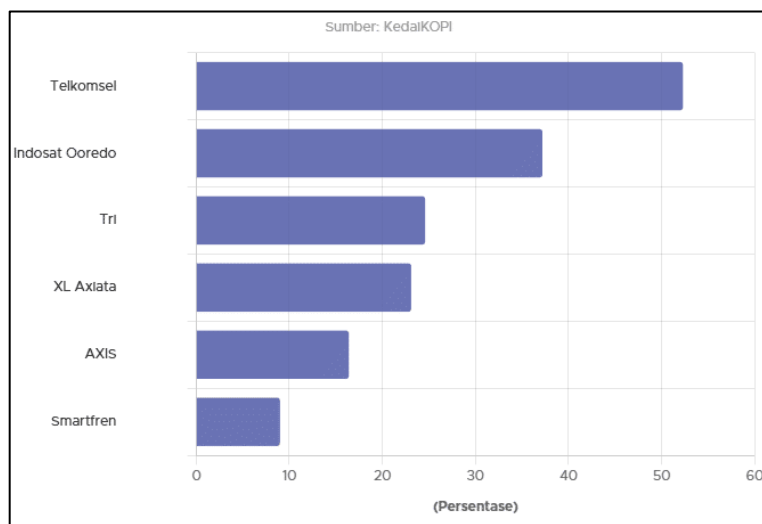


# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Industri telekomunikasi di Indonesia saat ini didominasi oleh provider besar, di antaranya yaitu Telkomsel, Indosat Ooredoo Hutchison (IM3), dan XL Axiata. Ketiganya bersaing tidak hanya dalam penyediaan jaringan, tetapi juga melalui inovasi digital yang menjadi kunci daya saing di era transformasi teknologi. Berdasarkan survei *GoodStats* pada tahun 2025, Telkomsel menempati posisi teratas dengan pangsa pasar 52,3%, diikuti Indosat dengan 37,2%, kemudian Tri dengan 24,6% dan XL Axiata dengan 23,1% [1]. Data ini mencerminkan tingginya kompetisi dalam mempertahankan dan memperluas basis pelanggan di pasar digital.



**Gambar 1.1 Provider Paling Banyak Digunakan di Indonesia 2025**

Gambar 1.1 menunjukkan pertumbuhan pesat pengguna layanan telekomunikasi yang diiringi dengan meningkatnya penetrasi internet memperkuat urgensi transformasi digital di sektor ini. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, penetrasi internet di Indonesia mencapai 72,78% dari total populasi, dan kepemilikan telepon seluler tercatat 68,65% [2]. Lebih lanjut, penelitian terkait perilaku konsumen Gen Z [3] menunjukkan bahwa generasi ini sangat bergantung pada *smartphone* dan platform digital dalam aktivitas pembelian *online*, termasuk layanan komunikasi dan paket data. Dengan perubahan perilaku tersebut, operator tidak lagi bersaing hanya dalam tarif dan jaringan, tetapi juga

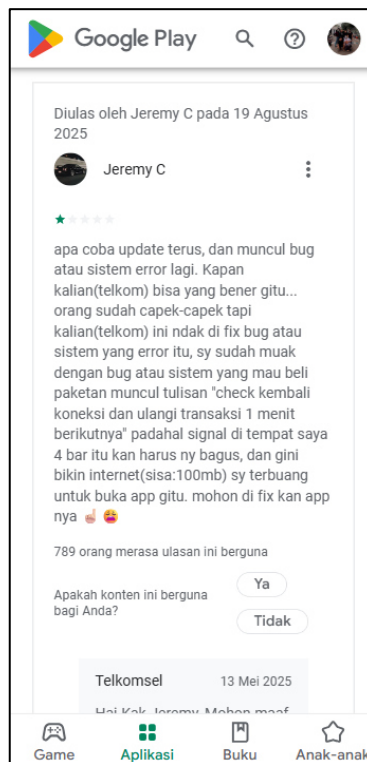
dalam pengalaman digital yang diberikan melalui aplikasi provider telekomunikasi, terutama bagi pengguna muda yang menuntut kemudahan, promo, dan personalisasi.

Laporan oleh [4] menunjukkan bahwa sekitar 75% pembelian layanan digital di Indonesia dilakukan melalui perangkat *mobile*, menandakan kanal aplikasi seluler semakin dominan dalam transaksi konsumen. Dalam konteks tersebut, semakin banyak pengguna memilih untuk membeli paket data atau layanan melalui aplikasi provider telekomunikasi karena kemudahan transaksi, promo eksklusif, dan integrasi pembayaran digital. Studi oleh [5] menunjukkan bahwa faktor *effort expectancy* dan *facilitating conditions* berpengaruh signifikan terhadap keputusan pembelian digital oleh Gen Z di Indonesia. Hal ini menandakan bahwa aplikasi tidak hanya sekadar saluran tambahan tetapi menjadi kanal utama interaksi dan transaksi, yang mampu menyajikan nilai lebih seperti diskon, *push-notification*, dan fitur loyalitas yang menarik bagi kaum Gen Z.

Sebagai bagian dari strategi digital tersebut, aplikasi provider telekomunikasi berperan sebagai antarmuka utama antara pelanggan dan layanan operator. Melalui aplikasi provider telekomunikasi seperti MyTelkomsel, MyIM3, dan MyXL pengguna dapat membeli paket data, memanfaatkan promo, melakukan *top-up*, memantau tagihan hingga menghubungi layanan pelanggan secara langsung. Keandalan, kemudahan penggunaan, dan kecepatan transaksi aplikasi menjadi faktor krusial dalam membentuk persepsi kualitas layanan. Sebuah studi oleh [6], menunjukkan bahwa kualitas layanan digital dan kemudahan penggunaan aplikasi berpengaruh signifikan terhadap kepuasan serta loyalitas pengguna jangka panjang. Dengan demikian, aplikasi yang andal dan responsif dapat menjadi faktor pembeda utama dalam industri telekomunikasi yang sangat kompetitif.

Di tengah tingginya adopsi aplikasi provider telekomunikasi, kualitas pengalaman pengguna masih menghadapi berbagai tantangan dalam hal klasifikasi emosi yang muncul dari interaksi pengguna. Berdasarkan data dari *42matters*, aplikasi MyTelkomsel, MyIM3, dan MyXL termasuk dalam tiga aplikasi dengan jumlah *download* tertinggi di Indonesia [7]. Meskipun jumlah pengguna masing-masing aplikasi cukup besar, hal tersebut tidak selalu berbanding lurus dengan tingkat kepuasan pelanggan. Studi oleh [8] mengenai persepsi publik di media

sosial mengungkapkan bahwa emosi negatif seperti *anger* mendominasi hingga 82,22% dari total respons di Facebook, disertai keluhan terkait lambatnya proses aktivasi hingga ketidaksesuaian harga dan kualitas layanan. Kondisi ini menandakan adanya kesenjangan antara ekspektasi dan pengalaman pelanggan terhadap layanan digital provider, sehingga perlu dilakukan pemodelan dan klasifikasi emosi yang lebih terstruktur untuk memetakan pola emosi pengguna.



**Gambar 1.2 Ulasan Pengguna MyTelkomsel di Google Play Store**

Masalah utama yang muncul dari kajian awal seperti pada Gambar 1.2 adalah ketidakmampuan analisis konvensional dalam menangkap variasi emosi pengguna secara lebih rinci dari teks ulasan yang kompleks, terutama yang mengandung berbagai keluhan dan pengalaman penggunaan yang beragam. Kondisi ini semakin kompleks karena ulasan pengguna sering menampilkan tantangan linguistik seperti *code-mixing* (campuran bahasa Indonesia-Inggris) dan bahasa informal, seperti “apa coba *update* terus, dan muncul *bug* atau sistem error lagi” & “mohon di *fix* kan *app* nya” yang menyulitkan pemrosesan otomatis bila model bahasa tidak dirancang untuk variasi tersebut. Oleh karena itu, kesenjangan ini memerlukan metode yang mampu melakukan klasifikasi emosi secara granular

dan mengidentifikasi topik utama yang muncul dalam ulasan pengguna sebagai bentuk eksplorasi terhadap pola diskusi yang berkembang.

Untuk menjawab tantangan bahasa dan variasi serta pemodelan topik, beberapa penelitian menunjukkan potensi model berbasis *transformer* seperti mBERT dan BERTopic dalam menangkap konteks dan variasi bahasa. Misalnya penelitian [9] melakukan *fine-tuning* mBERT untuk deteksi emosi pada bahasa Roman Urdu dan mencapai performa tinggi dimana rata-rata akurasi dan *F1-score* mencapai 91% pada dataset berlabel enam emosi dasar yaitu senang, marah, sedih, takut, jijik, dan terkejut, menunjukkan kemampuan model dalam memahami variasi linguistik yang kompleks. Sejalan dengan hal tersebut, penelitian [10] mengimplementasikan BERTopic untuk pemodelan topik pada ulasan aplikasi SATUSEHAT berbahasa Indonesia dan berhasil mengekstraksi kluster-kluster topik dominan dengan nilai koherensi 0,61. Kemampuan model BERTopic secara terstruktur dalam mengelompokkan dan mengidentifikasi keluhan spesifik pengguna menandakan bahwa arsitektur ini sangat efektif untuk memetakan konteks permasalahan dari teks ulasan aplikasi yang bervariasi. Temuan ini mendukung hipotesis bahwa perpaduan mBERT untuk klasifikasi dan BERTopic untuk pemodelan topik sangat layak diuji pada dataset ulasan aplikasi provider telekomunikasi yang penuh dengan *code-mixing*.

Sementara itu, terdapat studi-studi sebelumnya pada domain telekomunikasi menggunakan pendekatan berbasis aspek dan pemodelan topik untuk mengidentifikasi aspek layanan dan sentimen pengguna. Contohnya, penelitian [11] menganalisis sentimen pengguna aplikasi myIM3 serta mengidentifikasi topik utama menggunakan kombinasi *Support Vector Machine* (SVM) dan *Latent Dirichlet Allocation* (LDA). Penelitian tersebut menggunakan representasi TF-IDF dan menghasilkan akurasi sebesar 84,65% serta menemukan distribusi sentimen 43% positif dengan topik dominan seperti pengalaman pengguna, kegunaan aplikasi, serta masalah jaringan dan harga. Pendekatan lain [8] memanfaatkan LDA dan *NRC Lexicon* untuk mengidentifikasi dominasi emosi negatif di media sosial terhadap Telkomsel. Selain itu, penelitian [12] melakukan klasifikasi sentimen pada 39.493 ulasan MyTelkomsel menggunakan SVM, Naïve Bayes, Random Forest, dan Gradient Boosting, dengan SVM mencapai akurasi tertinggi sebesar 0,85

dimana 60,8% ulasan dikategorikan negatif. Meskipun berguna, metode-metode tersebut umumnya memisahkan pemetaan aspek dan emosi atau bergantung pada representasi kata statis, sehingga berisiko kehilangan konteks semantik pada teks yang informal dan bercampur bahasa.

Penelitian ini bertujuan membangun model klasifikasi emosi yang mengenali enam emosi utama (*anger, fear, joy, surprise, sadness, dan neutral*), serta menerapkannya pada ulasan multi-bahasa dari aplikasi MyTelkomsel, MyIM3, dan MyXL di Google Play Store. Fokus penelitian dibatasi pada ulasan dengan rating satu, tiga, dan lima bintang untuk merepresentasikan emosi negatif, netral, dan positif secara lebih jelas. Tahapan penelitian meliputi pelabelan pada dataset ulasan, dilanjutkan dengan proses *fine-tuning* model mBERT untuk klasifikasi emosi, kemudian dilakukan pemodelan topik menggunakan BERTopic untuk mengidentifikasi pola diskusi dan isu utama dalam ulasan. Pendekatan ini dipilih karena model berbasis *transformer* seperti mBERT terbukti efektif menangani teks *multilingual* yang memiliki variasi bahasa, ejaan, dan struktur kompleks, sehingga mampu menghasilkan klasifikasi emosi dan topik yang lebih mendalam.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, maka permasalahan yang dikaji dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana cara membangun model klasifikasi emosi dan pemodelan topik pada dataset multi-bahasa ulasan aplikasi provider telekomunikasi menggunakan mBERT dan BERTopic?
2. Bagaimana pengaruh variasi bahasa terhadap distribusi emosi pengguna pada aplikasi provider telekomunikasi?
3. Bagaimana cara menyajikan hasil klasifikasi emosi dan topik secara informatif sehingga pola emosi dan topik layanan pengguna dapat dipahami dengan jelas?

## **1.3 Batasan Masalah**

Untuk menjaga fokus dan keterukuran penelitian, beberapa batasan masalah telah ditetapkan sebagai berikut:

1. Data yang digunakan mencakup ulasan pengguna tiga aplikasi provider telekomunikasi dengan jumlah unduhan terbanyak di Indonesia, yaitu

MyTelkomsel, MyIM3, dan MyXL diambil dari Google Play Store yang berkisar antara tanggal 05 April 2025 hingga 05 Oktober 2025.

2. Ulasan yang diambil adalah yang memiliki rating satu, tiga, dan lima bintang, karena ulasan dengan nilai tersebut cenderung mengandung emosi yang jelas.
3. Bahasa yang dianalisis terbatas pada bahasa Indonesia dan bahasa Inggris, termasuk ulasan dengan fenomena *code-mixing*.
4. Emosi yang dianalisis dibatasi pada enam kategori utama, yaitu *anger*, *fear*, *joy*, *surprise*, *sadness*, dan *neutral*.
5. Dari dataset yang digunakan, topik yang diekstraksi merupakan kategori yang dihasilkan secara otomatis menggunakan metode BERTopic.

#### **1.4 Tujuan Penelitian**

Adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi emosi pada ulasan dataset multi-bahasa menggunakan mBERT dan mengekstrak topik layanan yang dibahas oleh pengguna menggunakan BERTopic.
2. Menganalisis pengaruh variasi bahasa terhadap distribusi emosi dan topik yang muncul dalam ulasan pengguna aplikasi provider telekomunikasi.
3. Menyajikan hasil klasifikasi emosi dan pemodelan topik secara informatif melalui visualisasi *dashboard* yang mudah dipahami.

#### **1.5 Manfaat Penelitian**

Adapun beberapa manfaat yang bisa didapatkan dari penelitian berikut:

1. Memberikan kontribusi empiris terhadap pengembangan dan evaluasi model multilingual (mBERT) dalam konteks *text mining*.
2. Memberikan informasi berbasis data bagi penyedia layanan aplikasi provider telekomunikasi tentang pola emosi pengguna dan topik layanan yang paling sering dibahas.
3. Menyajikan hasil klasifikasi emosi dan pemodelan topik melalui visualisasi informatif, sehingga pola emosi dan topik layanan dapat dipahami secara jelas.

## 1.6 Sistematika Penulisan

Penyusunan sistematika penulisan berfungsi untuk memberikan panduan dalam menyusun laporan agar jelas dan terarah. Dengan adanya struktur penulisan, proses penyusunan laporan dapat mengikuti kerangka yang konsisten dan sesuai dengan tujuan penelitian. Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

### **BAB I**

#### **PENDAHULUAN**

Bab ini berisi uraian umum tentang latar belakang yang mendasari dilakukannya penelitian ini, rumusan masalah yang nantinya akan diselesaikan, tujuan penelitian, batasan masalah, manfaat penelitian serta sistematika penulisan sebagai panduan keseluruhan isi skripsi.

### **BAB II**

#### **TINJAUAN PUSTAKA**

Bab ini membahas tentang daftar penelitian yang telah dilakukan sebelumnya serta landasan teori yang dapat menjadi pendukung penelitian.

### **BAB III**

#### **METODOLOGI PENELITIAN**

Bab ini menjelaskan rancangan tahapan penelitian dan metode yang akan digunakan.

### **BAB IV**

#### **HASIL DAN PEMBAHASAN**

Bab ini memuat hasil implementasi penelitian mulai dari pengumpulan data, pelabelan, pra-pemrosesan, hingga *fine-tuning* model mBERT serta pembahasan performa model berdasarkan metrik evaluasi dan analisis temuan utama.

### **BAB V**

#### **PENUTUP**

Bab ini berisi kesimpulan yang menjawab rumusan masalah berdasarkan hasil penelitian, disertai dengan saran pengembangan untuk penelitian selanjutnya.

### **DAFTAR PUSTAKA**

Bagian daftar pustaka memuat seluruh referensi yang digunakan dalam pengerjaan keseluruhan skripsi.

## **LAMPIRAN**

Bagian lampiran berisi dokumen pendukung penelitian yang berfungsi memberikan bukti dan transparansi terhadap proses pengerjaan skripsi.