

User Review Analysis: Psychological Impact of Duolingo Gamification Using Machine Learning

Septi Mulia Putri¹, Ken Ditha Tania^{2*}

Sistem Informasi, Universitas Sriwijaya
septimulia743@gmail.com¹, kenya.tania@gmail.com²

Article Info

Article history:

Received 2026-02-11

Revised 2026-02-26

Accepted 2026-04-08

Keyword:

*Duolingo,
Gamification,
Machine Learning,
Psychological Analysis.*

ABSTRACT

Gamification in Duolingo aims to enhance user motivation, yet its internal elements trigger diverse psychological responses. This study aims to analyze these responses through app reviews by comparing the effectiveness of four machine learning models, namely Naïve Bayes, Random Forest, Support Vector Machine (SVM), and Logistic Regression. The research design employs a quantitative comparative approach with review data classified into three labels, including Motivated, Frustrated, and Neutral. The methodology involves text preprocessing, label encoding, and addressing data imbalance using SMOTE and Class Weight techniques. The results indicate that all models achieved high accuracy, ranging from 92% to 93%. However, based on the Macro Average F1-Score metric, which measures the model's performance balance across all classes including minority ones, the SVM and Logistic Regression models delivered the most optimal performance with a score of 0.44, outperforming Naïve Bayes and Random Forest, which scored 0.40. Feature analysis through Feature Importance reveals that the Penalty feature is the dominant factor triggering frustrated conditions, whereas the Daily Challenge feature serves as a strong predictor despite its low frequency. In conclusion, SVM and Logistic Regression are the most reliable models for mapping user emotional responses, where the findings regarding the Penalty feature provide a critical evaluation for the development of gamification mechanics.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Gamifikasi, yaitu penambahan elemen permainan dalam konteks yang bukan permainan, muncul sebagai salah satu strategi ampuh dalam mendesain aplikasi Pembelajaran Bahasa Berbantuan Seluler (*Mobile-Assisted Language Learning/MALL*) untuk meningkatkan keterlibatan dan retensi pengguna. Pada kenyataannya, pembelajaran berbasis permainan interaktif dengan media digital dapat menghasilkan tingkat partisipasi, keterlibatan, dan motivasi belajar yang jauh lebih tinggi [1]. Secara teoretis, efektivitas gamifikasi berakar pada *Self-Determination Theory* yang menjelaskan bahwa motivasi manusia didorong oleh tiga kebutuhan psikologis dasar yaitu otonomi, kompetensi dan keterkaitan [2]. Duolingo, sebagai salah satu platform pembelajaran bahasa online terbesar di dunia, menggunakan pendekatan ini untuk memberikan fitur peningkatan level

seperti poin pengalaman (XP), papan peringkat, dan sistem beruntun. Penerapan gamifikasi dirancang untuk merangsang motivasi intrinsik dan ekstrinsik sehingga pengguna dapat mempertahankan konsistensi belajar. Namun, keberhasilan teknologi MALL ini bergantung pada persepsi pengguna tentang kegunaan dan manfaatnya untuk belajar [3]. Selain itu, penggabungan MALL yang efektif terbukti memiliki efek positif dalam meningkatkan keterampilan pemahaman bacaan (reading skills) bagi pembelajar bahasa [4].

Namun, tidak semua elemen gamifikasi memiliki dampak psikologis yang positif. Mekanisme permainan juga dapat menimbulkan tekanan psikologis, seperti kecemasan dan frustrasi di antara pengguna dalam beberapa kasus, terutama jika mekanisme kompetitif yang berlebihan lebih diutamakan daripada tujuan pembelajaran itu sendiri [5]. Dalam perspektif psikologi motivasi, fitur yang terlalu mengontrol dan bersifat menghukum dapat menyebabkan *need frustration*

yang menghambat pemenuhan kebutuhan dasar pengguna dan memicu respons emosional negatif [6]. Fenomena ini menunjukkan bahwa gamifikasi tidak hanya berfungsi sebagai alat motivasi, tetapi juga dapat menimbulkan respons emosional yang beragam tergantung pada persepsi pengguna terhadap instruksi dan pengalaman belajar yang mereka terima [7].

Meskipun dirancang dengan tujuan meningkatkan motivasi belajar, berbagai ulasan pengguna di platform Google Play Store mengindikasikan bahwa implementasi gamifikasi dalam aplikasi Duolingo berpotensi memunculkan “sisi gelap” tersendiri. Hasil analisis sentimen menggunakan algoritma Naïve Bayes memperlihatkan bahwa pengalaman pengguna tidak selamanya bersifat positif, mengingat adanya fluktuasi kepuasan serta munculnya sentimen negatif yang signifikan terkait fitur-fitur tertentu dalam aplikasi [8]. Beberapa fitur, seperti sistem penalti berbentuk “hati” atau “nyawa” yang membatasi aktivitas pembelajaran ketika pengguna melakukan kesalahan, serta kewajiban mempertahankan streak, kerap dipersepsikan sebagai masalah dalam kinerja dan instruksi pembelajaran, sehingga memicu ketidakpuasan pengguna [7]. Hal ini sejalan dengan kritik bahwa elemen gamifikasi dalam Duolingo terkadang lebih mengutamakan aspek permainan dibandingkan prinsip-prinsip pedagogis yang mendalam [5]. Akibatnya, pengguna yang pada awalnya merasa termotivasi dapat mengalami frustrasi apabila mereka mempersepsikan mekanika permainan tersebut sebagai sesuatu yang bersifat menghukum alih-alih mendukung proses pembelajaran. Situasi ini menggarisbawahi bahwa respons psikologis pengguna terhadap gamifikasi bersifat dinamis dan kompleks, serta sangat bergantung pada bagaimana elemen-elemen kompetitif dirancang agar tidak merepresentasikan beban mental yang kontraproduktif bagi pembelajar [5], [7].

Oleh karena itu, analisis sentimen pengguna perlu dilakukan secara lebih mendalam, tidak hanya terbatas pada klasifikasi polaritas sentimen, melainkan juga mengidentifikasi fitur-fitur spesifik yang berperan sebagai pemicu utama perubahan kondisi psikologis pengguna. Hal ini menjadi krusial mengingat kondisi afektif pengguna dalam sumber daya pembelajaran memberikan dampak signifikan terhadap proses kognitif, sehingga kegagalan dalam mengidentifikasi emosi secara tepat dapat menghambat efektivitas pendidikan itu sendiri [9]. Studi-studi terdahulu mengenai analisis sentimen aplikasi edukatif umumnya menggunakan pendekatan *Bag-of-Words* atau pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengklasifikasikan ulasan ke dalam kategori sentimen positif dan negatif. Namun demikian, pendekatan konvensional ini memiliki keterbatasan dalam menangkap konteks dan ambiguitas opini publik yang dinamis, karena model tersebut hanya merefleksikan sentimen umum tanpa mampu menjelaskan penyebab spesifiknya [10]. Untuk mengatasi keterbatasan tersebut, diperlukan pergeseran menuju analisis yang lebih granular. Penerapan *Aspect-Based Sentiment Analysis* (ABSA) merupakan solusi fundamental

dalam ekosistem pendidikan digital untuk mengevaluasi umpan balik pada tingkat aspek fitur, yang memungkinkan pengembang memahami fitur gamifikasi mana yang mendorong motivasi dan mana yang sebaliknya merepresentasikan beban psikologis bagi pengguna [10], [11]. Hal ini didukung oleh analisis sentimen ulasan Play Store, yang memperlihatkan bahwa ulasan pengguna kerap mengandung emosi kompleks (*mixed feelings*) yang membutuhkan akurasi lebih tinggi dalam algoritma klasifikasi [8].

Mengacu pada keterbatasan yang telah diuraikan, penelitian ini mengadopsi pendekatan *Feature-Based Opinion Mining* melalui ekstraksi delapan fitur gamifikasi, yakni *XP/Levelling*, *Streak*, *Badge/Achievement*, *Leaderboard*, *Reward/Gem*, *Daily Challenge*, *Reminder Notification*, dan *Penalti*. Proses ekstraksi diimplementasikan menggunakan metode *Rule-Based Keyword Extraction* yang disusun berlandaskan taksonomi gamifikasi serta terminologi spesifik dalam aplikasi Duolingo. Pendekatan ini memfasilitasi analisis sentimen yang lebih kontekstual, sehingga kondisi psikologis pengguna dapat dihubungkan secara langsung dengan fitur gamifikasi tertentu. Sebagaimana dipaparkan dalam [12], analisis sentimen pada tingkat fitur atau aspek terbukti sangat efektif untuk mengidentifikasi komponen spesifik yang memengaruhi kepuasan pengguna secara mendalam. Di samping itu, tren riset terbaru mengindikasikan bahwa pemetaan relasi antara opini dan entitas fitur tertentu merupakan kunci untuk menghasilkan wawasan yang dapat ditindaklanjuti (*actionable insights*) [13]. Dalam konteks data pendidikan, penerapan teknik ekstraksi fitur ini menjadi sangat krusial untuk memahami pengalaman belajar secara personal dan dinamis, yang tidak dapat ditangkap hanya melalui klasifikasi sentimen global [14].

Selanjutnya, ulasan pengguna diklasifikasikan ke dalam tiga kondisi psikologis, yaitu *Motivated*, *Frustrated*, dan *Neutral*, dengan membandingkan kinerja empat algoritma *machine learning*. Naïve Bayes dipilih karena efektivitasnya dalam menghasilkan klasifikasi yang cepat dengan beban komputasi rendah pada data teks ulasan aplikasi [15], [16]. *Random Forest* digunakan karena keunggulannya dalam memberikan stabilitas prediksi yang tinggi melalui arsitektur paralel dan optimasi pohon keputusan [10], [17], [18]. Selain itu, penelitian ini juga menerapkan *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Pemilihan kedua algoritma ini didasarkan pada efisiensi SVM dalam menangani dimensi fitur teks yang kompleks serta performa *Logistic Regression* yang stabil dalam tugas klasifikasi teks [19], [20], [21]. Perbandingan keempat algoritma ini bertujuan untuk menentukan model yang paling optimal dalam memetakan hubungan antara fitur gamifikasi dan respon emosional pengguna secara objektif.

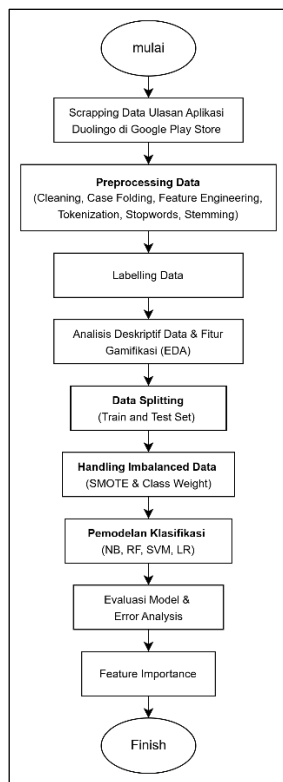
Lebih lanjut, penelitian ini juga menganalisis tingkat pengaruh masing-masing fitur gamifikasi terhadap kondisi psikologis pengguna melalui analisis *Feature Importance*. Analisis ini bertujuan untuk mengidentifikasi elemen permainan yang paling signifikan dalam membentuk motivasi

maupun frustrasi pengguna secara statistik. Hal ini selaras dengan urgensi untuk memahami bagaimana elemen gamifikasi secara spesifik memengaruhi pengalaman dan emosi pengguna dalam ekosistem digital yang semakin kompleks [22]. Dengan memanfaatkan kemampuan algoritma Random Forest dalam mengevaluasi kontribusi setiap variabel, penelitian ini dapat menentukan fitur mana yang paling dominan dalam memengaruhi akurasi klasifikasi sentimen [23]. Dengan demikian, penelitian ini tidak hanya berfokus pada akurasi klasifikasi, tetapi juga memberikan pemahaman yang lebih interpretatif mengenai hubungan timbal balik antara fitur gamifikasi tertentu dan kondisi psikologis pengguna [17].

Secara keseluruhan, temuan penelitian ini diharapkan dapat memberikan wawasan praktis bagi pengembang aplikasi pembelajaran bahasa dalam merancang sistem gamifikasi yang lebih adaptif, seimbang, dan memperhatikan aspek psikologis pengguna sehingga tujuan peningkatan motivasi belajar dapat tercapai tanpa menimbulkan tekanan emosional yang berlebihan.

II. METODE

Metodologi penelitian ini dirancang untuk mengekstraksi wawasan psikologis dari data teks ulasan pengguna dengan fokus pada fitur-fitur gamifikasi.



Gambar 1. Flowchart Alur Penelitian

A. Scraping Data

Objek penelitian dalam studi ini adalah ulasan pengguna aplikasi Duolingo pada platform Google Play Store. Data yang dimanfaatkan berupa ulasan terkini yang dikumpulkan secara otomatis dengan mengimplementasikan teknik *web scraping*, yakni metode pengambilan data dari laman web secara terstruktur menggunakan skrip pemrograman untuk mengekstraksi informasi tertentu. Teknik ini dipilih karena kapabilitasnya dalam mengekstraksi data teks secara tidak terstruktur dari berbagai sumber web secara efisien untuk kemudian ditransformasikan menjadi informasi yang terorganisir dan dapat dianalisis [24]. Proses pengumpulan data diimplementasikan dengan memanfaatkan bahasa pemrograman Python pada lingkungan Google Colab dengan total dataset terkumpul sebanyak 5.000 ulasan berbahasa Indonesia.

TABEL 1. SAMPEL DATA

username	score	content
Pengguna Google	5	terbaik
Pengguna Google	5	ya bagus
Pengguna Google	4	lebih enak pake hati sih daripada energi, soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah
Pengguna Google	5	bagus banget
Pengguna Google	1	Caturnya dipermudah donk gimana sih, aplikasi ga berguna

B. Preprocessing Data

Tahap *preprocessing data* bertujuan untuk mempersiapkan ulasan pengguna sebelum memasuki proses analisis. Proses ini krusial karena data mentah kerap rentan terhadap *noise*, korupsi, dan inkonsistensi yang dapat menurunkan akurasi model secara signifikan [25]. Di samping itu, implementasi strategi *preprocessing* yang tepat terbukti mampu meningkatkan kinerja model klasifikasi, khususnya pada algoritma seperti Random Forest, dengan meminimalkan bias dan meningkatkan kualitas data latih [26]. Rangkaian proses ini mencakup *Cleaning*, *Case Folding*, *Feature Engineering*, *Tokenization*, *Stopwords*, dan *Stemming*.

1. Cleaning dan Case Folding

Cleaning dilakukan untuk menghasilkan elemen-elemen yang tidak relevan dalam teks ulasan, seperti karakter non-alfabet, angka, tanda baca, serta simbol khusus yang tidak memiliki kontribusi terhadap proses analisis. Pembersihan data adalah langkah mendasar untuk mendeteksi dan mengoreksi data yang tidak akurat guna mencegah prediksi yang salah [25]. Selanjutnya, dilakukan *Case Folding* dengan mengonversi seluruh teks menjadi huruf kecil (*lowercase*) guna menyamakan bentuk data dan menjaga konsistensi.

TABEL 2. DATA SETELAH NORMALISASI TEKS

No	content	content_clean
1	terbaik	terbaik

2	ya bagus	ya bagus
3	lebih enak pake hati sih daripada energi, soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah	lebih enak pake hati sih daripada energi soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah
4	bagus banget	bagus banget
5	Caturnya dipermudah donk gimana sih, aplikasi ga berguna	caturnya dipermudah donk gimana sih aplikasi ga berguna

2. Feature Engineering (Ekstraksi Fitur Gamifikasi)

Tahap *Feature Engineering* dilakukan setelah proses normalisasi teks dan sebelum tahap *Tokenization*. Tujuan dari tahap ini adalah untuk mendeteksi kemunculan elemen-elemen gamifikasi dalam struktur kalimat ulasan yang masih utuh. Metode yang digunakan dalam penelitian ini adalah *Rule-based Keyword Extraction*. Secara fundamental, ekstraksi kata kunci bertujuan untuk mengidentifikasi unit kata atau frasa yang merepresentasikan topik utama atau informasi penting dalam sebuah teks [27]. Pada tahap ini, penelitian membangun delapan fitur gamifikasi yang diekstraksi berdasarkan kemunculan kata kunci tertentu dalam setiap ulasan pengguna, di mana setiap fitur memiliki daftar *keywords* yang merepresentasikan mekanisme gamifikasi yang dianalisis. Pemetaan fitur dilakukan secara biner, yaitu fitur akan bernilai 1 apabila salah satu kunci terdeteksi dalam ulasan, dan bernilai 0 apabila tidak ditemukan. Pendekatan berbasis aturan (*Rule Based*) ini dipilih karena kemampuannya dalam mengekstraksi informasi spesifik secara konsisten sesuai dengan taksonomi yang telah ditentukan [27].

TABEL 3.
KATA KUNCI PADA SETIAP FITUR

Kode Fitur	Nama Fitur	Kata Kunci (Keywords)
X1	F_xp_level	xp, level, poin, pengalaman
X2	F_streak	streak, beruntun, api
X3	F_badge_achievement	badge, pencapaian, medali, achievement, lencana, tropi
X4	F_leaderboard	leaderboard, liga, peringkat, rangking, papan skor
X5	F_reward_gem	reward, gem, permata, lingot, koin
X6	F_daily_challenge	challenge, tantangan, misi, quest, harian
X7	F_reminder_notification	notif, pengingat, reminder, notifikasi, pemberitahuan, burung
X8	F_penalti	penalti, nyawa, hati, heart, energi, energy

TABEL 4.
PEMETAAN FITUR GAMIFIKASI

<i>Content_clean</i>	X1	X2	X3	X4	X5	X6	X7	X8
terbaik	0	0	0	0	0	0	0	0
ya bagus	0	0	0	0	0	0	0	0
lebih enak pake hati sih daripada energi soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah	0	0	0	1	0	0	0	1
bagus banget	0	0	0	0	0	0	0	0
caturnya dipermudah donk gimana sih aplikasi ga berguna	0	0	0	0	0	0	0	0

3. Tokenization dan Stopwords Removal

Tokenization dilakukan untuk memecah teks ulasan menjadi unit-unit kata (token) setelah proses *Feature Engineering* selesai. Proses ini bertujuan untuk mengubah teks yang semula berbentuk kalimat utuh menjadi daftar token kata agar dapat diproses secara matematis oleh mesin [28]. Selanjutnya, penghapusan *stopwords* dilakukan setelah proses *tokenization* untuk menghilangkan kata-kata umum yang sering muncul namun tidak memiliki kontribusi signifikan terhadap proses analisis, seperti kata hubung dan kata depan.

TABEL 5.
DATA SETELAH PROSES *TOKENIZATION* DAN *STOPWORDS*

No	<i>content</i>	<i>content_stopwords</i>
1	terbaik	terbaik
2	ya bagus	ya bagus
3	lebih enak pake hati sih daripada energi, soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah	enak pake hati sih energi kalo energi pake ngerjain kalo salah turun kalo liga susah
4	bagus banget	bagus banget
5	Caturnya dipermudah donk gimana sih, aplikasi ga berguna	caturnya dipermudah donk gimana sih aplikasi ga berguna

4. Stemming

Tahap *stemming* dilakukan untuk mengembalikan kata-kata hasil *tokenization* ke bentuk dasarnya dengan menghilangkan imbuhan yang melekat. Sebagaimana dijelaskan dalam [28], proses ini bertujuan untuk menyatukan berbagai variasi kata yang memiliki makna dasar yang sama. Hal ini memungkinkan representasi fitur menjadi lebih sederhana dan konsisten, sehingga mengurangi dimensi data yang harus diolah dalam proses analisis dan klasifikasi sentimen [25].

TABEL 6.
DATA SETELAH PROSES *STEMMING*

No	content	content_final
1	terbaik	baik
2	ya bagus	ya bagus
3	lebih enak pake hati sih daripada energi, soalnya kalo energi kita pake buat ngerjain kalo salah kan ikut turun jadinya kalo mau naik liga susah	enak pake hati sih energi kalo energi pake ngerjain kalo salah turun kalo liga susah
4	bagus banget	bagus banget
5	Caturnya dipermudah donk gimana sih, aplikasi ga berguna	catur mudah donk gimana sih aplikasi ga guna

C. Labeling Data (Ground Truth)

Tahap pelabelan data dilakukan untuk menetapkan kategori kondisi psikologis pengguna. Dalam penelitian ini, pelabelan menggunakan skor rating Google Play Store sebagai label otomatis atau *ground truth*. Ulasan dengan rating 4 dan 5 dikategorikan sebagai *Motivated*, rating 1 dan 2 sebagai *Frustrated*, serta rating 3 sebagai *Neutral*. Secara operasional, label *Motivated* merepresentasikan kondisi psikologis positif yang ditandai dengan ekspresi kepuasan atau apresiasi terhadap fitur. Label *Frustrated* mencerminkan adanya hambatan psikologis atau *need frustration* yang ditandai dengan keluhan terhadap mekanisme aplikasi yang dianggap menghambat kemajuan pengguna. Sementara itu, label *Neutral* mengacu pada ulasan deskriptif tanpa muatan emosional yang kuat atau memiliki keseimbangan antara umpan balik positif dan negatif.

Pendekatan konversi rating menjadi tiga kelas ini merupakan strategi *Weak Supervision* untuk mendapatkan label objektif secara efisien tanpa melibatkan anotator ahli manusia guna menghindari subjektivitas antar penilai [29]. Untuk memastikan validitas label otomatis tersebut, dilakukan proses verifikasi manual melalui *random sampling* pada sebagian kecil data guna memastikan kesesuaian antara sentimen dalam teks ulasan dengan label rating yang diberikan. Selanjutnya, label kategori tersebut dikonversi menjadi nilai numerik melalui proses *label encoding* agar dapat diproses oleh algoritma pembelajaran mesin, khususnya model linear seperti SVM dan *Logistic Regression*.

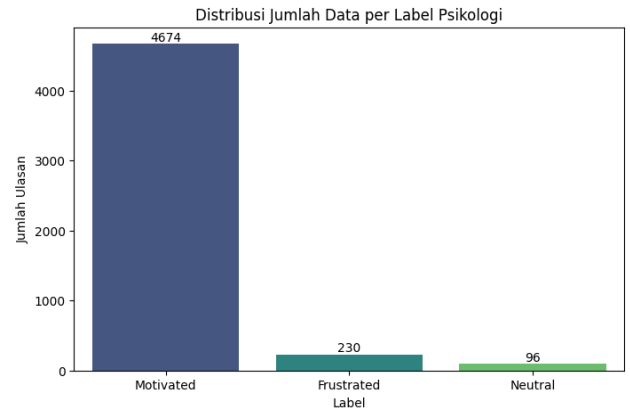
TABEL 7.
DATA SETELAH PELABELAN

Score	Label Psikologi	Label Psikologi Encoded
5	Motivated	1
5	Motivated	1
4	Motivated	1
5	Motivated	1
1	Frustrated	0

III. HASIL DAN PEMBAHASAN

A. Exploratory Data Analysis (EDA)

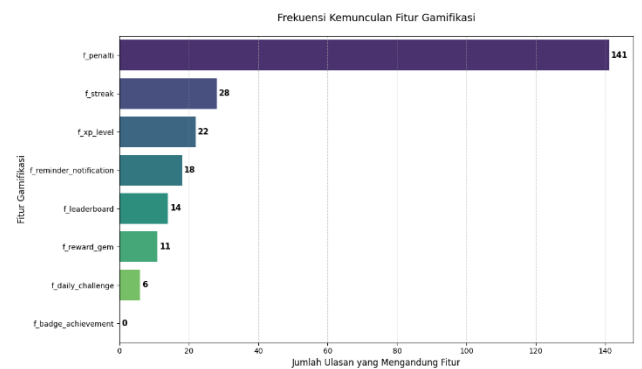
Exploratory Data Analysis (EDA) merupakan tahap krusial untuk mengeksplorasi dan mentransformasi data agar diperoleh informasi relevan yang mendukung akurasi prediksi model [30]. Berdasarkan distribusi ulasan pada Gambar 2, terlihat dominasi label *Motivated* sebanyak 4.674 data, yang menunjukkan kepuasan tinggi terhadap aplikasi Duolingo. Namun, adanya ketimpangan jumlah data (*Imbalanced Data*) pada kelas *Frustrated* dan *Neutral* menjadi landasan peneliti untuk menerapkan teknik *SMOTE* dan *Class Weight* guna memastikan model tetap objektif dalam mengenali pola ulasan negatif.



Gambar 2. Distribusi Jumlah Data per Label Psikologi

1. Distribusi Frekuensi Kemunculan Fitur Gamifikasi

Analisis dilakukan untuk mengidentifikasi fitur gamifikasi mana yang paling banyak dibahas oleh pengguna. Hal ini penting untuk memetakan elemen aplikasi yang paling memberikan kesan atau pengalaman bagi pengguna.

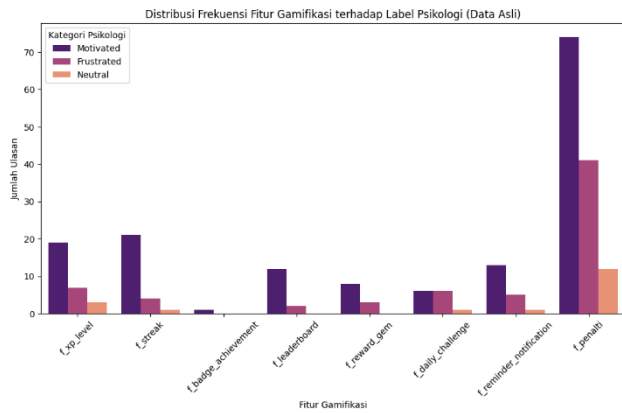


Gambar 3. Frekuensi Kemunculan Fitur

Berdasarkan analisis terhadap 5.000 ulasan, ditemukan bahwa fitur Penalti mendominasi pembicaraan pengguna dengan 141 ulasan. Angka ini jauh melampaui fitur *Streak* (28) dan *XP/Level* (22). Dominasi fitur Penalti mengindikasikan bahwa mekanisme “hukuman” atau pembatasan belajar memiliki dampak emosional yang lebih kuat dibandingkan fitur pencapaian. Sebaliknya, nihilnya pembahasan pada *Badge/Achievement* (0) menunjukkan fitur tersebut kurang memberikan kesan bagi pengguna dalam dataset ini. Hal ini membuktikan bahwa aspek

keberlangsungan belajar lebih krusial bagi pengguna daripada pengumpulan trofi digital.

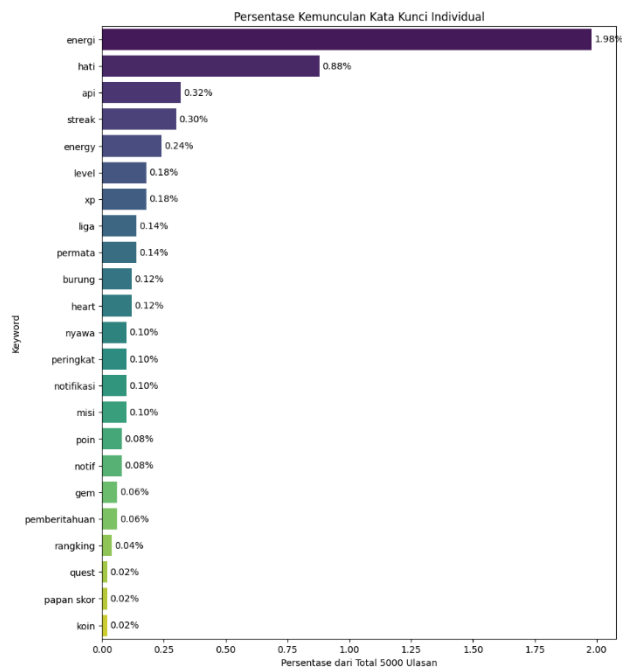
2. Distribusi Frekuensi Fitur Gamifikasi terhadap Respon Psikologi



Gambar 4. Frekuensi fitur gamifikasi terhadap label psikologi

Analisis distribusi fitur terhadap respon psikologi pada Gambar 4 menunjukkan bahwa fitur Penalti merupakan pemicu respon emosional yang paling kontras. Meskipun mendominasi pada label *Motivated*, fitur ini menyumbang jumlah ulasan *Frustrated* terbesar dibandingkan fitur lainnya. Hal ini mengonfirmasi bahwa mekanisme penalti adalah elemen paling sensitif yang mampu memicu motivasi sekaligus kekecewaan pengguna secara signifikan. Sebaliknya, fitur progresivitas seperti *Streak* dan *XP/Level* menunjukkan pola yang lebih stabil dengan dominasi pada label *Motivated* dan tingkat frustrasi yang rendah.

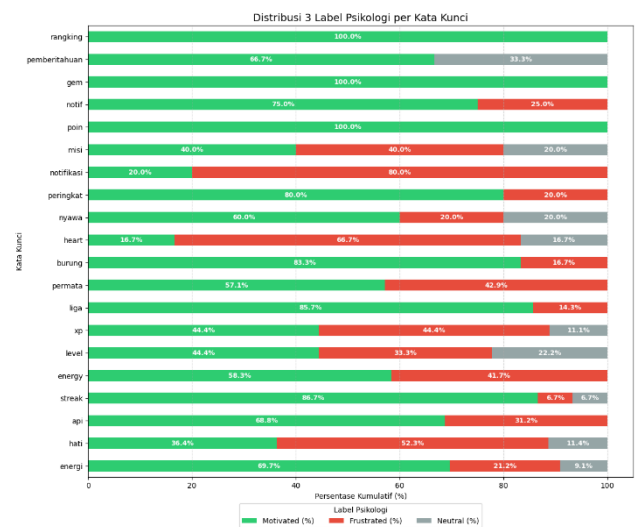
3. Analisis Kata Kunci Spesifik dan Persentase Kemunculan



Gambar 5. Persentase Kemunculan Kata Kunci Individual

Pada Gambar 5, kata kunci yang berkaitan dengan fitur Penalti mendominasi narasi pengguna secara signifikan. Kata kunci “energi” menempati posisi puncak dengan persentase kemunculan sebesar 1.98% diikuti oleh kata “hati” sebesar 0.88%. Sebaliknya, kata kunci yang merepresentasikan fitur persistensi seperti “api” (0.32%) dan “streak” (0.30%) muncul dengan frekuensi yang jauh lebih rendah. Selain itu, kata kunci seperti “poin” (0.08%) dan “permata” (0.14%) memiliki persentase yang sangat kecil, menandakan bahwa aspek ekonomi digital dalam aplikasi kurang menjadi perhatian utama pengguna dalam dataset ini.

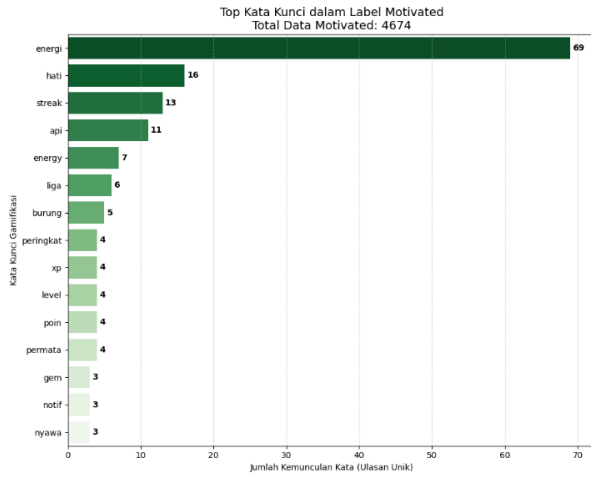
4. Distribusi 3 Label Psikologi per Kata Kunci



Gambar 6. Distribusi 3 Label Psikologi per Kata Kunci

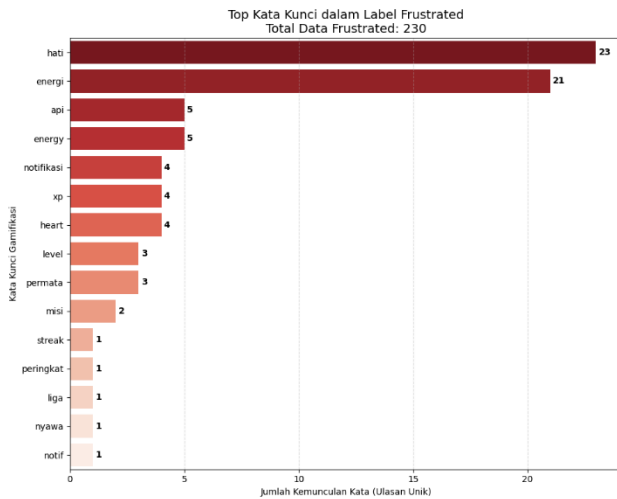
Berdasarkan analisis terhadap Gambar 6, fitur Penalti melalui kata kunci “hati” dan “heart” mendominasi label *Frustrated* (52.3% dan 66.7%) yang menunjukkan keterkaitan antara mekanisme pembatasan belajar dengan munculnya respon emosional negatif dalam ulasan pengguna. Sebaliknya, elemen kompetisi dan pencapaian seperti “rangking”, “gem”, dan “poin” seluruhnya muncul pada ulasan berlabel *Motivated* (100%).

5. Representasi Label Motivated



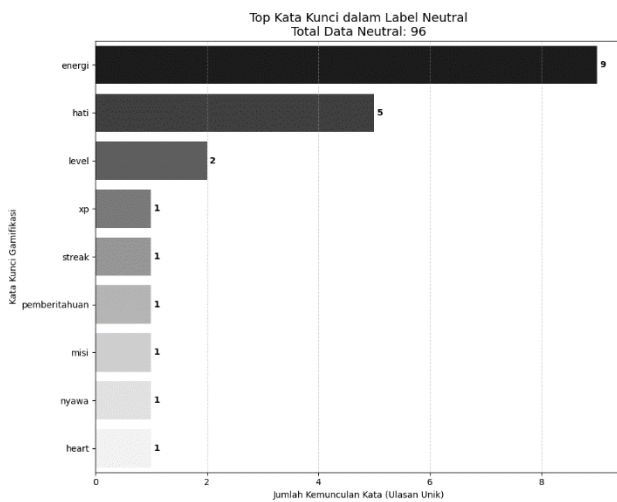
Gambar 7. Top Kata Kunci pada Label Motivated

6. Representasi Label Frustrated



Gambar 8. Top Kata Kunci pada Label Frustrated

7. Representasi Label Neutral



Gambar 9. Top Kata Kunci pada Label Neutral

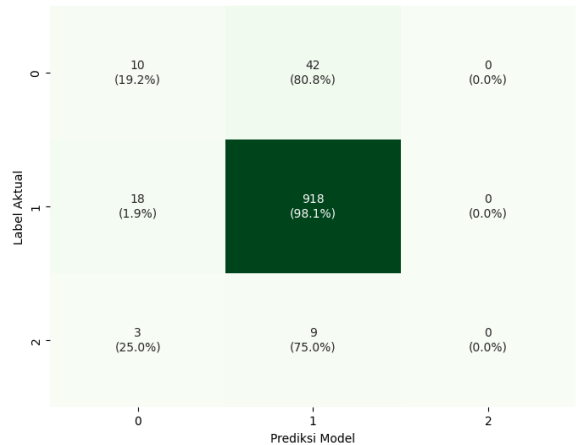
B. Implementasi Teknik Penyeimbangan Data

Dataset yang digunakan memiliki sebaran label yang tidak merata (*imbalanced*). Kondisi ini dapat menyebabkan model klasifikasi memiliki bias terhadap kelas mayoritas dan mengabaikan kelas minoritas yang justru seringkali mengandung informasi krusial [31]. Oleh karena itu, penelitian ini menerapkan dua strategi berbeda untuk mengatasi kendala *class-skewed* tersebut, yaitu manipulasi tingkat data melalui SMOTE dan penyesuaian tingkat algoritma melalui *Class Weight*.

1. SMOTE pada Naïve Bayes

Untuk algoritma Naïve Bayes, diterapkan teknik *Synthetic Minority Over-Sampling Technique* (SMOTE). Teknik ini bekerja dengan cara menghasilkan sampel sintesis pada kelas minoritas agar distribusi data menjadi seimbang, sehingga mencegah model dari kecenderungan memihak pada kelas mayoritas [32].

Confusion Matrix: Naive Bayes (SMOTE)
Akurasi: 92.80%

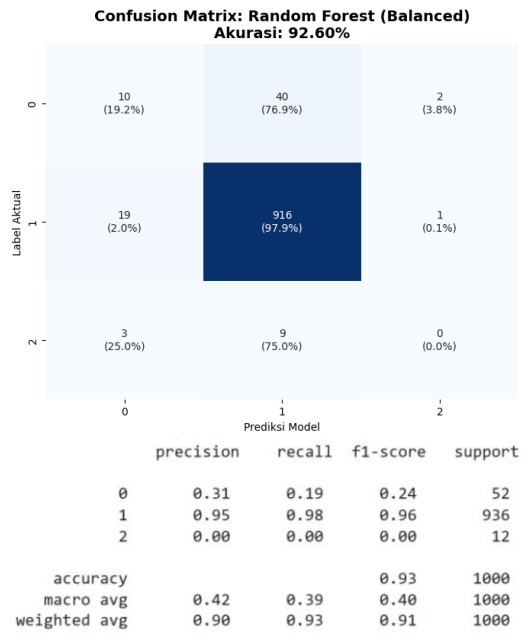


	precision	recall	f1-score	support
0	0.32	0.19	0.24	52
1	0.95	0.98	0.96	936
2	0.00	0.00	0.00	12
accuracy			0.93	1000
macro avg	0.42	0.39	0.40	1000
weighted avg	0.90	0.93	0.91	1000

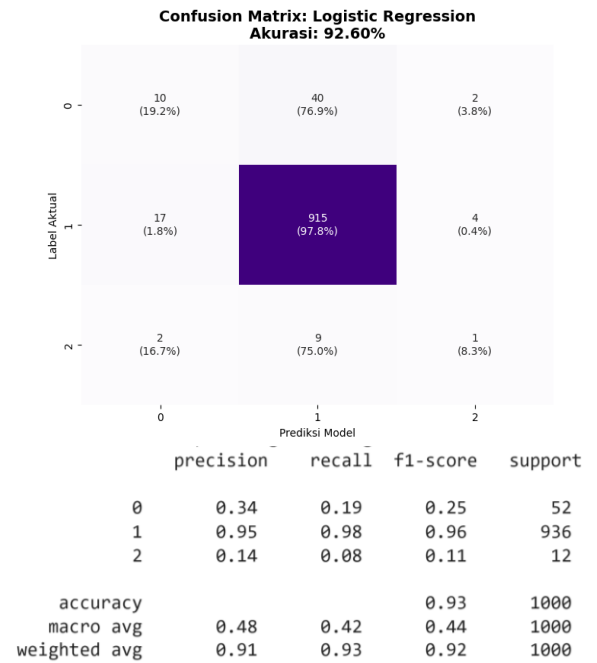
Gambar 10. Confusion Matrix dan Akurasi Naïve Bayes sesudah SMOTE

2. Class Weight pada Random Forest

Berbeda dengan Naïve Bayes, algoritma Random Forest menggunakan teknik *Class Weight* yang menerapkan prinsip *Cost-Sensitive Learning* [33]. Teknik ini bekerja dengan memodifikasi proses pembelajaran model melalui pemberian bobot penalti yang lebih besar pada kesalahan prediksi kelas minoritas (*Frustrated* dan *Neutral*).

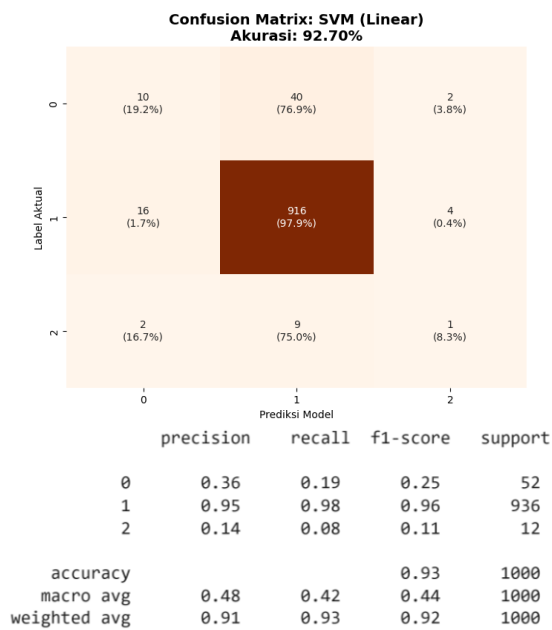


Gambar 11. Confusion Matrix dan Akurasi Random Forest setelah Class Weight



Gambar 13. Confusion Matrix dan Akurasi Logistic Regression

3. Support Vector Machine (SVM)



Gambar 12. Confusion Matrix dan Akurasi SVM

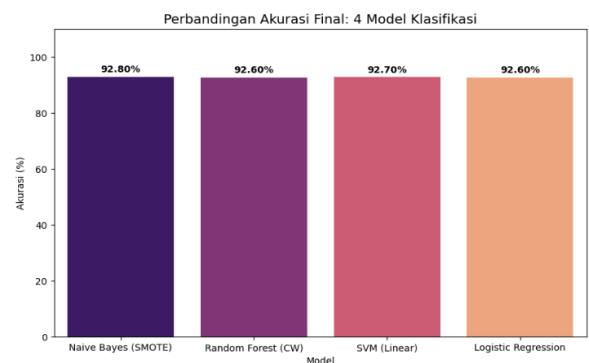
4. Logistic Regression (LR)

C. Perbandingan Performa Model

Pemilihan keempat model ini didasarkan pada karakteristik data teks ulasan yang memiliki variasi linguistik tinggi. Naïve Bayes dipilih sebagai *baseline* karena efisiensinya dalam menangani fitur teks pendek berdasarkan probabilitas kemunculan kata. *Random Forest* disertakan untuk menguji kemampuan model berbasis *ensemble tree* dalam menangani potensi hubungan non-linear antar fitur. Sementara itu, Support Vector Machine (SVM) dan Logistic Regression dipilih karena kemampuannya yang teruji dalam menangani ruang fitur yang bersifat *high-dimensional* dan *sparse* [19].

TABEL 8
RINGKASAN PERFORMA

Model	Akurasi	Macro Avg (F1)
Naïve Bayes	0.93	0.40
Random Forest	0.93	0.40
Support Vector Machine	0.93	0.44
Logistic Regression	0.93	0.44



Gambar 14. Perbandingan akurasi akhir model

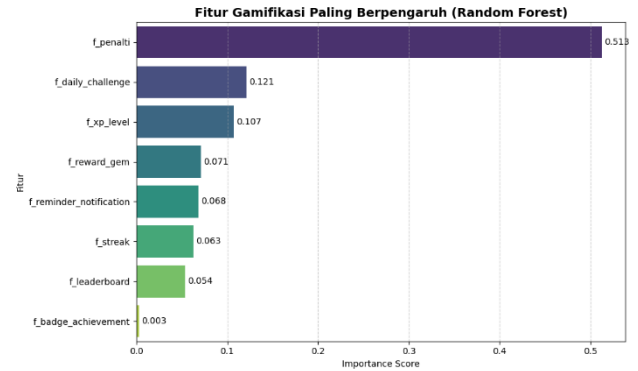
Berdasarkan hasil pengujian pada Tabel 8, keempat model menunjukkan performa yang kompetitif setelah penerapan teknik optimasi (SMOTE pada Naïve Bayes serta *Class Weight* pada Random Forest, SVM, dan Logistic Regression). Meskipun penggunaan SMOTE membawa potensi risiko overfitting karena penggunaan data sintesis, hal ini dimitigasi dengan melakukan evaluasi performa pada data uji asli yang tidak disintesis.

Secara akurasi, Naïve Bayes unggul tipis dengan skor 92.80%, diikuti oleh SVM (92.70%) serta Random Forest dan *Logistic Regression* (92.60%). Namun, jika meninjau nilai *Macro Average F1-Score*, model SVM dan *Logistic Regression* memberikan performa terbaik dengan skor 0.44, mengungguli Naïve Bayes dan Random Forest yang hanya mencapai 0.40. Hal ini membuktikan bahwa model linear seperti SVM dan LR memiliki sensitivitas yang lebih tinggi dalam mengklasifikasikan kelas minoritas (*Frustrated* dan *Neutral*). Temuan ini didukung oleh penelitian Hassan dkk. [19] yang mengonfirmasi bahwa karakteristik model linear secara konsisten memberikan performa klasifikasi yang lebih unggul dibandingkan algoritma berbasis probabilitas atau *tree*. Keunggulan ini terletak pada kemampuan model dalam menentukan *Maximum Margin Hyperplane* (MMH), yang memungkinkan pemisahan kelas secara optimal pada ruang fitur yang kompleks. Hal tersebut memberikan stabilitas prediksi yang lebih baik, terutama dalam menangani ambiguitas kata pada dataset dengan sebaran kelas yang tidak seimbang (*imbalanced*).

Meskipun demikian, nilai *Macro Average F1-Score* yang relatif rendah menunjukkan bahwa performa model belum merata. Berdasarkan analisis *confusion matrix* pada keempat model, pola kesalahan yang konsisten terlihat pada kegagalan klasifikasi kelas minoritas, khususnya kelas 2. Pada Naïve Bayes dan *Random Forest*, kelas ini tidak pernah terprediksi secara benar ($\text{recall} = 0.00$), sedangkan pada SVM dan *Logistic Regression* hanya terdapat satu prediksi benar ($\text{recall} = 0.08$). Selain itu, kelas 0 juga menunjukkan recall yang rendah (0.19) pada seluruh model dimana sebagian besar sampel salah diklasifikasikan ke kelas mayoritas. Hal ini menunjukkan bahwa rendahnya *Macro Average F1-Score* merupakan fenomena sistematis akibat ketidakseimbangan distribusi data dan tumpang tindih fitur antar kelas pada teks pendek. Dengan demikian, *Macro Average F1-Score* memberikan gambaran evaluasi yang lebih representatif dibandingkan akurasi dalam menilai performa model pada skenario multi-kelas yang tidak seimbang.

D. Evaluasi Feature Importance

Sebagai bagian dari interpretasi model, dilakukan analisis untuk melihat fitur mana yang paling menentukan hasil klasifikasi psikologis pengguna.



Gambar 15. Fitur Gamifikasi Paling Berpengaruh

Berdasarkan hasil pemodelan menggunakan Random Forest, diperoleh tingkat kepentingan fitur (*Feature Importance*) dalam mengklasifikasikan respon psikologis pengguna sebagaimana disajikan pada Gambar 15. Fitur Penalti menempati posisi puncak sebagai fitur yang paling berpengaruh dalam proses pengambilan keputusan model. Hal ini selaras dengan distribusi frekuensi, di mana sistem penalti merupakan elemen yang paling banyak menyita perhatian dan memicu emosi pengguna secara konsisten.

Temuan menarik muncul pada fitur *Daily Challenge* yang menempati urutan kedua dalam skor kepentingan meskipun frekuensi kemunculannya jauh di bawah fitur *Streak*. Hasil ini membuktikan bahwa nilai kepentingan sebuah fitur tidak hanya ditentukan oleh kuantitas atau frekuensi kemunculannya, melainkan oleh kekuatan makna dan konsistensi pola informasi yang dikandungnya sehingga memudahkan model dalam membedakan label psikologis secara akurat, sebagaimana dijelaskan dalam studi mengenai interpretabilitas model klasifikasi teks [34]. Dengan kata lain, fitur yang secara semantik merepresentasikan pengalaman spesifik pengguna dapat memiliki daya prediktif yang lebih tinggi meskipun jarang disebutkan. Temuan ini menegaskan pentingnya analisis fitur berbasis makna, bukan hanya berbasis frekuensi, dalam memahami respons psikologis pengguna.

E. Diskusi dan Temuan Utama

Berdasarkan serangkaian analisis yang telah dilakukan, penelitian ini menghasilkan beberapa temuan utama sebagai berikut:

1. Fitur Penalti (energi, hati, nyawa) merupakan elemen yang paling dominan dibahas oleh pengguna dengan 141 ulasan dan memiliki pengaruh terbesar dalam klasifikasi psikologis. Temuan ini mengonfirmasi adanya fenomena *loss aversion*, di mana secara psikologis pengguna cenderung memberikan respon emosional yang lebih kuat terhadap mekanisme kehilangan atau hukuman dibandingkan fitur pencapaian dengan nilai yang setara [35]. Dalam perspektif *Self-Determination Theory* (SDT), mekanisme penalti dapat memengaruhi motivasi pengguna karena berpotensi mengurangi persepsi otonomi dan kompetensi ketika pengguna merasa dikontrol atau

dihukum oleh sistem [2]. Selain itu, tekanan psikologis akibat sistem pembatasan belajar berkaitan erat dengan peningkatan beban kognitif yang mengganggu konsentrasi dan secara negatif mempengaruhi motivasi intrinsik pengguna [6]. Hal ini menjelaskan mengapa sistem penalti memicu tingkat frustrasi yang lebih tinggi dibandingkan perolehan trofi digital dalam dataset ini.

2. Terjadi polaritas respon yang tajam pada fitur Penalti. Kata kunci “hati” menjadi penyumbang pada label *Frustrated* tertinggi (52.3%) sementara kata kunci “energi” menunjukkan proses adaptasi dengan dominasi label *Motivated* (69.7%). Temuan ini mengindikasikan bahwa meskipun mekanisme pembatasan belajar cenderung memicu respon negatif, sebagian pengguna mulai menerima sistem “energi” sebagai bagian dari ekosistem motivasi dalam proses pembelajaran.
3. Meskipun fitur *Daily Challenge* memiliki frekuensi rendah, posisinya sebagai fitur terpenting kedua menunjukkan perannya sebagai titik cabang keputusan yang sangat informatif, sesuai dengan prinsip bahwa fitur dengan frekuensi rendah namun memiliki spesifisitas tinggi sering kali menjadi prediktor utama dalam model *machine learning* [34]. Secara perilaku, hal ini merepresentasikan fitur dengan dampak emosional instan, tantangan harian dapat menjadi pendorong motivasi sekaligus sumber frustrasi jika gagal dicapai dalam batas waktu tertentu.
4. Penerapan teknik SMOTE pada Naïve Bayes serta *Class Weight* pada Random Forest, SVM, dan *Logistic Regression* terbukti efektif meningkatkan kemampuan model dalam mengenali kelas minoritas (*Frustrated* dan *Neutral*). Hal ini ditandai dengan kenaikan *Macro Average F1-Score* dari 0.35 menjadi 0.40 pada Naïve Bayes dan Random Forest, serta hasil yang lebih unggul pada SVM dan *Logistic Regression* sebesar 0.44. Meskipun Naïve Bayes memiliki akurasi tertinggi, model SVM dan *Logistic Regression* menjadi yang paling optimal dalam memberikan keseimbangan prediksi dan sensitivitas terhadap respon psikologi pengguna dalam konteks gamifikasi ini.

Meskipun penelitian ini memberikan temuan yang signifikan terkait pengaruh fitur gamifikasi terhadap kondisi psikologis pengguna, terdapat beberapa keterbatasan yang perlu diperhatikan. Pertama, data ulasan yang dikumpulkan dari Google Play Store cenderung memiliki bias ekstrem, di mana pengguna umumnya hanya memberikan ulasan ketika merasa sangat puas atau sangat kecewa, sehingga sebaran data kelas netral menjadi lebih sedikit. Kedua, analisis kesalahan menunjukkan adanya tumpang tindih fitur yang tinggi antar kelas, hal ini menyebabkan model seringkali salah mengklasifikasikan kelas minoritas sebagai kelas mayoritas karena kemiripan diksi atau kata kunci yang digunakan oleh pengguna. Terakhir, analisis ini terbatas pada aplikasi Duolingo, sehingga temuan mengenai dinamika psikologis elemen gamifikasi mungkin memerlukan validasi lebih lanjut

jika diterapkan pada platform pembelajaran digital dengan karakteristik pengguna yang berbeda.

IV. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa optimasi model melalui teknik penanganan *imbalanced data* sangat krusial dalam mengklasifikasikan respon psikologi pengguna secara akurat. Penerapan SMOTE dan *Class Weight* pada keempat model yang diuji mampu meningkatkan nilai *Macro Average F1-Score* secara signifikan. Meskipun seluruh model memiliki akurasi kompetitif di kisaran 92%, model SVM dan *Logistic Regression* memberikan performa paling optimal dengan nilai F1-Score sebesar 0.44, mengungguli Naïve Bayes dan Random Forest dengan score sebesar 0.40. Ini membuktikan bahwa model linear lebih objektif dalam mendeteksi kelas minoritas yang membawa informasi kritis mengenai titik lemah fitur gamifikasi. Secara keseluruhan, SVM dan *Logistic Regression* direkomendasikan sebagai model terbaik untuk klasifikasi ini, sementara Random Forest tetap memberikan keunggulan dalam aspek interpretabilitas fitur melalui analisis *Feature Importance*.

Dari sisi pengalaman pengguna, temuan ini mengungkapkan bahwa fitur Penalti merupakan pemicu emosi yang paling dominan. Hal ini mengonfirmasi adanya fenomena *loss aversion*, di mana pengguna cenderung merespon mekanisme kehilangan secara lebih intens dibandingkan sebuah pencapaian. Selain itu, fitur *Daily Challenge* teridentifikasi sebagai titik cabang keputusan yang sangat informatif bagi model karena konsistensi pola respons pengguna yang kontras antara motivasi dan frustrasi. Temuan ini memberikan gambaran bagi pengembang aplikasi edukasi untuk lebih menyeimbangkan mekanisme penalti agar tidak menjadi hambatan psikologis, melainkan tetap menjadi pendorong motivasi yang sehat dalam proses belajar.

Secara praktis, hasil penelitian ini menunjukkan bahwa desain gamifikasi sebaiknya mempertimbangkan proporsi antara mekanisme *reward* dan *punishment* secara lebih adaptif. Mekanisme penalti yang terlalu sering atau tidak disertai umpan balik yang konstruktif berpotensi meningkatkan frustrasi dan menurunkan keterlibatan pengguna dalam jangka panjang. Oleh karena itu, pengembang dapat mempertimbangkan penerapan sistem penalti yang bersifat progresif, pemberian kesempatan pemulihan, atau umpan balik reflektif yang membantu pengguna memahami kesalahan sebagai bagian dari proses belajar.

DAFTAR PUSTAKA

- [1] F. Y. Pratama, V. Y. Erviana, I. Suwartini, and M. N. Antono, "Gamifikasi Dalam Kelas Bahasa Indonesia Bagi Penutur Asing (BIPA) Anak," *Scholaria: Jurnal Pendidikan dan Kebudayaan*, vol. 15, no. 1, pp. 60-67, Jan. 2025.
- [2] F. Gao, *Advancing Gamification Research and Practice with Three Underexplored Ideas in Self-Determination Theory*, TechTrends, vol. 68, pp. 661-671, Jun. 2024. doi: 10.1007/s11528-024-00968-9.
- [3] S. A. Pangestu, P. Setyosari, and D. Kuswandi, "Analisis Persepsi Mobile Learning Readiness dengan Menggunakan Technology

- Acceptance Model dalam Pembelajaran Bahasa Jawa," JKTP: Jurnal Kajian Teknologi Pendidikan, vol. 6, no. 3, pp. 152-162, Agt. 2023.
- [4] W. D. P. F. Dewi and M. H. Santosa, "A Systematic Review: The Implementation of MALL toward Students' Reading Skill," Jurnal Pendidikan Bahasa Inggris Undiksha, vol. 10, no. 2, pp. 219-225, 2022.
- [5] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020," Computer Assisted Language Learning, vol. 36, no. 3, pp. 517-554, 2023.
- [6] P. Evans, M. Vansteenkiste, P. D. Parker, A. Kingsford-Smith, dan S. Zhou, Cognitive Load Theory and Its Relationships with Motivation: a Self-Determination Theory Perspective, Educational Psychology Review, vol. 36, art. 7, 2024. doi: 10.1007/s10648-023-09841-2.
- [7] T. A. Sandy, A. Ghufron, A. Muhtadi, and Pujiriyanto, "Text Classification of Duolingo Reviews on Google Play: Insights for Enhancing M-Learning Applications," International Journal of Interactive Mobile Technologies (IJIM), vol. 19, no. 7, pp. 210-225, 2025.
- [8] A. Serlina, A. Rahim, and Arbansyah, "Comparative Analysis of Naïve Bayes Algorithm Performance in English and Indonesian Text Sentiment Classification on Duolingo Application in Playstore," Informatics Engineering, Science and Technology, vol. 1, no. 1, pp. 165-171, Feb. 2025.
- [9] C. Salazar, E. Montoya-Múnica, and J. Aguilar, "Sentiment analysis in learning resources," J. Comput. Educ., vol. 10, no. 4, pp. 637-662, 2023, doi: 10.1007/s40692-022-00237-9.
- [10] A. Lighthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," Artif. Intell. Rev., vol. 54, pp. 4997-5053, 2021, doi: 10.1007/s10462-021-09973-3.
- [11] W. A. Awadh, R. B. Sulaiman, and M. A. Mahmoud, "Aspect-based sentiment analysis in MOOCs: a systematic literature review introducing the MASC-MEF framework," J. King Saud Univ. - Comput. Inf. Sci., vol. 37, no. 2, 2025, doi: 10.1007/s44443-025-00018-1.
- [12] L. Davoodi, J. Mezei, and M. Heikkilä, "Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms," Electron. Commer. Res., 2025, doi: 10.1007/s10660-025-09948-4.
- [13] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," Artif. Intell. Rev., vol. 57, no. 296, 2024, doi: 10.1007/s10462-024-10906-z.
- [14] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," Nat. Lang. Process. J., vol. 2, 100003, 2023, doi: 10.1016/j.nlp.2023.100003.
- [15] H. Junianto, P. Arsi, B. A. Kusuma, and D. I. S. Saputra, "Evaluasi Aplikasi Raileo Melalui Analisis Sentimen Ulasan Playstore Dengan Metode Naive Bayes," SINTECH J., vol. 7, no. 1, pp. 1-10, 2024.
- [16] M. N. Ramadhani, K. D. Tania, and M. Afrina, "Knowledge discovery in sharia mobile banking reviews using aspect-based sentiment analysis and machine learning," Journal of Applied Informatics and Computing (JAIC), vol. 10, no. 1, pp. 640-650, Feb. 2026.
- [17] D. N. Fahria, K. D. Tania, and R. D. Kurnia, "Analisis komparatif algoritma Random Forest, XGBoost, dan CatBoost untuk klasifikasi tingkat stres pengguna media sosial," RABIT: Jurnal Teknologi dan Sistem Informasi Univrab, vol. 11, no. 1, pp. 1843-1853, Jan. 2026, doi: 10.36341/rabit.v11i1.7449.
- [18] N. Jalal, A. Mehmood, G. S. Choi, and I. Ashraf, "A novel improved random forest for text classification using feature ranking and optimal number of trees," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 5, pp. 2733-2742, 2022.
- [19] S. U. Hassan, J. Ahamed, dan K. Ahmad, "Analytics of machine learning-based algorithms for text classification," Sustainable Operations and Computers, vol. 3, pp. 238-248, 2022. doi: 10.1016/j.susoc.2022.03.001.
- [20] E. T. Pardede, K. D. Tania, and M. Afrina, "Knowledge discovery: Analisis sentimen dan emosi WhatsApp Business dengan machine learning dan deep learning," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 11, no. 3, pp. 310-318, Jan. 2026, doi: 10.25077/TEKNOSI.v11i3.2025.310-318.
- [21] M. Sahira, K. D. Tania, and M. Afrina, "Comparative analysis of machine learning algorithms for cosmetic sales prediction on Tokopedia," JURTEKSI (Jurnal Teknologi dan Sistem Informasi), vol. 11, no. 4, pp. 701-708, Sep. 2025, doi: 10.33330/jurteksiv11i4.4187.
- [22] R. Saha et al., "The psychological and emotional factors of gamification and customer experience. A systematic literature review and future research agenda," Acta Psychologica, vol. 260, 105659, 2025.
- [23] M. C. Untoro and M. Farhan, "Sentiment Analysis of Public Opinion on BAWASLU Using Random Forest and Particle Swarm Optimization," Sci. J. Informatics, vol. 12, no. 1, pp. 182-192, 2025.
- [24] V. Pichiyen et al., "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis," Procedia Comput. Sci., vol. 230, pp. 193-202, 2023, doi: 10.1016/j.procs.2023.12.075.
- [25] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," Glob. Transitions Proc., vol. 3, no. 1, pp. 91-99, 2022.
- [26] C. C. Oisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," Comput. Methods Programs Biomed., vol. 220, p. 106773, 2022.
- [27] T. Nomoto, "Keyword Extraction: A Modern Perspective," SN Comput. Sci., vol. 4, no. 1, p. 92, 2023, doi: 10.1007/s42979-022-01481-7.
- [28] S. Albahra et al., "Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts," Semin. Diagn. Pathol., vol. 40, no. 2, pp. 71-87, 2023.
- [29] G. P. Satria, K. S. Gunawan, and Z. E. Rasjida, "Aspect-Based Sentiment Analysis of Multilingual Customer Product Reviews: Leveraging Unsupervised Aspect Extraction with ABAE for XLM-ROBERTa-based Classification," Procedia Comput. Sci., vol. 269, pp. 239-248, 2025, doi: 10.1016/j.procs.2025.01.031.
- [30] T. K. Thivakaran and M. Ramesh, "Exploratory Data analysis and sales forecasting of bigmart dataset using supervised and ANN algorithms," Meas. Sensors, vol. 23, p. 100388, 2022, doi: 10.1016/j.measen.2022.100388.
- [31] M. Z. Abedin, C. Guotai, P. Hajek, and T. Zhang, "Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk," Complex Intell. Syst., vol. 9, no. 4, pp. 3559-3579, 2023.
- [32] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, "RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 8, pp. 5059-5074, 2022.
- [33] I. Araf, A. Idri, and I. Chairi, "Cost-sensitive learning for imbalanced medical data: a review," Artif. Intell. Rev., vol. 57, no. 4, p. 80, 2024.
- [34] T. Ling, J. Luo, J. Adams, K. Osinski, X. Liu, and D. Friedland, "Interpretable machine learning text classification for clinical computed tomography reports: a case study of temporal bone fracture," Comput. Methods Programs Biomed. Update, vol. 3, p. 100104, 2023, doi: 10.1016/j.cmpbup.2023.100104.
- [35] S. Gächter, E. J. Johnson, and A. Herrmann, "Individual-level loss aversion in riskless and risky choices," Theory Decis., vol. 92, no. 3, pp. 599-624, 2022, doi: 10.1007/s11238-021-09839-8.