e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 9 No. 2, Agustus, 2025, **Hal. 412-421**
DOI: 10.29408/edumatic.v9i2.30133

Analisis Sentimen Ulasan Game dengan KNN: Perbandingan Rating dan Kamus Sentimen

Wisesa Sat Sunyaruri [1,*] , Novita Kurnia Ningrum [1]

[1] Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia
[*] Correspondence: wisesasats@gmail.com

Received: 2 Mei 2025 | Revised: 13 Juni 2025 | Accepted: 11 Juli 2025 | Published: 11 Agustus 2025

Abstrak
Pertumbuhan industri *game* global menjadikan analisis sentimen ulasan sebagai alat krusial untuk memahami kepuasan dan mengidentifikasi masalah teknis. Penelitian ini bertujuan mengevaluasi tiga metode pelabelan (*rating*, Sentiwords_id, dan InSet) dalam mengklasifikasikan sentimen ulasan *game* Zenless Zone Zero (ZZZ) berbahasa Indonesia menggunakan algoritma K-Nearest Neighbour (KNN). Penelitian ini menganalisis 4.282 ulasan *Google Play Store* melalui tahap Pra-pemrosesan Data, meliputi *null handling, cleaning, case folding*, tokenisasi, *stopword removal*, dan *stemming*. Kinerja KNN dari setiap metode pelabelan dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* pada data uji dengan rasio 80:20. Hasil pelabelan menunjukkan persepsi sentimen yang berbeda: metode berbasis *rating* cenderung positif, InSet cenderung negatif, sementara *Sentiwords_id* didominasi oleh kelas positif dan netral. Evaluasi kinerja KNN menunjukkan *labelling rating* mencapai akurasi tertinggi (72%), unggul pada kelas positif (*recall* 86%) namun lemah pada kelas netral (*recall* 9%). Sebaliknya, *labelling* leksikon/kamus sentimen (keduanya akurasi 69%) memiliki kekuatan spesifik: InSet pada deteksi negatif (*recall* 81%) dan Sentiwords_id pada kelas netral (*recall* 83%). Tantangan utama penelitian ini meliputi keterbatasan leksikon dalam menangani slang dan istilah *game*, serta ketidakselarasan antara *rating* dan teks. Penelitian ini diharapkan memberikan bukti empiris mengenai *trade-off* kinerja antar metode pelabelan otomatis untuk membantu identifikasi kepuasan pemain guna memajukan kualitas pengembangan *game*.

Kata kunci: analisis sentimen; ulasan game; bahasa indonesia; knn; sentiwords_id

Abstract
The growth of the global gaming industry makes sentiment analysis of user reviews a crucial tool for understanding satisfaction and identifying technical issues. This study aims to evaluate three labelling methods (rating-based, Sentiwords_id, and InSet) for classifying the sentiment of Indonesian-language reviews for the game Zenless Zone Zero (ZZZ) using the K-Nearest Neighbor (KNN) algorithm. The study analyzes 4,282 reviews from the Google Play Store, which underwent a Data Preprocessing stage, including Null Handling, Cleaning, Case Folding, Tokenization, Stopword Removal, and Stemming. The KNN's performance for each labelling method was evaluated using accuracy, precision, recall, and F1-score metrics on 80:20 train-test split. The labelling results reveal different sentiment perceptions: the rating-based method tends toward positive, InSet toward negative, while Sentiwords_id is dominated by the positive and neutral classes. The KNN performance evaluation shows that rating-based labelling achieved the highest accuracy (72%), excelling on the positive class (86% recall) but performing poorly on the neutral class (9% recall). Conversely, the lexicon-based labelling methods (both 69% accuracy) have specific strengths: InSet in negative detection (81% recall) and Sentiwords_id in recognizing the neutral class (83% recall). Main challenges of this study include the lexicon's limitations in handling slang and game-specific terms, as well as the inconsistency between ratings and text. This study is expected to provide empirical evidence on performance trade-offs among

automatic labelling methods to aid in identifying player satisfaction and advancing the quality of game development.

Keywords: *sentiment analysis; game review; indonesian language; knn; sentiwords_id*

PENDAHULUAN

Industri *game* global mengalami pertumbuhan pesat, dengan miliaran pengguna aktif. Salah satu platform distribusi utamanya, *Google Play Store*, menyediakan ratusan ribu judul *game* yang dapat diakses, diunduh, dan diperbarui (Zhang, 2024). Platform ini juga dilengkapi sistem ulasan terintegrasi yang memungkinkan pengguna memberikan masukan, mulai dari pujian hingga keluhan terkait performa atau konten *game* (Ulya et al., 2022). Mengingat volume ulasan yang masif dan terus bertambah, analisis otomatis terhadap ulasan pengguna menjadi krusial untuk memahami kepuasan serta mengidentifikasi masalah teknis guna meningkatkan kualitas produk (Saninah et al., 2025).

Role-Playing Game (RPG) populer seperti Zenless Zone Zero (ZZZ) menjadi contoh relevan, yang menerima puluhan ribu ulasan dalam berbagai bahasa, termasuk Bahasa Indonesia. Ulasan *game* sering kali lebih kompleks dibandingkan ulasan produk biasa karena mengandung informasi penting seperti laporan *bug* atau permintaan fitur (Cahyaningtyas et al., 2021) yang menjadi sumber kritis untuk meningkatkan kualitas produk (Mustofa & Idris, 2024). Selain itu, ulasan ini sarat dengan bahasa komunitas yang dinamis, termasuk *slang*, istilah teknis (*bug*, *lag*), dan jargon *gameplay* yang umum di kalangan *gamer* (Magria et al., 2021). Volume yang besar dan kompleksitas bahasa yang unik ini menuntut pendekatan evaluasi otomatis yang efektif. Untuk dapat menganalisis data dalam skala besar ini, teknik seperti *web scraping* (Abodayeh et al., 2023; Khder, 2021) umumnya digunakan untuk mengumpulkan data ulasan secara otomatis, sering kali dengan memanfaatkan *library* Python seperti google-play-scraper (Nur et al., 2024).

Salah satu tantangan utama dalam menganalisis sentimen pengguna terletak pada ketidakselarasan *rating* numerik dengan sentimen tekstual, dimana *rating* tinggi tidak selalu berarti ulasan positif (Sadiq et al., 2021). Sebagai contoh, seorang *user* dapat memberikan *rating* bintang 5 tetapi menuliskan keluhan tajam seperti: "Power musuh terlalu sakit dan tombol Dodge yang sangat delay... gausah bikin game kalo ini game terlalu hard buat f2p... Uninstall". Fenomena *mismatch* ini menyebabkan *noise* pada data latih dan menunjukkan bahwa metode ini rentan terhadap bias (Khamket & Polpinij, 2023), meskipun dapat merepresentasikan penilaian langsung dari pengguna (Sadiq et al., 2021). Di sisi lain, pelabelan berbasis kamus sentimen, yang mengandalkan identifikasi kata kunci, menghadapi tantangan berbeda, terutama untuk Bahasa Indonesia. Permasalahan utamanya adalah ketergantungan pada leksikon yang jumlahnya terbatas, seperti SentiStrength_ID (Kusumastuti et al., 2022) dan InSet (Artana et al., 2023), yang dirancang untuk domain umum dan sering kali gagal menangkap konteks spesifik ulasan *game*.

Sebagai solusi menghadapi tantangan tersebut, penelitian ini menerapkan analisis sentimen atau *opinion mining* yang berupa proses komputasional untuk mengidentifikasi dan mengekstrak opini subjektif dari data teks (Rahayu et al., 2022). Langkah teknis pertama adalah merepresentasikan teks ulasan menjadi format numerik melalui ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata dalam ulasan berdasarkan signifikansinya, yaitu dimana bobot akan bernilai tinggi jika sebuah kata sering muncul dalam satu ulasan (*Term Frequency*) namun jarang muncul di ulasan-ulasan lain (*Inverse Document Frequency*), sehingga menonjolkan kata-kata yang paling representatif (Cahyani & Patasik, 2021; Mahendra et al., 2023).

Setelah teks diubah menjadi vektor TF-IDF, untuk klasifikasi kelas sentimen menggunakan algoritma *supervised learning* K-Nearest Neighbor (KNN). KNN adalah metode

non-parametrik yang mengklasifikasikan sebuah ulasan baru dengan cara melihat sentimen mayoritas dari 'K' atau jumlah tetangga terdekatnya di dalam ruang fitur/dimensi. Kedekatan antar data dalam penelitian ini diukur menggunakan *cosine similarity* (Halder et al., 2024).

Berbagai penelitian sebelumnya telah menerapkan analisis sentimen pada ulasan berbahasa Indonesia. Terdapat beberapa penelitian yang menggunakan pelabelan otomatis berbasis *rating* untuk melatih model KNN (Mustaqim et al., 2024), sementara penelitian lain memanfaatkan leksikon sentimen bahasa Indonesia seperti InSet (Artana et al., 2023) dan *SentiStrength_ID* (Firmansyah et al., 2024), atau *SentiWordNet* melalui proses translasi (Pamungkas & Cahyono, 2024; Rahayu et al., 2022). Terdapat pula penelitian yang berfokus membandingkan performa leksikon dengan label manual yang dianggap sebagai 'kebenaran dasar' (Abdillah et al., 2021; Artana et al., 2023). Meskipun demikian, dari penelusuran literatur tersebut, belum ditemukan adanya kajian yang secara langsung membandingkan efektivitas antara metode pelabelan otomatis berbasis *rating* dengan metode pelabelan otomatis berbasis leksikon Bahasa Indonesia pada satu *dataset* yang sama, khususnya dalam domain ulasan *game* yang dinamis.

Penelitian kami dilakukan selain membantu pengembang *game* untuk mengidentifikasi kepuasan pemain, juga untuk memperkaya kajian analisis sentimen. Tujuan penelitian kami adalah untuk mengevaluasi kinerja tiga metode pelabelan data, yaitu berbasis *rating*, kamus sentimen Sentiwords_id, dan InSet. Dengan demikian, penelitian ini diharapkan dapat membantu pengembang *game* mengidentifikasi kepuasan pemain guna memajukan kualitas pengembangan *game,* khususnya ZZZ.

METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan kerangka studi komparatif untuk mengevaluasi tiga metode *labelling* (*rating*, Sentiwords_id, dan InSet) dalam analisis sentimen ulasan *game* ZZZ berbahasa Indonesia. Proses penelitian meliputi enam tahap utama: Pengumpulan Data, Pra-pemrosesan Data, Pelabelan Data, Ekstraksi Fitur, Pemodelan, dan Evaluasi. Proses pengumpulan data ulasan (komentar teks dan *rating* 1-5 bintang) diperoleh dari Google Play Store melalui *web scraping* dengan *library Python google-play-scraper*.

Data mentah kemudian melewati tahap pra-pemrosesan untuk membersihkan dan mempersiapkan teks. Tahapan ini dilakukan secara berurutan, meliputi *Null Handling*, *Cleaning* (penghapusan tanda baca dan karakter non-alfabet), *Case Folding* (mengubah teks menjadi huruf kecil), Tokenisasi (pemisahan kata), *Stopword Removal* (menghilangkan kata tidak bermakna penting), dan *Stemming* (menghilangkan imbuhan). Untuk proses *Stopword Removal* dan *Stemming* spesifik Bahasa Indonesia, digunakan *library* Sastrawi.

Data yang telah bersih kemudian diberi label menggunakan dua pendekatan utama. Pertama, metode berbasis *rating*, dimana ulasan dengan *rating* 4-5 diberi label ‘positif’, *rating* 3 ‘netral’, dan *rating* 1-2 ‘negatif’. Kedua, metode berbasis leksikon, yang menerapkan kamus InSet dan Sentiwords_id untuk menghasilkan skor sentimen, dimana skor > 0 menjadi ‘positif’, skor = 0 ‘netral’, dan skor < 0 ‘negatif’. Sebagai batasan, penelitian ini tidak melakukan normalisasi slang atau kata tidak baku dan hanya menggunakan Sentiwords_id tanpa kamus pelengkapnya yaitu SentiStrength_ID untuk menjaga kesetaraan perbandingan dengan InSet. Hasil label yang bisa didapatkan dari tahap *Data Labelling* dapat dilihat pada Tabel 1.

Setelah pelabelan, data teks diubah menjadi vektor numerik menggunakan TF-IDF, dengan batasan hanya menggunakan *unigram* dan tanpa seleksi fitur. *Dataset* kemudian dibagi menjadi 80% data latih dan 20% data uji melalui skema *single split* tanpa *stratified sampling*. Model klasifikasi K-NN selanjutnya dilatih dengan parameter k=3 dan metrik *cosine similarity*, meskipun tanpa optimasi *hyperparameter* untuk nilai k. Kinerja akhir model dari setiap metode pelabelan dievaluasi menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan dianalisis melalui *confusion matrix* untuk mendapatkan gambaran yang komprehensif.

Tabel 1. Hasil *data labelling*

Labelling Method	**Bahan**	**Hasil**
Rating	Ulasan Bintang 4 & 5	Positif
	Ulasan Bintang 3	Netral
	Ulasan Bintang 1 & 2	Negatif
Sentiwords_id	['gameplay', 'senang', 'cerita', 'ga', 'bosan', 'ikut', 'developer', 'baik', 'qol', 'terus', 'tingkat', 'nyaman', 'player', 'best', 'lah']	0 + 4 + 0 + 0 – 4 + 0 + 0 + 4 + 0 + 0 + 0 + 4 + 0 + 0 + 0 = 8 → Positif
	['tambah', 'kamen', 'rider', 'zero', 'on', 'karakter']	0 + 0 + 0 + 0 + 0 + 0 = 0 → Netral
	['game', 'kikir', 'ngasih', 'bansos']	0 - 4 + 0 + 0 = - 4 → Negatif
InSet	['sehatt', 'game', 'baikk', 'makasii', 'udahh', 'wangiin', 'akun', 'aku', 'cintaa', 'deh', 'sama', 'game', 'ini']	0 + 2 + 0 + 0 + 0 + 0 + 0 - 1+ 0 + 0 + 3 + 2 + 0 = 6 → Positif
	['alamak', 'perangkat', 'kompetitable']	0 + 0 + 0 = 0 → Netral
	['gacha', 'nya', 'suka', 'zonk', 'mulu']	0 + 0 - 1 + 0 - 1 = 0 → Negatif

HASIL DAN PEMBAHASAN

Hasil

Data scraping dari laman aplikasi ZZZ di *Google Play Store*, yang berhasil mengumpulkan total 4.282 ulasan berbahasa Indonesia selama periode 4 Juli 2024 hingga 31 Maret 2025 disajikan pada tabel 2. Data yang dibutuhkan dari proses ini terdiri dari kolom *username* yaitu nama pengguna, *content* yaitu isi kata ulasan, dan *score* yaitu numerik *rating* dari ulasan.

Tabel 2. Hasil *data scraping* ulasan *game* zzz

No	***userName***	***content***	***score***
1	Aldhi Ridwan	🤝	5
2	Amerta Wijaya	aplikasinya bagus dan seru untuk di mainkan, pas gacha dapet di 80pull kurang jadi gak terlalu kikir kayak genshin, dan dapetin itemnya juga sangat mudah	5
...	...	...	...
4281	Dadan Firmani	Bisa bisanya game baru rilis sizenya lebih gede dari pada game open word pas rilis rilis 🤣🤣🤣 lawak ini game	1
4282	Reyvan J.A	Keren visual memanjangkan mata 💗	5

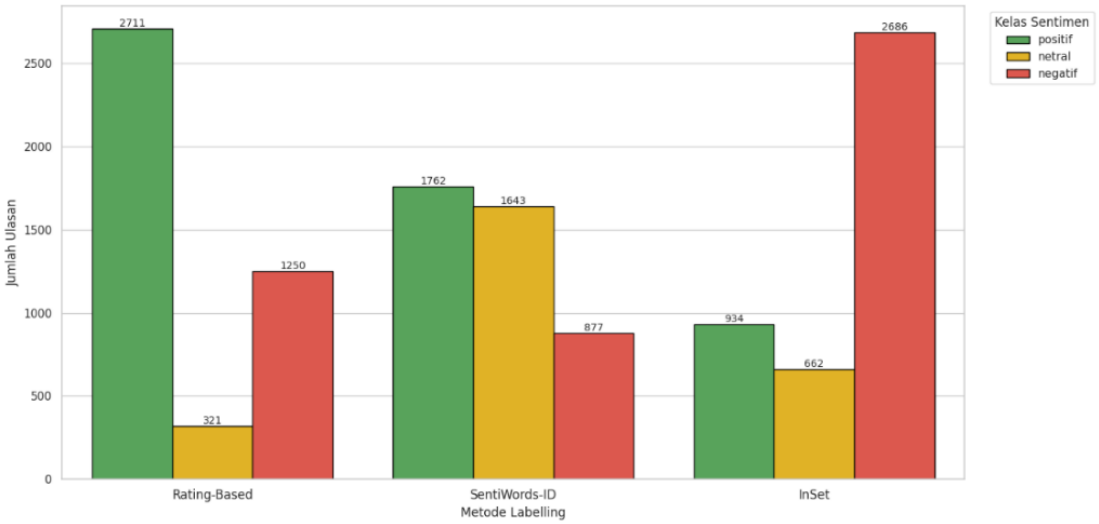
Data ulasan yang telah dikumpulkan kemudian melalui tahap Pra-pemrosesan Data yang terdiri dari *Null Handling*, *Cleaning*, *Case Folding, Tokenisation, Stopword Removal,* dan *Stemming*. Tahapan ini secara signifikan mereduksi kompleksitas data. Pada 48 ulasan dalam dataset menunjukkan terdapat karakteristik yang hanya terdiri dari simbol atau emoji tanpa disertai teks informatif. Tidak ditemukan adanya ulasan yang sepenuhnya kosong. Sebagai batasan dalam penelitian ini, tidak dilakukan proses eliminasi terhadap 48 ulasan tersebut. Penelitian ini juga tidak menerapkan filter manual untuk spam atau duplikasi, dengan asumsi bahwa platform Google Play Store telah memiliki mekanisme penyaringan internal untuk anomali semacam itu. Proses *stopword removal* berhasil mengurangi volume kata sebesar 19,78%, dari total 61.676 kata menjadi 49.475 kata. Secara keseluruhan, ukuran kosakata

(jumlah kata unik) dalam *dataset* berkurang dari 6.415 kata unik menjadi 5.344 kata unik setelah semua tahapan pra-pemrosesan.

Evaluasi kualitatif terhadap hasil *stemming* dari *library* Sastrawi menunjukkan efektivitasnya dalam mengubah kata berimbuhan menjadi kata dasar untuk kosakata bahasa Indonesia. Namun, teridentifikasi batasan dimana proses ini tidak dapat menangani kata dalam bahasa Inggris (misalnya, *'gameplay'*, *'crash'*) atau istilah/slang *game* (misalnya, *'gacha'*), sehingga kata-kata tersebut dipertahankan dalam bentuk aslinya. Tabel 3 menyajikan contoh hasil akhir dari keseluruhan tahap pra-pemrosesan, dengan menampilkan perbandingan antara teks ulasan orisinal pada kolom 'content' dan *token* kata-kata dasar pada kolom 'Stemming'.

Tabel 3. Hasil pra-pemrosesan data

No	*Content*	*Stemming*
1	🤝	[]
2	aplikasinya bagus dan seru untuk di mainkan, pas gacha dapet di 80pull kurang jadi gak terlalu kikir kayak genshin, dan dapetin itemnya juga sangat mudah	['aplikasi', 'bagus', 'seru', 'main', 'pas', 'gacha', 'dapet', 'pull', 'kurang', 'jadi', 'gak', 'terlalu', 'kikir', 'kayak', 'genshin', 'dapetin', 'item', 'sangat', 'mudah']
...	...	...
4281	Bisa bisanya game baru rilis sizenya lebih gede dari pada game open word pas rilis rilis 🤣🤣🤣 lawak ini game	['bisa', 'game', 'baru', 'rilis', 'sizenya', 'lebih', 'gede', 'game', 'open', 'word', 'pas', 'rilis', 'rilis', 'lawak', 'game']
4282	Keren visual memanjangkan mata 💗	['keren', 'visual', 'panjang', 'mata']



Gambar 1. Perbandingan distribusi label/kelas sentimen berdasarkan metode pelabelan

Tahap *data labelling* menghasilkan distribusi sentimen yang bervariasi tergantung pada metode yang digunakan. Perbandingan distribusi ini disajikan pada gambar 1. Pelabelan berbasis *rating* menghasilkan 2.711 data positif, 321 netral, dan 1.250 negatif; metode Sentiwords_ID menghasilkan 1.762 data positif, 1.643 netral, dan 877 negatif; sementara metode InSet menghasilkan 934 data positif, 662 netral, dan 2.686 negatif.

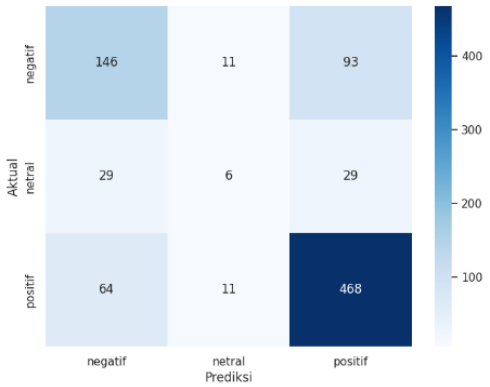
Setelah proses *data labelling*, data tersebut diproses menggunakan ekstraksi fitur TF-IDF, kemudian dibagi menjadi data latih (80%) dan data uji (20%) untuk persiapan klasifikasi model K-NN. Perbedaan signifikan pada distribusi label pada Gambar 1 merefleksikan bias setiap metode. Pelabelan berbasis *rating* cenderung menghasilkan label positif karena

aturannya yang mengonversi *rating* tinggi (bintang 4-5). Sebaliknya, Sentiwords_ID didominasi oleh label netral, kemungkinan akibat keterbatasan leksikonnya dalam mengenali istilah spesifik *game*. Sementara itu, InSet menghasilkan mayoritas label negatif, yang menunjukkan sensitivitasnya yang tinggi terhadap kata keluhan yang umum pada ulasan *game*.

Evaluasi model dengan pelabelan berbasis *rating* pada tabel 4 menunjukkan akurasi 72%. Kinerja terbaik dicapai pada kelas positif (*recall* 86%, presisi 79%), yang mengindikasikan efektivitas tinggi. Sebaliknya, performa pada kelas netral sangat lemah (*recall* 9%), yang utamanya disebabkan oleh ketidakseimbangan data yang signifikan (hanya 64 sampel). Sementara itu, kelas negatif menunjukkan *trade-off* yang moderat antara presisi (61%) dan *recall* (58%). Gambar 2 merinci bahwa 468 ulasan berhasil diprediksi dengan tepat sebagai positif atau *True Positive* (TP), namun terdapat juga kesalahan signifikan, seperti 64 ulasan positif yang keliru diklasifikasikan sebagai negatif atau *False Negative* (FN) dan 93 ulasan negatif yang salah diprediksi sebagai positif atau *False Positive* (FP).

Tabel 4. Laporan klasifikasi pelabelan berbasis *rating*

	Precision	***Recall***	***F1-Score***	***Support***
Negatif	61%	58%	60%	250
Netral	21%	9%	13%	64
Positif	79%	86%	83%	543
Accuracy		72%		857
Macro Average	54%	51%	52%	857
Weighted Average	70%	72%	71%	857



Gambar 2. *Confusion matrix* dari pelabelan *rating*

Tabel 5. Laporan klasifikasi pelabelan berbasis *sentiwords_id*

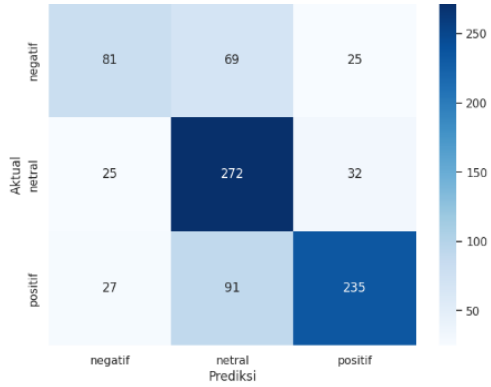
	Precision	***Recall***	***F1-Score***	***Support***
Negatif	61%	46%	53%	175
Netral	63%	83%	71%	329
Positif	80%	67%	73%	353
Accuracy		69%		857
Macro Average	68%	65%	66%	857
Weighted Average	70%	69%	68%	857

Sementara itu, pada pelabelan berbasis Sentiwords_ID yang tertera di tabel 5, akurasi yang dicapai adalah 69%. Berbeda dengan metode lain, model ini justru unggul dalam mengklasifikasikan kelas netral (*recall* 83%), didukung oleh distribusi data yang lebih seimbang. *Trade-off* jelas terlihat pada kelas positif (presisi tinggi 80%, namun *recall* lebih rendah 67%). Performa terlemah ada pada kelas negatif dengan *recall* 46%, yang menegaskan

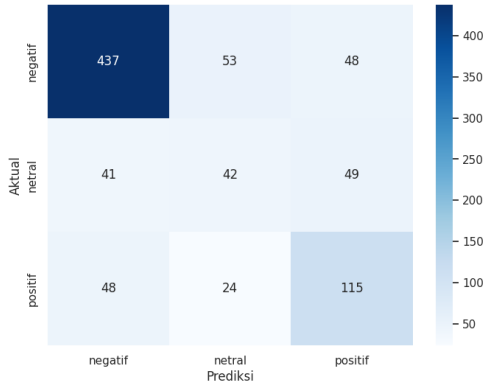
keterbatasan leksikon ini dalam mendeteksi istilah negatif spesifik *game*. Gambar 3 merinci performa model, dimana kelas netral menunjukkan True Positive (TP) atau prediksi tepat yang tinggi (272 ulasan), namun juga menarik banyak ulasan non-netral sebagai False Positive (FP), seperti 91 ulasan positif dan 69 ulasan negatif yang salah diklasifikasikan sebagai netral.

Tabel 6. Laporan klasifikasi pelabelan berbasis *inset*

	Precision	***Recall***	***F1-Score***	***Support***
Negatif	83%	81%	82%	538
Netral	35%	32%	33%	132
Positif	54%	61%	58%	187
Accuracy		69%		857
Macro Average	58%	58%	58%	857
Weighted Average	69%	69%	69%	857



Gambar 3. *Confusion Matrix* dari pelabelan kamus sentimen *sentiwords_id*



Gambar 4. *Confusion matrix* dari pelabelan kamus sentimen *inset*

Hasil dari pelabelan *InSet* pada Tabel 6 juga mencapai akurasi 69%. Model ini menunjukkan kekuatan utamanya pada kelas negatif, dengan presisi (83%) dan *recall* (81%) yang sangat tinggi dan seimbang. Seperti pada metode *rating*, kelas netral kembali menjadi yang terlemah (*recall* 32%) akibat jumlah data yang tidak seimbang. Kelas positif menunjukkan kinerja moderat dengan *recall* yang sedikit lebih tinggi (61%) daripada presisi (54%). Rincian pada *confusion matrix* pada gambar 4 menunjukkan kekuatan model dalam memprediksi ulasan negatif secara tepat (*true positive*) sebanyak 437 kali, namun juga menunjukkan adanya kesalahan klasifikasi, seperti 48 ulasan positif yang keliru diidentifikasi sebagai negatif *(false negative*).

Hasil evaluasi pada Tabel 7 merangkum perbandingan kinerja klasifikasi dari ketiga metode pelabelan berdasarkan metrik evaluasi utama. Berdasarkan tabel 7, metode pelabelan berbasis *rating* menunjukkan performa tertinggi di antara ketiganya, dengan mencapai akurasi 72%, presisi 70%, *recall* 72%, dan *F1-score* 71%. Metode InSet mengikuti dengan akurasi 69%, presisi 69%, *recall* 69%, dan *F1-score* 69%. Sementara itu, metode Sentiwords_ID mencatatkan hasil sebanding, dengan akurasi 69%, presisi 70%, *recall* 69%, dan *F1-score* 68%.

Tabel 7. Evaluasi kinerja klasifikasi berbasis metode pelabelan

Metode Pelabelan	***Precision***	***Recall***	***Accuracy***	***F1_Score***
Rating	70%	72%	72%	71%
Sentiwords_ID	70%	69%	69%	68%
InSet	69%	69%	69%	69%

Pembahasan

Hasil penelitian ini menunjukkan bahwa metode pelabelan berbasis *rating* memang menunjukkan performa akurasi tertinggi (72%). Namun, tingginya akurasi ini perlu ditinjau secara kritis, karena penggunaan *rating* sebagai 'kebenaran dasar' (*ground truth*) dalam konteks ulasan *game* memiliki keterbatasan signifikan. *Rating* dapat dipengaruhi oleh berbagai bias eksternal seperti loyalitas penggemar (*fan loyalty*), ekspektasi berlebih (*hype*), atau bahkan strategi ulasan yang tidak tulus.

Fenomena bias ini diakibatkan ketidakselarasan antara skor *rating* dan isi teks ulasan. Sebagai contoh, ditemukan kasus dimana pengguna memberikan *rating* bintang 5 namun menuliskan keluhan tajam seperti, "Power musuh terlalu sakit dan tombol Dodge yang sangat delay... gausah bikin game kalo in game terlalu hard buat f2p... Uninstall". Motivasi pengguna yang kompleks, ditambah dengan penggunaan sarkasme atau ironi yang tidak terdeteksi oleh TF-IDF, menyebabkan munculnya kasus *false negative*. Meskipun unggul secara metrik, keandalan metode berbasis *rating* sangat bergantung pada konsistensi antara skor numerik dan sentimen tekstual, yang sering kali tidak selaras dalam ulasan *game*.

Metode *leksikon Sentiwords_id* memiliki *recall* kelas negatif yang rendah yaitu 46%, menunjukkan adanya keterbatasan kamus dalam mengakomodasi istilah spesifik game. Leksikon Sentiwords_id juga gagal mendeteksi kata slang/istilah *game*. Sementara itu, metode InSet cenderung *over-predict* kelas negatif dengan skor presisi yaitu 83%, karena kamus InSet lebih banyak memuat kata dengan polaritas negatif kuat seperti "kecewa" atau "bug", tetapi kurang sensitif terhadap konteks netral seperti istilah teknis (misal: "update patch 1.6").

Rendahnya kinerja metode leksikon (keduanya akurasi 69%) menyoroti ketidaksesuaian antara kamus umum dan bahasa dinamis dalam ulasan *game*. Istilah spesifik seperti "gacha", "*nerf*", dan "*lag*" yang bernilai sentimen kuat bagi pemain, tidak terdaftar dalam leksikon standar. Akibatnya, esensi ulasan gagal ditangkap, yang menurunkan performa seperti pada *recall* negatif Sentiwords_ID yang hanya 46%.

Sulitnya mengklasifikasikan kelas netral (*recall* 9% pada metode *rating*) disebabkan oleh beberapa faktor. Ulasan netral sering kali tidak memiliki kata kunci sentimen yang jelas dan hanya berisi fakta atau pertanyaan, sehingga sulit diolah oleh metode berbasis frekuensi kata seperti TF-IDF. Selain itu, definisi 'netral' pada pelabelan otomatis tidak selalu konsisten, dan kecenderungan ulasan yang sangat positif atau negatif membuat kelas ini secara inheren kurang terwakili dalam data.

Arsitektur model juga memiliki keterbatasan. Representasi fitur TF-IDF yang berbasis 'kumpulan kata' gagal menangkap konteks semantik untuk mendeteksi sarkasme. Saat dipasangkan dengan KNN, model ini rentan terhadap 'kutukan dimensionalitas' akibat tingginya jumlah fitur. Algoritma seperti SVM atau naïve bayes bisa lebih *robust* untuk data teks. Ke depannya, eksplorasi *word embeddings* (seperti Word2Vec) atau *model transformator* (seperti BERT) direkomendasikan untuk pemahaman konteks yang lebih baik.

Meskipun berbagai penelitian sebelumnya telah menunjukkan efektivitas analisis sentimen pada produk atau aplikasi *google play store* (Abdillah et al., 2021; Artana et al., 2023; Firmansyah et al., 2024; Mustaqim et al., 2024; Pamungkas & Cahyono, 2024; Rahayu et al., 2022), belum ada penelitian yang secara spesifik membahas analisis sentimen pada *game* ZZZ ataupun melakukan evaluasi komparatif antara metode pelabelan berbasis *rating* dan kamus sentimen Bahasa Indonesia. Penelitian mereka umumnya berfokus pada penerapan satu model tunggal, mengandalkan pelabelan manual, atau menggunakan kamus sentimen Bahasa Inggris. Oleh karena itu, penelitian ini dapat membantu mengidentifikasi kepuasan pemain yang akan membantu para pengembang *game* ZZZ, sekaligus menyajikan adanya kebutuhan (*demand*) kamus sentimen *game-domain-specific* berbahasa Indonesia.

SIMPULAN

Penelitian ini menemukan bahwa metode pelabelan terbaik merupakan berbasis *rating*, yang meskipun mencapai akurasi tertinggi (72%), memiliki validitas yang problematik akibat ketidakselarasan antara skor dan teks ulasan. Sebaliknya, metode berbasis leksikon (*InSet* dan *SentiStrength_id* memiliki akurasi 69%) terhambat oleh keterbatasan cakupan kosakata untuk domain *game*. Temuan ini menunjukkan bagi para developer *game* bahwa untuk mendapatkan pemahaman yang komprehensif, diperlukan kombinasi metode: analisis berbasis *rating* untuk gambaran kepuasan umum secara cepat, dan analisis berbasis teks yang lebih mendalam untuk mengidentifikasi keluhan spesifik. Lebih lanjut, penelitian ini menegaskan bahwa pemilihan metode pelabelan dengan leksikon secara fundamental memengaruhi kinerja model. Oleh karena itu, langkah krusial berikutnya untuk meningkatkan akurasi analisis sentimen pada ulasan *game* berbahasa Indonesia adalah pengembangan leksikon domain-spesifik dan penerapan model berbasis konteks seperti *deep learning*.

REFERENSI

Abdillah, W. F., Premana, A., & Bhakti, R. M. H. (2021). Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, *3*(02), 160–170. https://doi.org/10.46772/intech.v3i02.556

Abodayeh, A., Hejazi, R., Najjar, W., Shihadeh, L., & Latif, R. (2023). Web Scraping for Data Analytics: A BeautifulSoup Implementation. *Proceedings - 2023 6th International Conference of Women in Data Science at Prince Sultan University, WiDS-PSU 2023*, 65–69. Riyadh, Saudi Arabia: IEEE. https://doi.org/10.1109/WiDS-PSU57071.2023.00025

Artana, I. K. A. B., Aditra, P. G., & Darmawiguna, I. G. M. (2023). Analisis Sentimen Twitter Untuk Menilai Kesiapan Pembelajaran Tatap Muka Terbatas Dengan Inset Lexicon Dan Levenshtein Distance. *Jurnal Pendidikan Teknologi dan Kejuruan*, *20*(2), 200–209. https://doi.org/10.23887/jptkundiksha.v20i2.64579

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780–2788. https://doi.org/10.11591/eei.v10i5.3157

Cahyaningtyas, C., Nataliani, Y., & Widiasari, I. R. (2021). Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE. *AITI: Jurnal Teknologi Informasi*, *18*(2), 173–184. https://doi.org/10.24246/aiti.v18i2.173-184

Firmansyah, Y., Kurniawan, R., & Wijaya, Y. A. (2024). Analisis Data Sentimen Pemain Game Role-Playing Game (RPG) Honkai Star Rail dengan Algoritma Naive Bayes. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, *6*(1), 127–135.

Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 1-55. https://doi.org/10.1186/s40537-024-00973-y

Khamket, T., & Polpinij, J. (2023). Automatically Correcting Noisy Labels for Improving Quality of Training Set in Domain-specific Sentiment Classification. *Current Applied Science and Technology*, *23*(2), 1–17. https://doi.org/10.55003/cast.2022.02.23.006

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications*, *13*(3), 144–168. https://doi.org/10.15849/ijasca.211128.11

Kusumastuti, R., Utami, E., & Yaqin, A. (2022). Detection of Sarcasm Sentences in Indonesian Tweets using SentiStrength. *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability,* 93–98.

Yogyakarta, Indonesia: IEEE. https://doi.org/10.1109/ICITISEE57756.2022.10057904

Magria, V., Asridayani, A., & Sari, R. W. (2021). Word Formation Process of Slang Word Used by Gamers In The Game Online "Mobile Legend." *Jurnal Ilmiah Langue and Parole*, *5*(1), 38–53. https://doi.org/10.36057/jilp.v5i1.497

Mahendra, M. H., Murdiansyah, D. T., & Lhaksmana, K. M. (2023). Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer. *DIKE : Jurnal Ilmu Multidisiplin*, *1*(2), 37–43. https://doi.org/10.69688/dike.v1i2.35

Mustaqim, K., Amaresti, F. A., & Dewi, I. N. (2024). Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor (KNN). *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 11–20. https://doi.org/10.29408/edumatic.v8i1.24779

Mustofa, Y. A., & Idris, I. S. K. (2024). Pendekatan Ensemble pada Analisis Sentimen Ulasan Aplikasi Google Play Store. *Jambura Journal of Electrical and Electronics Engineering*, *6*, 181–188. https://doi.org/10.37905/jjeee.v6i2.25184

Nur, A., Zulkifli, A., & Shafie, N. A. (2024). Review of the Lazada application on Google Play Store: Sentiment Analysis. *Journal of Computing Research and Innovation*, *9*(1), 43–55. https://doi.org/10.24191/jcrinn.v9i1.412

Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 1–10. https://doi.org/10.29408/edumatic.v8i1.24114

Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 98–106. https://doi.org/10.29408/edumatic.v6i1.5433

Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. *Expert Systems with Applications*, *181*. https://doi.org/10.1016/j.eswa.2021.115111

Saninah, A., Prihartono, W., Rohmat, C. L., & Cirebon, K. (2025). Analisis Sentimen Pengguna Terhadap Aplikasi Duolingo Dengan Naïve Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, *13*(1), 619–628. https://doi.org/10.23960/jitet.v13i1.5691

Ulya, S., Ridwan, A., Cholid Wahyudin, W., & Hana, F. M. (2022). Text Mining Sentimen Analisis Pengguna Aplikasi Marketplace Tokopedia Berdasar Rating dan Komentar Pada Google Play Store. *Jurnal Bisnis Digital dan Sistem Informasi*, *3*(2), 33–40.

Zhang, J. (2024). Catching the unlikely gambler : how and why gacha games appeal to conscientious consumers [Master's thesis, Lingnan University]. Lingnan University Digital Commons. https://commons.ln.edu.hk/otd/234/