

# Pemodelan Emosi Pengguna Berbasis Ulasan Digital Melalui Integrasi *Natural Language Processing* dan Ilmu Sosial Komputasional

Adinda Riska Safitri<sup>1\*</sup>, Andi Arniaty Arsyad<sup>1</sup>

<sup>1</sup>Informatika, Fakultas Sains dan Teknologi, Universitas Al-Azhar Indonesia, Jl. Sisingamangaraja, Kebayoran Baru, DKI Jakarta, 12110.

Penulis untuk Korespondensi/E-mail: [adindariskasafitri@if.uai.ac.id](mailto:adindariskasafitri@if.uai.ac.id)

**Abstract** - Internal company apps such as PT Astra Honda Motor's AHM Mobile receive a wide range of reviews on the Google Play Store that contain users' emotional expressions. However, the emotional dimension of these reviews has rarely been systematically analyzed, as most previous studies have focused only on positive, negative, and neutral sentiments. This study aims to analyze the emotions of AHM Mobile users through the integration of Natural Language Processing (NLP) and Computational Social Science perspectives. The dataset consists of 2.117 reviews obtained via web scraping and classified into six emotional categories: angry, sad, afraid, neutral, surprised, and happy. The annotation process was conducted by two annotators in the fields of clinical psychology and linguistics using a Seniority-Based Tie-Breaking mechanism with a Cohen's Kappa value of 0.636. Emotion classification was performed using a combination of TF-IDF and Logistic Regression as classical models, as well as IndoBERT as the main model. Evaluation results show that the classical model achieved an accuracy of 0.43 and a macro-F1 score of 0.183, while IndoBERT reached an accuracy of 0.7831 and a macro-F1 score of 0.5582. Collective emotion analysis indicates that anger dominates user reviews and is strongly correlated with user ratings, as evidenced by a Spearman correlation coefficient of 0.7141. These results indicate that sentiment analysis using the IndoBERT model can provide more effective insights for evaluating the quality of usage of internal corporate applications in Indonesia.

**Abstrak** – Aplikasi internal perusahaan seperti AHM Mobile milik PT Astra Honda Motor menerima beragam ulasan di Google Play Store yang mengandung ekspresi emosional pengguna, namun dimensi emosional pada ulasan masih jarang dianalisis secara sistematis karena sebagian besar penelitian sebelumnya hanya berfokus pada sentimen positif, negatif dan netral. Penelitian ini bertujuan untuk menganalisis emosi pengguna AHM Mobile melalui integrasi *Natural Language Processing* (NLP) dan perspektif Ilmu Sosial Komputasional. Dataset terdiri dari 2.117 ulasan yang diperoleh melalui *web scraping* dan diklasifikasikan ke dalam enam kategori emosi yaitu marah, sedih, takut, netral, terkejut dan senang. Proses anotasi dilakukan oleh dua anotator di bidang psikologi klinis dan linguistik menggunakan mekanisme *Seniority-Based Tie-Breaking* dengan nilai Cohen's Kappa sebesar 0.636. Klasifikasi emosi dilakukan menggunakan kombinasi TF-IDF dan *Logistic Regression* sebagai model klasik serta IndoBERT sebagai model utama. Hasil evaluasi menunjukkan bahwa model klasik memperoleh akurasi 0.43 dan macro-F1 0.183, sedangkan IndoBERT mencapai akurasi 0.7831 dan macro-F1 0.5582. Analisis emosi kolektif menunjukkan bahwa emosi marah mendominasi ulasan pengguna dan berkorelasi kuat dengan rating yang diberikan pengguna berdasarkan koefisien korelasi Spearman sebesar 0.7141. Hasil ini menunjukkan bahwa analisis emosi menggunakan model IndoBERT dapat memberikan informasi yang lebih efektif dalam mengevaluasi kualitas penggunaan aplikasi internal perusahaan di Indonesia.

**Keywords** - Computational Social Science (CSS), Emotion Classification, IndoBERT, Natural Language Processing (NLP), User Reviews of the AHM Mobile Application.

## PENDAHULUAN

Digitalisasi yang berkembang pesat telah mendorong perusahaan untuk berinovasi dalam menyediakan layanan digital yang lebih efisien, adaptif dan terintegrasi. Digitalisasi tidak hanya menjadi kebutuhan operasional, tetapi juga strategi penting dalam meningkatkan produktivitas, efektivitas kerja, serta kepuasan pengguna layanan digital [1]. Dalam konteks organisasi modern, sistem digital internal memegang peranan strategis karena digunakan secara langsung oleh karyawan dalam aktivitas kerja sehari-hari.

Salah satu perusahaan yang bergerak di sektor otomotif yaitu PT Astra Honda Motor (AHM) merespons perkembangan tersebut melalui pengembangan aplikasi internal AHM Mobile [2]. AHM Mobile merupakan aplikasi berbasis *mobile* yang dirancang khusus untuk karyawan PT Astra Honda Motor (AHM). Melalui aplikasi ini, karyawan dapat mengakses berbagai layanan internal perusahaan seperti pencatatan kehadiran, slip gaji elektronik, pengajuan cuti, pengelolaan informasi personal, hingga informasi seputar PT Astra Honda Motor (AHM). Ditinjau dari intensitas penggunaannya pada Google Play Store, AHM Mobile telah diunduh lebih dari 100.000 kali, namun hanya memperoleh rating 2,4 dari 5. Hal ini tercermin dari beragam pengalaman karyawan yang tertuang dalam kolom ulasan aplikasi [2].

Ulasan karyawan menunjukkan adanya dua sisi yang kontras. Di satu sisi, terdapat apresiasi terhadap kemudahan dan manfaat aplikasi dalam mendukung aktivitas sehari-hari karyawan. Di sisi lain, muncul pula keluhan terkait kendala teknis, seperti masalah *login* hingga lambatnya respons sistem. Ulasan-ulasan tersebut tidak hanya memuat evaluasi fungsional, tetapi juga dapat mengandung ekspresi emosional karyawan seperti marah, sedih, terkejut, takut, senang, maupun netral. Ekspresi emosi ini merepresentasikan respons afektif karyawan terhadap layanan digital yang mereka gunakan [2].

Berdasarkan kontras ulasan tersebut, terdapat dimensi emosional yang selama ini belum dianalisis secara sistematis oleh perusahaan. Padahal emosi yang diekspresikan karyawan dalam interaksi dengan sistem digital dapat mencerminkan tingkat kepuasan, frustrasi, maupun kepercayaan terhadap aplikasi perusahaan. Tanpa analisis yang terstruktur, informasi emosional tersebut berpotensi terabaikan sehingga peluang untuk meningkatkan kualitas

penggunaan dan pengalaman pengguna tidak dimanfaatkan secara optimal.

Dilihat dari sisi akademis, kesenjangan penelitian semakin terlihat pada studi terdahulu yang masih berfokus pada analisis sentimen berbasis polaritas sederhana. Pendekatan yang umum digunakan yaitu membatasi klasifikasi ulasan ke dalam kategori positif dan negatif, sehingga hasil analisis cenderung merepresentasikan respons pengguna secara biner. Pendekatan ini belum mampu menangkap kompleksitas emosi pengguna yang lebih beragam, seperti marah, sedih, terkejut, takut, senang, maupun netral terhadap suatu sistem digital [3].

Keterbatasan serupa juga terlihat pada berbagai penelitian yang mengklasifikasikan ulasan pengguna ke dalam kategori positif, negatif dan netral. Pendekatan ini memang mampu memberikan gambaran umum mengenai tingkat kepuasan pengguna, namun masih belum cukup menangkap variasi emosi yang lebih spesifik seperti marah, sedih, terkejut, takut, senang, maupun netral. Akibatnya, dinamika emosional pengguna belum teridentifikasi secara mendalam. Selain itu, kajian yang berfokus pada analisis emosi dalam aplikasi internal perusahaan juga masih relatif terbatas. Oleh karena itu, diperlukan pendekatan klasifikasi emosi yang lebih *fine-grained* untuk memperoleh pemahaman yang lebih komprehensif terhadap pengalaman pengguna [4-5].

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan pendekatan analisis emosi berbasis *Natural Language Processing* (NLP) untuk mengidentifikasi dan memetakan ekspresi emosional pengguna aplikasi internal perusahaan. Penelitian ini menerapkan dua pendekatan *Natural Language Processing* (NLP) yaitu metode klasik dan metode modern, guna mengklasifikasikan emosi pengguna ke dalam beberapa kategori emosi yang lebih rinci, seperti marah, sedih, takut, terkejut, senang dan netral. Pendekatan ini diharapkan mampu memberikan pemahaman yang lebih mendalam terhadap pengalaman dan persepsi pengguna terhadap layanan digital internal.

Selanjutnya, ulasan pengguna diposisikan sebagai jejak digital yang merefleksikan ekspresi sosial dan emosional karyawan dalam berinteraksi dengan sistem digital perusahaan. Oleh karena itu, penelitian ini ditempatkan dalam kerangka Ilmu Sosial Komputasional yaitu pendekatan interdisipliner yang menggabungkan kajian perilaku manusia dalam ilmu sosial dengan teknik komputasi untuk

memahami fenomena sosial dalam skala besar [6]. Melalui integrasi *Natural Language Processing* (NLP) dan perspektif Ilmu Sosial Komputasional, penelitian ini tidak hanya mengidentifikasi emosi pengguna, tetapi juga menganalisis hubungan antara emosi kolektif dengan *rating* yang diberikan oleh pengguna saat menyampaikan ulasan sebagai indikator persepsi terhadap kualitas penggunaan. Dalam penelitian ini, emosi dipahami sebagai representasi respons afektif pengguna terhadap interaksi dengan aplikasi, seperti kemudahan penggunaan hingga responsivitas aplikasi [6].

Dengan demikian, penelitian ini memberikan kontribusi praktis berupa pemanfaatan analisis berbasis data untuk memahami pengalaman penggunaan aplikasi internal oleh karyawan yang dapat menjadi dasar dalam evaluasi serta pengembangan aspek antarmuka dan fungsionalitas sistem. Selain itu, penelitian ini juga memberikan kontribusi teoritis dalam memperkaya literatur mengenai analisis emosi digital dan pemanfaat jejak digital untuk memahami dinamika interaksi pengguna pada aplikasi internal perusahaan di Indonesia.

**METODE**

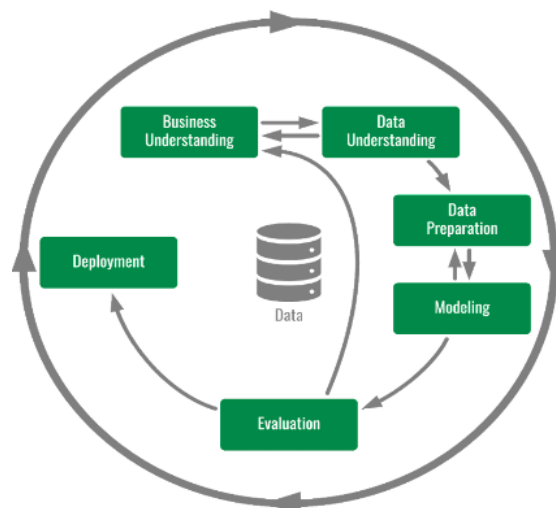
Penelitian ini menggunakan Pendekatan Kuantitatif dengan mengintegrasikan teknik *Natural Language Processing* (NLP) dan perspektif Ilmu Sosial Komputasional untuk menganalisis emosi pengguna yang terekspresikan dalam ulasan aplikasi digital.

**Kerangka Kerja Penelitian**

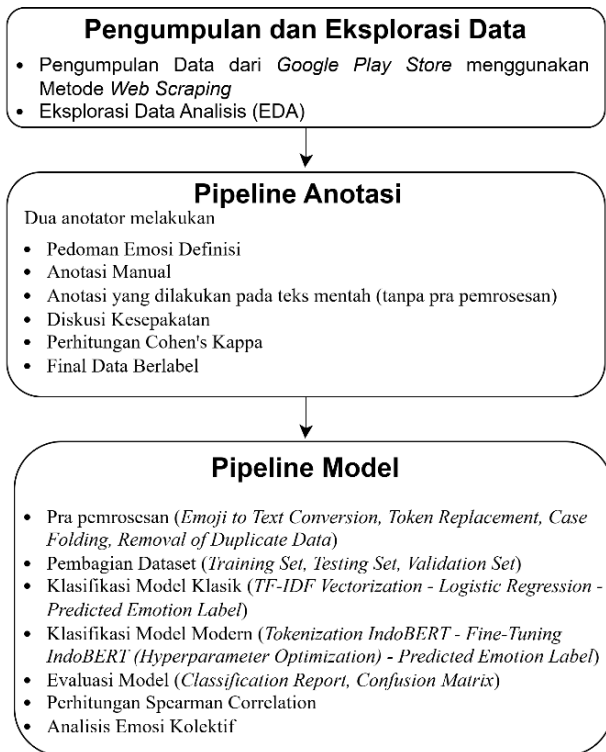
Penelitian ini menggunakan Pendekatan Kuantitatif dengan mengintegrasikan Metode *Natural*

*Language Processing* (NLP) dan perspektif Ilmu Sosial Komputasional untuk menganalisis emosi pengguna yang terekspresikan dalam ulasan aplikasi internal perusahaan. Integrasi kedua pendekatan ini tidak hanya berfokus pada aspek linguistik dan klasifikasi otomatis, tetapi juga pada pemaknaan emosi sebagai fenomena sosial yang muncul dari interaksi kolektif pengguna dengan sistem digital. Kerangka metodologis yang digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang dipilih karena menyediakan alur kerja analisis data yang sistematis, iteratif dan fleksibel, sehingga sesuai untuk penelitian berbasis data ulasan yang bersifat dinamis dan tidak terstruktur [7].

Dalam penelitian ini, kerangka kerja CRISP-DM pada Gambar 1, diadaptasi secara kontekstual untuk mendukung analisis emosi ulasan aplikasi sebagai bentuk interaksi sosial digital [7]. Tahap *business understanding* berperan dalam merumuskan permasalahan dan tujuan analisis emosi pengguna, *data understanding* difokuskan pada eksplorasi karakteristik dan kualitas data ulasan, sedangkan *data preparation* mencakup proses penyiapan data teks dan pelabelan emosi sebagai dasar pemodelan. Tahap *modeling* digunakan untuk membangun model klasifikasi emosi yang kemudian dievaluasi pada tahap *evaluation* guna menilai kinerja dan keandalan model. Terakhir, tahap *deployment* diarahkan pada penyajian temuan analisis dan interpretasi emosi kolektif pengguna sebagai dasar pemahaman persepsi sosial terhadap kualitas penggunaan aplikasi [6]. Adaptasi CRISP-DM ini memastikan bahwa proses analisis emosi dilakukan secara terstruktur sekaligus memungkinkan penyesuaian metodologis sesuai dengan kebutuhan penelitian.



Gambar 1. CRISP-DM



Gambar 2. Tahap Penelitian

### Pengumpulan dan Eksplorasi Data

Data penelitian berupa ulasan pengguna aplikasi yang dikumpulkan dari Google Play Store menggunakan metode *web scraping* dengan pustaka Python *google-play-scraper* (Gambar 2). Pengambilan data difokuskan pada ulasan berbahasa Indonesia dan diurutkan berdasarkan ulasan tanggal 18 April 2017 hingga 8 Desember 2025. Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk mengevaluasi kualitas data, meliputi pemeriksaan kelengkapan data, distribusi temporal ulasan, distribusi *rating*, serta identifikasi potensi *noise* seperti data duplikasi. Tahap ini bertujuan memastikan data layak digunakan pada proses analisis lanjutan.

### Pipeline Anotasi

Dalam memperoleh dataset berlabel yang valid dan reliabel, dilakukan proses anotasi emosi secara manual oleh dua *annotator* dibidang psikologi klinis dan linguistik. Anotasi didasarkan pada pedoman emosi yang disusun bersama dengan mempertimbangkan aspek psikolinguistik [8], semantik [9] dan sintaksis [10]. Enam kategori emosi yang digunakan dalam penelitian ini meliputi marah, sedih, takut, netral, terkejut dan senang.

Proses anotasi dilakukan langsung pada teks mentah tanpa melalui tahapan *pre-processing* dengan tujuan menjaga keutuhan dan konteks asli ekspresi

emosional yang disampaikan pengguna. Dalam mengevaluasi konsistensi pelabelan antar-*annotator*, digunakan Metode *Cohen's Kappa* [11] sebagai ukuran reliabilitas yang mempertimbangkan kemungkinan kesepakatan yang terjadi secara acak. Perhitungan dilakukan secara global menggunakan rumus sebagai berikut (1).

$$Kc = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (1)$$

$\Pr(a)$  merupakan probabilitas kesepakatan aktual antar-*annotator* dan  $\Pr(e)$  merupakan probabilitas kesepakatan yang terjadi secara acak. Nilai Kappa ( $Kc$ ) berkisar antara nilai 0 hingga 1, berarti semakin mendekati nilai 1 menandakan bahwa tingkat kesepakatan antar penilai semakin kuat.

Apabila setelah perhitungan *Cohen's Kappa* masih terdapat perbedaan label yang tidak dapat diselesaikan melalui diskusi, maka diterapkan mekanisme *Seniority-Based Tie Breaking* yaitu penentuan label akhir mengikuti keputusan *annotator* senior. Tahap ini memastikan bahwa setiap data memiliki satu label final yang telah melalui proses validasi, sehingga dataset berlabel yang dihasilkan dapat digunakan sebagai *ground truth* yang reliabel dalam proses pelatihan dan evaluasi model [12].

### Pipeline Model

Model pipeline pada penelitian ini terdiri atas tiga tahapan utama yaitu *pre-processing*, *data splitting* klasifikasi emosi, serta evaluasi dan analisis emosi kolektif. Sebelum proses pemodelan, data teks melalui tahap pra-pemrosesan untuk meningkatkan konsistensi fitur dan mengurangi *noise*. Tahapan pra-pemrosesan meliputi konversi emoji ke bentuk teks, baik untuk emoji Unicode maupun emotikon non-Unicode, guna mempertahankan makna emosional [13], penggantian pola tanda baca seperti pengulangan tanda seru dan titik menjadi token teks yang terstandarisasi [14], penerapan *case folding* dengan mengubah seluruh teks menjadi huruf kecil [15] serta penghapusan data duplikasi untuk mencegah bias dalam proses pelatihan model [16].

Proses klasifikasi emosi dilakukan menggunakan dua pendekatan untuk memungkinkan perbandingan antara Metode *Natural Language Processing* (NLP) klasik dan modern dengan pembagian dataset menggunakan *stratified split*. Pendekatan klasik digunakan sebagai model *baseline* dengan menerapkan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) yang dibatasi

hingga 3.000 fitur guna menjaga keseimbangan antara kemampuan representasi dan efisiensi komputasi. Vektor TF-IDF selanjutnya digunakan sebagai masukan bagi algoritma *Logistic Regression* yang dipilih karena sifatnya yang sederhana, mudah diinterpretasikan dan umum digunakan dalam tugas klasifikasi teks.

Pendekatan modern menggunakan Model *transformer* IndoBERT (*indobenchmark/indobert-base-p1*) yang telah dilatih pada korpus berbahasa Indonesia. Teks ulasan di tokenisasi menggunakan *tokenizer* IndoBERT dengan panjang maksimum 128 *token*, kemudian dilakukan proses *fine-tuning* dengan menambahkan lapisan klasifikasi untuk memetakan representasi teks ke dalam enam kelas emosi. Proses pelatihan model dilakukan menggunakan *optimizer* Adam dengan *learning rate* sebesar  $2e-05$ , *batch size* 32, dan enam *epoch*, untuk mengatasi ketidakseimbangan kelas emosi, nilai *macro-F1* digunakan sebagai matriks utama dalam penerapan *early stopping* dengan *patience* 2.

Evaluasi performa model dilakukan pada data uji menggunakan matriks *Accuracy*, *Precision*, *Recall*, *F1-score* [17] serta *Confusion Matrix* [18] untuk menilai kinerja klasifikasi secara keseluruhan dan pada masing-masing kelas emosi.

Secara matematis, *Accuracy* mengukur proporsi keseluruhan prediksi yang benar terhadap total data uji, yang dirumuskan sebagai berikut (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

*Precision* mengukur ketepatan model dalam memprediksi suatu kelas tertentu, yaitu proporsi prediksi positif yang benar terhadap seluruh prediksi positif yang dirumuskan pada persamaan (3).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

*Recall* mengukur kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam suatu kelas, dirumuskan pada persamaan (4).

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Sedangkan *F1-score* merupakan rata-rata harmonik antara *Precision* dan *Recall* yang digunakan untuk memberikan keseimbangan antara kedua matriks tersebut, terutama pada kondisi distribusi kelas yang tidak seimbang, dirumuskan pada persamaan (5).

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Sementara itu, *Confusion Matrix* merupakan representasi tabular yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan label aktual dan label prediksi (Tabel 1). Dalam bentuk biner, matriks ini terdiri dari empat komponen utama yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN) dan *False Negative* (FN).

Tabel 1. *Confusion Matrix*

<i>Predicted</i>	<i>Actual</i>	
	<i>True</i>	<i>False</i>
<i>True</i>	TP ( <i>True Positive</i> )	FP ( <i>False Positive</i> )
<i>False</i>	TN ( <i>True Negative</i> )	FN ( <i>False Negative</i> )

*True Positive* (TP) menunjukkan jumlah data yang benar-benar termasuk dalam kelas positif dan berhasil diprediksi dengan tepat sebagai positif oleh model. Sebaliknya *False Positive* (FP) merupakan jumlah data yang sebenarnya termasuk dalam kelas negatif, namun secara keliru diprediksi sebagai positif.

Sementara itu, *True Negative* (TN) merepresentasikan jumlah data yang memang termasuk dalam kelas negatif dan berhasil diprediksi dengan benar sebagai negatif. Adapun *False Negative* (FN) menunjukkan jumlah data yang sebenarnya positif, tetapi diprediksi secara keliru sebagai negatif oleh model.

Keempat komponen tersebut menjadi dasar dalam perhitungan berbagai matriks evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*. Dengan memahami distribusi nilai TP, FP, TN dan FN, peneliti dapat mengidentifikasi pola kesalahan model secara lebih spesifik, termasuk kecenderungan model dalam menghasilkan prediksi positif yang berlebihan atau kegagalan dalam mendeteksi kelas tertentu.

Selanjutnya, analisis emosi kolektif dilakukan dengan memetakan hasil prediksi emosi ke dalam skala ordinal [19] yang merepresentasikan tingkat ketidakpuasan hingga kepuasan pengguna. Hubungan antara skor emosi kolektif dan *rating* aplikasi dianalisis menggunakan *Spearman Correlation* yang menunjukkan pengukuran hubungan berbasis peringkat tanpa asumsi linearitas atau distribusi data tertentu yang memungkinkan pengukuran hubungan berbasis peringkat (*rank-*

based correlation) tanpa asumsi linearitas maupun distribusi normal pada data [20]. Metode ini sesuai digunakan ketika data bersifat ordinal atau tidak memenuhi asumsi parametris.

Koefisien *Spearman Correlation* ( $\rho$ ) dihitung menggunakan rumus (6).

$$\rho = 1 - \frac{(6\sum d_i^2)}{(n(n^2 - 1))} \quad (6)$$

$d_i$  merupakan selisih antara peringkat (*rank*) skor emosi dan peringkat *rating* pada observasi ke- $i$ , sedangkan  $n$  merupakan jumlah pasangan data yang dianalisis. Nilai koefisien ( $\rho$ ) berada pada rentang -1 hingga +1. Apabila nilai +1 menunjukkan hubungan positif yang kuat, nilai mendekati -1 menunjukkan hubungan negatif yang kuat, dan nilai mendekati 0 menunjukkan tidak adanya hubungan monotonik yang signifikan antara kedua variabel. Analisis ini memberikan pemahaman mengenai keterkaitan antara ekspresi emosi pengguna secara kolektif dan persepsi mereka terhadap kualitas penggunaan aplikasi.

## HASIL DAN PEMBAHASAN

Implementasi kerangka analisis emosi pada penelitian ini menghasilkan pendekatan terstruktur untuk mengidentifikasi dan menginterpretasikan ekspresi emosional pengguna dalam ulasan aplikasi AHM Mobile berbasis *Natural Language Processing* (NLP) dan Ilmu Sosial Komputasional. Penelitian ini menerapkan model kontekstual IndoBERT sebagai pendekatan utama dalam klasifikasi emosi, dengan metode klasik berbasis TF-IDF dan *Logistic Regression* digunakan sebagai *baseline* pembandingan.

### Karakteristik Dataset

Dataset penelitian diperoleh melalui *web scraping* yang merupakan teknik otomatis untuk mengekstraksi data dari halaman *web* sehingga data daring dapat dikumpulkan dan diolah menjadi dataset terstruktur untuk kebutuhan analisis penelitian [21]. Berdasarkan metode tersebut, penelitian ini memperoleh 2.117 ulasan pengguna aplikasi AHM Mobile pada *Google Play Store* dan hingga 8 Desember 2025. Struktur dataset terdiri dari beberapa atribut dengan kolom utama yang digunakan dalam analisis meliputi waktu ulasan "at", *rating* pengguna "score" dan teks ulasan "content" sebagai sumber utama analisis emosi berbasis *Natural Language Processing* (NLP).

Distribusi ulasan menunjukkan fluktuasi partisipasi pengguna sepanjang periode 2017-2025, dengan peningkatan signifikan pada tahun 2019 dan penurunan relatif pada tahun-tahun terakhir. Variasi temporal ini merefleksikan dinamika interaksi dan intensitas pengalaman pengguna terhadap aplikasi. Dataset tersebut selanjutnya menjadi dasar dalam proses anotasi dan pemodelan klasifikasi emosi.

### Karakteristik Anotasi Emosi

Tahap anotasi emosi dilakukan untuk membangun *ground truth* yang merepresentasikan ekspresi emosional pengguna secara sistematis. Proses ini mengacu pada *Emotion Guidelines* yang disusun melalui diskusi terstruktur antara dua *annotator* dengan mempertimbangkan landasan teori emosi dan variasi konteks linguistik dalam ulasan. Pedoman tersebut menghasilkan enam kategori emosi utama yaitu marah, sedih, takut, netral, terkejut dan senang yang masing-masing memiliki definisi konseptual serta indikator kebahasaan khas (Tabel 2).

Tabel 2. Karakteristik Anotasi Emosi

Emosi	Contoh Kosakata	Karakteristik Emosi
Marah	tidak jelas, aneh, goblok, lemot dan sebagainya.	Respons negatif terhadap hambatan atau perlakuan yang merugikan, ditandai kosakata konfrontatif, bernada keras, dan ekspresif. Respons terhadap kehilangan,
Sedih	sering <i>error</i> , <i>logout</i> sendiri dan sebagainya.	direpresentasikan melalui bahasa reflektif, naratif, dan bernuansa pasrah. Antisipasi terhadap ancaman atau risiko, ditandai ungkapan kekhawatiran dan penekanan pada kemungkinan konsekuensi negatif.
Takut	<i>password diupdate</i> justru tidak bisa masuk dan sebagainya.	Teks informatif atau deskriptif tanpa dominasi muatan afektif tertentu.
Netral	Umumnya informatif, deskriptif, faktual, instruksional, Kirain, ternyata, baru tahu, kok bisa, loh, bukan cuma saya dan sebagainya.	Reaksi terhadap peristiwa tak terduga; diekspresikan melalui ungkapan spontan atau interjeksi singkat.
Senang	Senang, puas, mantap, keren, bagus, sangat bermanfaat dan sebagainya.	Respons terhadap kepuasan atau keberhasilan, ditunjukkan melalui kosakata afirmatif dan evaluasi positif.

Setiap kategori emosi memiliki karakteristik linguistik yang berbeda dan digunakan sebagai dasar dalam proses anotasi manual. Emosi marah ditandai oleh kosakata konfrontatif dan bernada keras sebagai respons terhadap hambatan atau perlakuan yang

merugikan. Emosi sedih direpresentasikan melalui bahasa reflektif dan bernuansa pasrah akibat kehilangan. Emosi takut muncul melalui ungkapan kekhawatiran terhadap risiko atau konsekuensi negatif. Emosi senang ditunjukkan melalui evaluasi positif dan kosakata afirmatif yang mencerminkan kepuasan. Emosi terkejut diekspresikan dalam bentuk ungkapan spontan terhadap peristiwa tak terduga, sedangkan emosi netral bersifat informatif dan deskriptif tanpa dominasi muatan afektif tertentu. Karakteristik ini menjadi landasan konseptual dalam memastikan konsistensi pelabelan emosi pada seluruh dataset.

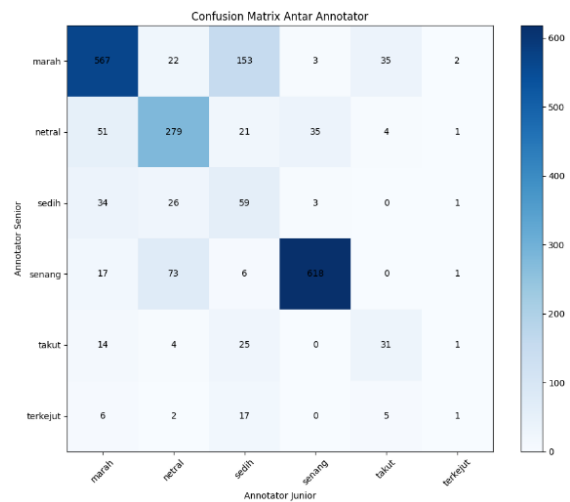
Tabel 3. Distribusi Label Emosi

Emosi	Jumlah Label Emosi
Marah	782
Sedih	715
Takut	391
Netral	123
Terkejut	75
Senang	31

Distribusi jumlah data pada setiap kategori emosi menunjukkan variasi yang cukup signifikan (Tabel 3). Emosi marah menjadi kategori dengan jumlah data terbanyak yaitu sebanyak 782 ulasan, diikuti oleh emosi senang sebanyak 715 ulasan. Selanjutnya, emosi netral tercatat sebanyak 391 ulasan. Sementara itu, emosi sedih berjumlah 123 ulasan, emosi takut sebanyak 75 ulasan dan emosi terkejut menjadi kategori dengan jumlah paling sedikit yaitu 31 ulasan. Pola distribusi tersebut menunjukkan adanya ketidakseimbangan kelas pada data berlabel. Emosi marah dan sedih mendominasi sebagian besar dataset, sedangkan emosi terkejut dan senang memiliki proporsi yang relatif kecil. Kondisi ini mengindikasikan bahwa ekspresi emosional pengguna dalam ulasan cenderung terkonsentrasi pada emosi dengan muatan evaluatif yang kuat, khususnya dalam bentuk respons negatif.

Dalam mengevaluasi konsistensi pelabelan antar-annotator, dilakukan perhitungan *Cohen's Kappa* secara global. Hasil perhitungan menunjukkan nilai *Cohen's Kappa* sebesar 0.636, yang termasuk dalam kategori baik. Nilai ini mengindikasikan bahwa tingkat kesesuaian label antar-annotator tidak terjadi secara kebetulan, melainkan didasarkan pada penerapan pedoman anotasi yang konsisten serta pemahaman yang relatif selaras terhadap definisi setiap kategori emosi. Dengan tingkat kesepakatan tersebut, hasil anotasi dapat dinilai cukup reliabel untuk digunakan sebagai *ground truth* dalam proses pelatihan dan evaluasi model klasifikasi emosi.

Interpretasi terhadap nilai *Cohen's Kappa* tersebut diperkuat melalui analisis *confusion matrix* antar-annotator yang memberikan gambaran rinci mengenai pola kesesuaian dan perbedaan label pada setiap kategori emosi (Gambar 3). Nilai pada diagonal utama yang relatif tinggi pada beberapa kelas menunjukkan bahwa sebagian besar ulasan memperoleh label yang sama dari kedua *annotator*. Dengan demikian, kombinasi antara nilai *Cohen's Kappa* sebesar 0.636 dan pola kesesuaian yang terlihat pada *confusion matrix* mempertegas bahwa proses anotasi telah dilakukan secara konsisten dan dapat dipertanggungjawabkan secara metodologis.



Gambar 3. Confusion Matrix Anotasi Emosi Antar Annotator

### Klasifikasi Model Klasik

Klasifikasi emosi ulasan pengguna aplikasi AHM Mobile, menggunakan model *baseline* yang terdiri dari metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai ekstraksi fitur dan *Logistic Regression* sebagai algoritma klasifikasi. Model *baseline* digunakan sebagai acuan awal untuk menggambarkan performa dasar klasifikasi emosi sebelum diterapkannya model IndoBERT pada sub bab selanjutnya, untuk mengevaluasi performa model *baseline* secara kuantitatif, dilakukan pengukuran menggunakan *classification report* (Tabel 4) yang mencakup matriks *precision*, *recall*, *F1-score* dan akurasi pada setiap kategori emosi.

Tabel 4. Hasil *Classification Report* Model Klasik

	Precision	Recall	F1-Score	Support
Marah	0.474	0.682	0.559	107.00
Netral	0.333	0.093	0.145	54.00
Sedih	0.000	0.000	0.000	9.00
Senang	0.387	0.403	0.395	72.00
Takut	0.000	0.000	0.000	6.00
Terkejut	0.000	0.000	0.000	1.00
Accuracy	0.430	0.430	0.430	0.43
Macro avg	0.199	0.196	0.183	249.00

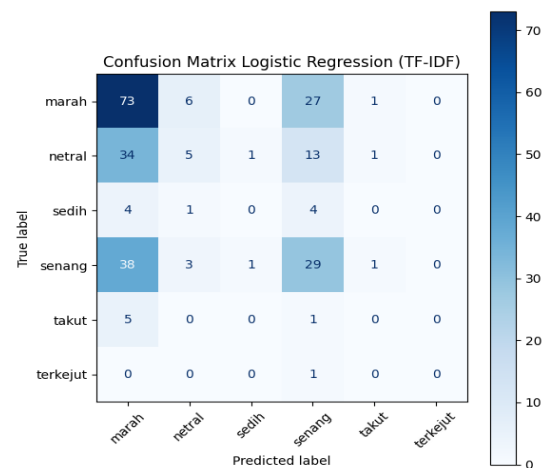
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Weighted avg</i>	0.388	0.430	0.386	249.00

Hasil klasifikasi menggunakan pendekatan TF-IDF dan *Logistic Regression* menunjukkan performa yang relatif rendah, khususnya dalam skenario multi-kelas dengan distribusi data yang tidak seimbang. Model memperoleh akurasi sebesar 0.43, yang menunjukkan bahwa kurang dari setengah data uji berhasil diklasifikasikan dengan benar. Perbedaan yang cukup besar antara *macro average F1-Score* sebesar 0.183 dan *weighted average F1-Score* sebesar 0.386 yang mengindikasikan bahwa kinerja model lebih dipengaruhi oleh kelas mayoritas, sementara performa pada kelas minoritas sangat terbatas. Secara kategori, emosi marah menunjukkan performa terbaik dengan *recall* sebesar 0.682 dan *F1-Score* sebesar 0.559 yang disebabkan oleh keberadaan indikator linguistik yang eksplisit dan bernada negatif sehingga mudah dikenali oleh model berbasis frekuensi kata. Emosi senang menunjukkan performa sedang dengan nilai *recall* 0.403 dan *F1-Score* 0.395, namun masih mengalami kesulitan akibat variasi ekspresi kepuasan yang tidak selalu eksplisit. Sebaliknya, emosi netral memiliki performa rendah dengan nilai *recall* 0.093 dan *F1-Score* 0.145 karena tidak memiliki ciri leksikal khas dan sering tumpang tindih dengan kosakata bernuansa keluhan. Emosi sedih, takut dan terkejut tidak berhasil diprediksi dengan baik yang tercermin dari nilai evaluasi nol, terutama akibat keterbatasan jumlah data serta sifat ekspresi yang lebih halus dan kontekstual. Secara keseluruhan, hasil ini menunjukkan bahwa model berbasis fitur leksikal memiliki keterbatasan dalam menangkap kompleksitas dan ambiguitas ekspresi emosi pada ulasan pengguna, khususnya pada kelas dengan representasi data yang kecil.

Dalam melengkapi analisis berdasarkan *classification report*, dilakukan evaluasi lebih lanjut menggunakan *confusion matrix* guna memperoleh gambaran yang lebih rinci mengenai pola kesalahan dan kesesuaian prediksi pada setiap kategori emosi.

Berdasarkan *confusion matrix*, model TF-IDF dan *Logistic Regression* menunjukkan kemampuan yang relatif lebih baik dalam mengklasifikasikan emosi marah dibandingkan kategori lainnya, dengan 73 prediksi benar, meskipun masih terdapat kesalahan klasifikasi ke emosi senang dan netral (Gambar 4). Emosi senang dan netral menunjukkan tingkat kesalahan yang cukup tinggi, sebagian besar ulasan pada kedua kategori tersebut salah diklasifikasikan sebagai marah. Hal ini mengindikasikan bahwa

model cenderung bias terhadap kelas dominan serta kesulitan membedakan ekspresi emosional yang ambigu atau tidak eksplisit. Performa pada emosi sedih, takut dan terkejut sangat rendah, bahkan nol, dipengaruhi oleh jumlah data uji yang terbatas serta karakteristik ekspresi yang lebih halus dan kontekstual. Secara keseluruhan, hasil ini menunjukkan bahwa pendekatan berbasis fitur leksikal lebih efektif dalam mengenali emosi dengan indikator linguistik yang eksplisit dan berintensitas tinggi, namun mengalami keterbatasan dalam menangkap kompleksitas makna emosional yang implisit. Hal ini memperkuat kebutuhan akan model berbasis pemahaman kontekstual, seperti IndoBERT yang lebih mampu merepresentasikan hubungan semantik dan konteks kalimat secara komprehensif untuk meningkatkan performa klasifikasi emosi.



Gambar 4. *Confusion Matrix* Model Klasik

**Hasil Klasifikasi Model Modern**

Klasifikasi emosi yang digunakan adalah model IndoBERT dengan arsitektur *indobenchmark/indobert-base-p1* yaitu model *transformer* yang dilatih menggunakan korpus berbahasa Indonesia. Model ini dipilih karena kemampuannya dalam menangkap representasi kontekstual teks berbahasa Indonesia secara lebih mendalam dan komprehensif dibandingkan pendekatan klasik berbasis fitur statistik, untuk menilai efektivitas model IndoBERT dalam melakukan klasifikasi emosi, dilakukan evaluasi menggunakan *classification report* yang mencakup matriks *precision*, *recall*, *F1-score* dan akurasi pada setiap kategori emosi (Tabel 5).

Tabel 5. Hasil *Classification Report* Model Modern

<i>Epoch</i>	<i>Accuracy</i>	<i>Macro-F1</i>	<i>Validation Took</i>	<i>Patience</i>
1	0.7430	0.3542	0:00:02	-
2	0.7229	0.4250	0:00:02	-
3	0.7590	0.4888	0:00:02	-

Epoch	Accuracy	Macro-F1	Validation Took	Patience
4	0.7831	0.5582	0:00:02	-
5	0.7711	0.5361	0:00:02	1/2
6	0.7631	0.5291	0:00:02	2/2

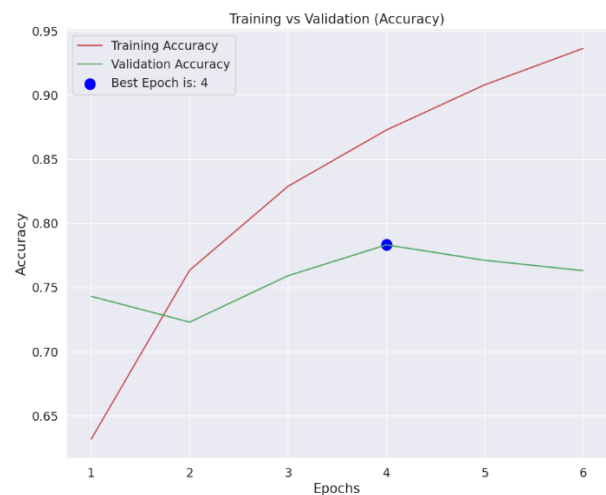
Model modern dievaluasi selama enam *epoch* pelatihan menggunakan matriks *accuracy* dan *macro-F1* (Tabel 5). Waktu validasi pada setiap *epoch* tercatat stabil, menunjukkan proses evaluasi yang efisien dan konsisten. Pada *epoch* pertama, model telah menunjukkan performa awal yang cukup baik dengan *accuracy* sebesar 0.7430, meskipun nilai *macro-F1* masih rendah sebesar 0.3542 yang mengindikasikan bahwa model belum mampu mengenali seluruh kelas emosi secara seimbang, khususnya pada kelas minoritas.

Performa meningkat secara bertahap pada *epoch* kedua dan ketiga, ditandai dengan kenaikan *macro-F1* hingga 0.4888 dan *accuracy* hingga 0.7590 yang menunjukkan bahwa model mulai menangkap representasi kontekstual kalimat secara lebih efektif. Performa terbaik dicapai pada *epoch* keempat dengan *accuracy* sebesar 0.7831 dan *macro-F1* sebesar 0.5582 yang mencerminkan keseimbangan optimal antara ketepatan klasifikasi dan pemerataan performa antar kelas. Setelah itu, pada *epoch* kelima dan keenam, terjadi penurunan nilai *accuracy* dan *macro-F1* yang mengindikasikan mulai berkurangnya kemampuan generalisasi model.

Mekanisme *early stopping* aktif pada *epoch* kelima dan mencapai batas maksimum pada *epoch* keenam, sehingga pelatihan dihentikan untuk mencegah *overfitting*. Dengan demikian, model pada *epoch* keempat dipilih sebagai model terbaik untuk tahap evaluasi dan analisis selanjutnya.

Dalam hal memperkuat hasil *classification report*, dilakukan analisis tambahan menggunakan grafik *training* dan *validation accuracy* guna mengamati pola pembelajaran model serta mengevaluasi kemampuan generalisasinya.

Grafik pada Gambar 5 menunjukkan bahwa akurasi data latih meningkat secara konsisten dari *epoch* pertama hingga *epoch* keenam yang mengindikasikan bahwa model IndoBERT semakin mampu mempelajari representasi semantik dari data ulasan. Sementara itu, akurasi data validasi meningkat hingga mencapai titik optimal pada *epoch* keempat sebelum mengalami penurunan ringan pada *epoch* berikutnya. Pola ini menunjukkan bahwa kemampuan generalisasi model berada pada kondisi terbaik pada *epoch* keempat, sejalan dengan hasil *classification report*.



Gambar 5. Grafik *Training vs Validation (Accuracy)*

Perbedaan yang semakin besar antara akurasi *training* dan *validation* pada *epoch-epoch* akhir mengindikasikan adanya kecenderungan *overfitting* ringan, akibat peningkatan performa pada data latih tidak sepenuhnya diikuti oleh peningkatan pada data validasi. Oleh karena itu, *epoch* keempat dipilih sebagai titik pelatihan yang paling optimal karena memberikan keseimbangan terbaik antara kemampuan pembelajaran dan generalisasi model. Untuk memastikan bahwa pemilihan *epoch* tersebut juga didukung oleh pola konvergensi kesalahan prediksi selama pelatihan, analisis selanjutnya dilakukan melalui pengamatan terhadap grafik *Training vs Validation (Loss)*.

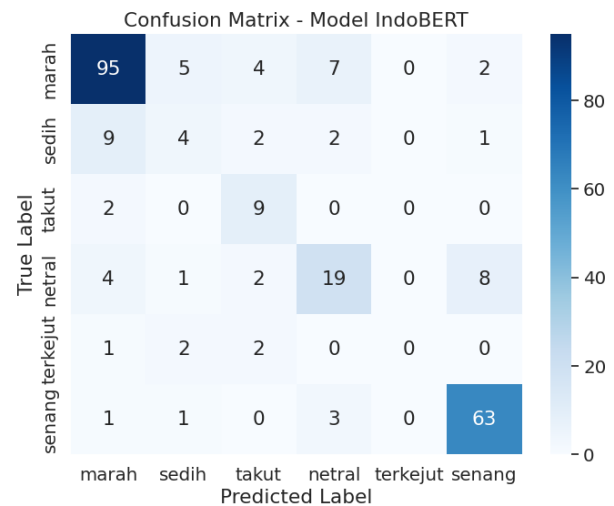


Gambar 6. Grafik *Training vs Validation (Loss)*

Berdasarkan grafik *Training vs Validation (Loss)* pada Gambar 6, terlihat bahwa nilai *training loss* menurun secara konsisten dari *epoch* pertama hingga *epoch* keenam, menunjukkan bahwa model semakin mampu meminimalkan kesalahan prediksi pada data latih. Sementara itu, *validation loss* menurun hingga mencapai titik terendah pada *epoch* ketiga, kemudian cenderung stagnan dan meningkat secara perlahan pada *epoch* berikutnya. Pola ini mengindikasikan bahwa setelah titik optimal, penambahan *epoch* tidak lagi meningkatkan kemampuan generalisasi model meskipun performa pada data latih terus membaik, sehingga muncul indikasi *overfitting* ringan. Kondisi tersebut selaras dengan aktifnya mekanisme *early stopping* pada *epoch* kelima dan keenam yang menghentikan pelatihan ketika tidak terdapat perbaikan signifikan pada data validasi. Secara keseluruhan, konsistensi antara hasil *classification report*, grafik akurasi dan grafik *loss* menunjukkan bahwa model IndoBERT mampu mempelajari representasi emosional secara efektif serta mempertahankan kemampuan generalisasi yang baik. Temuan ini memperkuat bahwa pendekatan berbasis konteks seperti IndoBERT lebih sesuai untuk klasifikasi emosi pada ulasan pengguna yang kompleks dan tidak seimbang secara distribusi kelas.

Berdasarkan *confusion matrix*, model modern (IndoBERT) menunjukkan peningkatan performa yang jelas dibandingkan model *baseline* yang ditandai dengan jumlah prediksi benar yang dominan pada diagonal utama (Gambar 7). Emosi marah dan senang memiliki jumlah prediksi benar tertinggi, masing-masing sebanyak 95 dan 63 ulasan yang menunjukkan kemampuan model dalam

mengenali emosi dengan indikator linguistik yang relatif kuat. Emosi netral dan takut juga menunjukkan performa yang lebih baik dibandingkan pendekatan klasik, dengan masing-masing 19 dan 9 prediksi benar. Sebaliknya, emosi sedih masih memiliki jumlah prediksi benar yang terbatas pada 4 ulasan, sementara emosi terkejut belum berhasil diprediksi dengan tepat.



Gambar 7. *Confusion Matrix* Klasifikasi Model Modern

Hasil ini memperkuat temuan pada *classification report* bahwa IndoBERT mampu meningkatkan ketepatan klasifikasi, terutama pada kelas mayoritas dan emosi yang memiliki konteks linguistik yang lebih jelas. Meskipun demikian, tantangan tetap terlihat pada kategori minoritas dan emosi dengan kedekatan semantik yang tinggi, sehingga distribusi data dan kompleksitas ekspresi emosional masih menjadi faktor yang memengaruhi performa model.

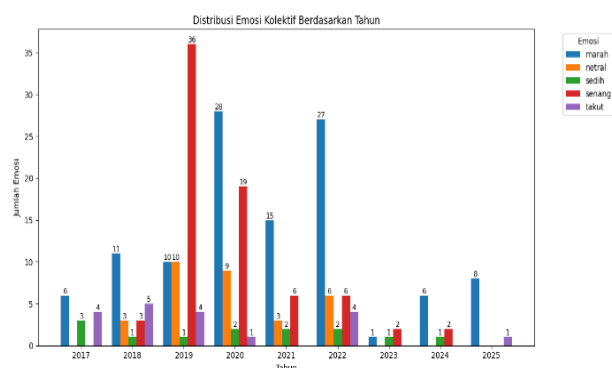
### Analisis Emosi Kolektif

Analisis distribusi emosi kolektif pengguna dilakukan untuk mengidentifikasi kecenderungan emosional terhadap aplikasi AHM Mobile selama periode pengamatan. Berdasarkan hasil visualisasi (Tabel 6), emosi marah menjadi kategori yang paling dominan dengan 112 kemunculan, diikuti oleh senang sebanyak 74 ulasan, netral 31 ulasan, takut 19 ulasan dan sedih 13 ulasan.

Tabel 6. Distribusi Emosi Kolektif

Emosi	Jumlah Emosi Kolektif
Marah	112
Senang	74
Netral	31
Takut	19
Sedih	13

Dominasi emosi marah mengindikasikan bahwa sebagian besar pengguna mengekspresikan ketidakpuasan, yang umumnya berkaitan dengan gangguan teknis, ketidakstabilan sistem, atau ketidaksesuaian kinerja aplikasi dengan ekspektasi pengguna. Meskipun demikian, kemunculan emosi senang dalam jumlah yang signifikan menunjukkan bahwa aplikasi tetap mampu memberikan pengalaman positif bagi sebagian pengguna. Emosi netral merepresentasikan ulasan yang bersifat deskriptif tanpa kecenderungan afektif yang kuat, sedangkan emosi takut dan sedih mencerminkan bentuk kekhawatiran serta kekecewaan yang lebih implisit. Secara keseluruhan, distribusi ini menunjukkan bahwa pengalaman fungsional aplikasi memiliki keterkaitan langsung dengan ekspresi emosi pengguna yang berarti pengalaman negatif cenderung memicu emosi negatif, sementara pengalaman yang berjalan baik mendorong munculnya emosi positif. Tidak ditemukannya emosi terkejut dalam hasil prediksi juga mengindikasikan keterbatasan representasi atau karakteristik data yang kurang mendukung kemunculan emosi tersebut.



Gambar 8. Distribusi Emosi Kolektif Berdasarkan Tahun

Setelah mengidentifikasi distribusi emosi secara kolektif, analisis selanjutnya difokuskan pada dinamika emosi berdasarkan tahun untuk memahami perubahan kecenderungan emosional pengguna dari waktu ke waktu.

Analisis dinamika emosi berdasarkan tahun dilakukan untuk mengamati pola perubahan distribusi emosi pengguna terhadap aplikasi AHM Mobile selama periode pengamatan (Gambar 8). Hasil visualisasi menunjukkan adanya fluktuasi kecenderungan emosional pada setiap tahun. Pada periode awal (2017-2018), emosi marah mendominasi, yang mengindikasikan fase adaptasi pengguna terhadap fungsi dasar aplikasi. Pada tahun 2019 terjadi peningkatan signifikan pada emosi senang yang menjadi emosi dominan, mencerminkan peningkatan kualitas pengalaman pengguna, namun pada periode 2020-2022, emosi marah kembali meningkat secara konsisten dan menjadi emosi yang paling banyak muncul, disertai kemunculan emosi netral, takut dan sedih dalam proporsi yang lebih kecil. Pola ini menunjukkan adanya tantangan fungsional atau peningkatan ekspektasi pengguna yang belum sepenuhnya terpenuhi. Pada tahun-tahun akhir pengamatan (2023-2025), jumlah ulasan cenderung menurun, meskipun emosi negatif tetap muncul. Secara keseluruhan, dinamika emosi ini mencerminkan hubungan yang erat antara perkembangan aplikasi dan persepsi pengguna, sehingga dapat berfungsi sebagai indikator evaluatif terhadap efektivitas pengembangan aplikasi serta dasar dalam merancang strategi perbaikan yang lebih berorientasi pada pengalaman pengguna.

Analisis selanjutnya difokuskan pada hubungan antara ekspresi emosi dalam ulasan dengan rating numerik yang diberikan pengguna terhadap aplikasi.

Tabel 7. Hubungan Emosi dan Rating Ulasan

Ulasan	Rating	Skor Emosi
joossss <ellipsis>	5	5
sebelum di upgrade bisa masuk sekarang jadi gabisa <ellipsis> tulisan nya kalo ga id salah malah nyuruh cek jaringan internet padahal udah coba pake wifi	1	1
masih banyak yang perlu ditingkatkan	1	3
tolong terus di perbaiki agar karyawan mudah aksesnya diperbarui malah gak bisa dibuka <comma> password salah terus <ellipsis> tolong gimana min <question>	5	3
	2	2

Analisis hubungan emosi dengan *rating* pengguna bertujuan untuk memahami sejauh mana ekspresi emosional dalam ulasan berkaitan dengan penilaian numerik terhadap aplikasi AHM Mobile. Secara deskriptif, pola umum menunjukkan bahwa emosi positif cenderung disertai dengan *rating* tinggi, sedangkan emosi negatif berasosiasi dengan *rating* rendah. Namun demikian, ditemukan beberapa pengecualian yang menunjukkan bahwa hubungan tersebut tidak selalu linear, di mana terdapat ulasan dengan ekspresi emosional relatif netral namun disertai *rating* rendah, maupun sebaliknya, untuk menguji hubungan ini secara kuantitatif, dilakukan analisis korelasi menggunakan *Spearman Correlation*.

Tabel 8. Hasil Nilai Koefisien *Spearman Correlation*

Hasil Nilai Koefisien <i>Spearman Correlation</i>	p-value	Interpretasi <i>Spearman Correlation</i>
0.7141	0.0000	Kuat

Hubungan antara emosi dan *rating* pengguna diuji secara kuantitatif menggunakan *Spearman Correlation*. Hasil analisis menunjukkan koefisien korelasi sebesar 0.7141 dengan *p-value* 0.0000 yang mengindikasikan adanya hubungan positif yang kuat dan signifikan secara statistik antara emosi yang diekspresikan dalam ulasan dan *rating* yang diberikan pengguna. Temuan ini menunjukkan bahwa semakin positif emosi yang disampaikan, semakin tinggi kecenderungan *rating* yang diberikan, sedangkan emosi negatif berasosiasi dengan *rating* yang lebih rendah. Meskipun demikian, hubungan yang kuat ini tidak sepenuhnya bersifat deterministik, karena penilaian pengguna juga berpotensi dipengaruhi oleh faktor kontekstual lainnya di luar ekspresi emosional yang tertuang dalam teks ulasan.

### Analisis Perspektif Ilmu Sosial Komputasional

Dalam perspektif Ilmu Sosial Komputasional, hasil klasifikasi emosi tidak hanya dipahami sebagai keluaran teknis model, tetapi sebagai representasi fenomena sosial yang mencerminkan interaksi kolektif pengguna dengan sistem digital. Ulasan pengguna dipandang sebagai ekspresi sosial yang terekam secara digital yang ketika dianalisis secara agregat membentuk pola emosi kolektif. Dominasi emosi negatif, seperti marah dan takut, dapat diinterpretasikan sebagai indikasi permasalahan sistem yang dirasakan secara luas, sementara peningkatan emosi positif mencerminkan fase pemenuhan ekspektasi pengguna. Reliabilitas anotasi manual yang ditunjukkan oleh nilai *Cohen's Kappa* sebesar 0.636 menjadi fondasi penting dalam

memastikan kualitas data, yang selanjutnya tercermin pada performa IndoBERT dengan akurasi 0.7831 dan *Macro-F1* sebesar 0.5582. Keterkaitan antara reliabilitas anotasi, performa model dan korelasi *Spearman* yang kuat ( $\rho = 0.7141$ ) antara emosi dan *rating* menunjukkan bahwa hasil klasifikasi tidak hanya konsisten secara metodologis, tetapi juga relevan secara empiris terhadap perilaku evaluatif pengguna. Dengan demikian, analisis ini menegaskan bahwa emosi pengguna berfungsi sebagai indikator sosial terhadap kualitas interaksi antara pengguna dan aplikasi, serta mencerminkan dinamika persepsi dan kepercayaan kolektif dalam ekosistem digital.

### Implikasi Penelitian

Temuan mengenai distribusi emosi, dinamika emosi tahunan, serta hubungan antara emosi dan *rating* pengguna memberikan dasar empiris yang relevan bagi evaluasi dan pengambilan keputusan strategis dalam pengembangan aplikasi AHM Mobile. Dominasi emosi negatif pada periode tertentu mengindikasikan adanya aspek fungsional yang perlu menjadi prioritas perbaikan, seperti stabilitas sistem dan keandalan fitur, sementara kemunculan emosi positif dapat menjadi acuan dalam mempertahankan elemen yang telah memberikan pengalaman memuaskan. Keterkaitan yang kuat antara emosi dan *rating* menunjukkan bahwa analisis emosi dapat berfungsi sebagai indikator komplementer terhadap matriks konvensional, sekaligus sebagai sinyal dini terhadap potensi penurunan kualitas penggunaan aplikasi. Integrasi analisis emosi ke dalam proses evaluasi internal memungkinkan pengelola aplikasi untuk memahami kebutuhan dan persepsi pengguna secara lebih komprehensif serta merancang strategi peningkatan penggunaan yang berorientasi pada pengalaman pengguna. Secara lebih luas, pendekatan analisis emosi berbasis komputasional dapat juga diterapkan pada aplikasi digital lainnya sebagai kerangka evaluasi untuk meningkatkan kualitas penggunaan dan kepuasan pengguna secara berkelanjutan.

## KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan emosi yang terkandung dalam ulasan pengguna aplikasi internal perusahaan melalui *Natural Language Processing* (NLP). Berdasarkan tujuan tersebut, penelitian ini menunjukkan bahwa ekspresi emosional dalam ulasan berbasis teks dapat direpresentasikan secara sistematis dan terukur ke dalam kategori emosi yang

terdefinisi secara konseptual. Metode *Natural Language Processing* (NLP) berperan sebagai kerangka analitis untuk mengubah ekspresi emosional yang bersifat subjektif dan tidak terstruktur menjadi informasi terstruktur yang dapat dianalisis secara komputasional dalam konteks evaluasi pengalaman pengguna aplikasi digital.

Tujuan kedua penelitian yaitu menganalisis hubungan antara emosi kolektif pengguna dan *rating* yang diberikan untuk menegaskan bahwa emosi tersebut memiliki keterkaitan dengan penilaian pengguna terhadap aplikasi. Emosi yang muncul secara agregat membentuk pola persepsi bersama yang merefleksikan pengalaman sosial pengguna dalam berinteraksi dengan sistem digital. Dalam konteks ini, emosi kolektif tidak hanya merepresentasikan respons individual, tetapi juga menggambarkan dinamika interaksi dan evaluasi bersama terhadap kualitas penggunaan aplikasi internal perusahaan.

Selanjutnya, penelitian ini memberikan kontribusi teoretis melalui integrasi pendekatan *Natural Language Processing* (NLP) dengan kerangka Ilmu Sosial Komputasional. Ulasan pengguna diposisikan sebagai *digital traces* yang merekam pengalaman sosial dan emosional pengguna dalam ekosistem aplikasi digital. Integrasi kedua pendekatan tersebut menempatkan analisis emosi sebagai bagian dari kajian sosial berbasis data, yang berarti ekspresi emosional pengguna dipahami sebagai indikator persepsi kolektif terhadap sistem digital. Dengan demikian, penelitian ini memperkaya kajian integrasi *Natural Language Processing* (NLP) dengan Ilmu Sosial Komputasional, khususnya dalam konteks aplikasi internal perusahaan atau layanan digital serupa di Indonesia dengan menunjukkan bahwa data ulasan berbasis teks dapat digunakan untuk memahami dinamika sosial dan emosional pengguna secara agregat.

#### UCAPAN TERIMA KASIH

Para penulis ingin mengucapkan terima kasih kepada Universitas Al-Azhar Indonesia atas dukungan dan fasilitas institusional yang membuat penelitian ini menjadi mungkin.

#### REFERENSI

[1] A. A. Vărzaru and C. G. Bocean, "Digital Transformation and Innovation: The Influence of Digital Technologies on Turnover from

Innovation Activities and Types of Innovation," *Systems*, vol. 12, no. 9, 2024, doi: 10.3390/systems12090359.

- [2] Welovehonda, "AHM Mobile," Google Play Store. [Online]. Available: <https://play.google.com/store/apps/details?id=com.ahm.hr>
- [3] S. K. Lubis, M. H. Dar, and F. A. Nasution, "Analisis Sentimen Ulasan Pengguna Aplikasi pada Google Play Store Menggunakan Algoritma Support Vector Machine," *J. Surya Inform.*, vol. 11, no. 2, pp. 120–128, 2024, doi: 10.36987/informatika.v11i2.5860.
- [4] N. P. Husain, S. Sukirman, and S. SAJIAH, "Analisis Sentimen Ulasan Pengguna Tiktok pada Google Play Store Berbasis TF-IDF dan Support Vector Machine," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 91–102, 2024, doi: 10.61628/jsce.v5i1.1105.
- [5] N. F. Hilmi and F. Irwiensyah, "Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naïve Bayes," *SMATIKA STIKI Inform. J.*, vol. 14, no. 01, pp. 146–156, 2024, doi: 10.32664/smatika.v14i01.1210.
- [6] D. Lazer *et al.*, "Computational Social Science," *Science (80-. )*, vol. 323, pp. 721–723, 2009, [Online]. Available: <https://www.science.org/doi/10.1126/science.1167742>
- [7] P. C. Ncr *et al.*, "Step-by-step data mining guide," 2000. [Online]. Available: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>
- [8] M. Z. E. Putra, "Representasi Emosi Dalam Cuitan Twitter Komunitas Marah-Marah: Kajian Psikolinguistik," *J. Pendidik. Bhs. dan Sastra Indones.*, vol. 14, pp. 349–358, 2024, doi: <https://doi.org/10.23887/jpbsi.v14i3.82042>.
- [9] N. Zahra, Y. Sonia, S. Adilla, R. A. Mardiyah, and D. Amelia, "Semantik Dalam Bahasa Indonesia," *Morfol. J. Ilmu Pendidikan, Bahasa, Sastra dan Budaya*, vol. 2, no. 6, pp. 156–164, 2024, doi: 10.61132/morfologi.v2i6.1163.
- [10] N. F. Ramadhani, R. A. Hamzah, and R. Arief, "Struktur Kebahasaan Bahasa Indonesia Sebagai Rujukan Penggunaan Bahasa (Sitaksis)," *Kande J. Ilm. Pendidik. Bhs. dan Sastra Indones.*, vol. 06, no. 01, pp. 1–14, 2025, doi: <https://doi.org/10.29103/jk.v6i1.21412>.
- [11] R. Parlita, R. F. Taufiqurrahman, H. I. Farhana, and R. Dimas, "Pengukuran Validitas Aplikasi

- Akademik Siswa Berbasis Website Menggunakan Metode R-Tabel Dan Cohen's Kappa," *J. Teknol. Inf.*, vol. 6, no. 1, 2022, [Online]. Available: <https://jurnal.una.ac.id/index.php/jurti/article/view/2617>
- [12] J. Klie and R. E. De Castilho, "Analyzing Dataset Annotation Quality Management in the Wild," *Comput. Linguist.*, 2024, doi: [https://doi.org/10.1162/coli\\_a\\_00516](https://doi.org/10.1162/coli_a_00516).
- [13] M. V. S. Reddy, B. C. Samantha, B. Madhavi, S. Likhita, and K. Sumathi, "Convert Emoji Into Text Using Python," *Int. J. Progress. Res. Eng. Manag. Sci.*, pp. 1711–1715, 2025, [Online]. Available: <https://www.ijprems.com/research-paper/convert-emoji-into-text-using-python>
- [14] F. Stahlberg and S. Kumar, "Seq2Edits: Sequence transduction using span-level edit operations," in *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2020, pp. 5147–5159. doi: 10.18653/v1/2020.emnlp-main.418.
- [15] P. M. S. Ardinata, A. A. J. Permana, and I. N. S. W. Wijaya, "Identifikasi Dan Normalisasi Teks Slang Dengan Fasttext Pada Twitter Dalam Bahasa Indonesia," *J. Pendidik. Teknol. dan Kejuru.*, vol. 21, no. 1, pp. 34–44, 2024, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JPTK/article/view/66381>
- [16] K. E. Saputra, "Multilabel multiclass sentiment and emotion dataset from indonesian mobile application review," *Data Br.*, vol. 50, 2023, doi: 10.1016/j.dib.2023.109576.
- [17] I. B. Jatiarso and Y. Azhar, "Klasifikasi Emosi Pada Tweet Pengguna Platform X Menggunakan Metode LSTM-GloVe Berbasis SMOTE," *J. Repos.*, vol. 7, no. 3, pp. 401–406, 2025, [Online]. Available: <https://ejournal.umm.ac.id/index.php/repositor/article/view/36608>
- [18] I. I. A. Nadhiroh, M. Z. Sarwani, and M. Udin, "Mendeteksi Emosi Berdasarkan Postingan Sosial Media X Menggunakan Algoritma Long Short-Term Memory," *Techno.Com*, vol. 24, no. 3, pp. 745–755, 2025, doi: 10.62411/tc.v24i3.13509.
- [19] J. Wu, M. Barthet, D. Melhart, and G. N. Yannakakis, "Emotions as Ambiguity-aware Ordinal Representations," in *International Conference on Affective Computing and Intelligent Interaction (ACII) Emotions*, 2025.
- [20] A. F. Suherman, P. P. Lisnaeni, S. A. Izqiatullailiyah, T. Herlinawati, and Ahman, "A Comparative Analysis of Spearman and Pearson Correlation Using SPSS," *Guid. J. Guid. Couns.*, vol. 6, no. 1, pp. 1–10, 2016, [Online]. Available: <https://ejournal.upi.edu/index.php/optima/article/view/79415>
- [21] C. Lotfi, S. Srinivasan, M. Ertz, and I. Latrous, "Web Scraping Techniques and Applications: A Literature Review," in *Scrs Conference Proceedings on Intelligent Systems*, 2021, pp. 381–394. doi: 10.52458/978-93-91842-08-6-38.