lunak.
2. Mengevaluasi pengaruh *class weighting* terhadap ketidakseimbangan kelas melalui perbandingan terkontrol dengan model IndoBERT *baseline* menggunakan arsitektur, parameter pelatihan, dan pembagian data yang sama, sehingga trade-off antara keseimbangan kinerja antarkelas dan kinerja agregat dapat dianalisis.
3. Menyediakan analisis hubungan antara label kelas aktual berdasarkan makna tekstual ulasan dan peringkat bintang aktual melalui tabulasi silang dan Uji Kruskal-Wallis, sehingga peringkat bintang dapat dievaluasi sebagai indikator yang perlu dilengkapi dengan analisis semantik teks.

Penelitian ini diharapkan dapat memberikan manfaat yang berarti, baik dari segi akademis maupun penerapan praktis di industri, meliputi:

1. Manfaat teoritis
 - a. Memberikan kontribusi ilmiah pada bidang *Natural Language Processing* (NLP) mengenai penerapan IndoBERT untuk klasifikasi teks berbahasa Indonesia dengan taksonomi hibrida sentimen dan intensi.
 - b. Memperkaya literatur mengenai klasifikasi teks multi-kelas pada kondisi *imbalanced data* melalui evaluasi *class weighting* dan perbandingannya dengan model *baseline*.
2. Manfaat praktis
 - a. Bagi pengembang (Picsart): penelitian ini memberikan dasar pengembangan alat bantu ulasan untuk mengelompokkan laporan *bug*, saran fitur, dan pola sentimen pengguna, sehingga berpotensi mendukung proses penyaringan umpan balik dan penyusunan prioritas pemeliharaan perangkat lunak.
 - b. Bagi pengguna: hasil penelitian ini berpotensi mendukung perbaikan aplikasi yang lebih responsif melalui pemanfaatan ulasan sebagai sumber informasi bagi pengembang.
 - c. Bagi peneliti lain: menjadi referensi dalam pengembangan sistem analitik ulasan pada *platform* distribusi digital dengan mempertimbangkan aspek anomali peringkat.

1.5 Ruang Lingkup

Untuk menjaga agar investigasi penelitian tetap terfokus, terukur dan tidak menyimpang dari tujuan awal, ditetapkan ruang lingkup batasan masalah sebagai berikut:

1. *Dataset* bersumber dari ulasan aplikasi Picsart berbahasa Indonesia pada Google Play Store. Pengumpulan awal menghasilkan 428.966 ulasan, kemudian diterapkan teknik *purposive sampling* secara nonacak berdasarkan kriteria berupa ulasan dari periode 1 Januari 2023 sampai dengan 25 Maret 2026 dan panjang ulasan minimal empat kata. Data yang memenuhi kriteria selanjutnya melalui tahap prapemrosesan. Setelah itu, dilakukan teknik *disproportionate stratified random sampling* berdasarkan peringkat bintang hingga diperoleh 10.000 ulasan sebagai *dataset final*. *Dataset* tersebut digunakan untuk proses pelabelan, pelatihan, dan pengujian model IndoBERT, serta analisis hubungan antara label kelas aktual dan peringkat bintang.
2. Prapemrosesan teks (*text preprocessing*) dibatasi pada penyeragaman huruf (*case folding*), pembersihan derau (*cleansing*) dan normalisasi (*normalization*), tanpa melakukan penghapusan kata hubung (*stopword removal*) maupun pemotongan

imbuhan (*stemming*) guna menjaga integritas konteks kalimat bagi arsitektur *transformer*.

3. Pelabelan data menggunakan lima kelas target dalam skema hibrida sentimen dan intensi, yaitu positif, negatif, netral, laporan *bug*, dan saran fitur. Pelabelan awal terhadap 10.000 ulasan awalnya dilakukan dengan bantuan *Large Language Model* (LLM) berdasarkan pedoman klasifikasi hibrida lima kelas. Seluruh label awal kemudian diverifikasi dan dikoreksi oleh Penilai I, sehingga label akhir Penilai I digunakan sebagai *operational ground truth* penelitian. Reliabilitas penerapan pedoman pelabelan diuji pada tiga sampel berstrata dengan membandingkan label akhir Penilai I dan hasil pelabelan independen secara buta oleh Penilai II menggunakan Cohen's Kappa.
4. Pelatihan model IndoBERT disertai penanganan ketidakseimbangan data (*imbalanced data*) pada penelitian ini dibatasi pada teknik pembobotan kelas (*class weighting*) pada fungsi kerugian (*loss function*), tanpa melakukan manipulasi jumlah sampel data (*resampling*).
5. Dataset dibagi dengan rasio 80:20 pada tiga eksperimen menggunakan *random state* 100, 42, dan 99. Model IndoBERT *baseline* dan model IndoBERT dengan *class weighting* dibandingkan menggunakan *accuracy*, *precision macro*, *recall macro*, *macro F1-score*, *precision weighted*, *recall weighted*, dan *F1 weighted*. Selanjutnya, *classification report* per kelas dianalisis pada model IndoBERT dengan *class weighting* untuk mengetahui nilai *precision*, *recall*, dan *F1-score* pada setiap kelas. *Macro F1-score* digunakan sebagai metrik utama untuk menilai keseimbangan kinerja antarkelas.
6. Analisis ketidaksesuaian dilakukan antara label kelas aktual (*operational ground truth*) dan peringkat bintang aktual menggunakan tabulasi silang. Anomali didefinisikan sebagai ulasan laporan *bug* atau negatif yang memperoleh bintang 4 atau 5 serta ulasan positif yang memperoleh bintang 1 atau 2. Uji Kruskal-Wallis H digunakan untuk menguji perbedaan distribusi peringkat bintang di antara lima kelas ulasan.