

# Analisis Sentimen Fenomena “Brewek” Kartu Pokémon Pada Platform Reddit Menggunakan Arsitektur RoBERTa

Siti Fatimah Az Zahrah <sup>1</sup>, Klaudius Audie Irsansaputra<sup>2</sup> and Muhammad Rizky Pribadi <sup>3,\*</sup>

<sup>1,2,3</sup>Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang; sitifatimahazzahrah\_2327250055@mhs.mdp.ac.id; klaudiusaudieirsansaputra\_2327250018@mhs.mdp.ac.id; rizky@mdp.ac.id

\* Korespondensi : sitifatimahazzahrah\_2327250055@mhs.mdp.ac.id

## Info Artikel:

Dikirim: 10 Mei 2026

Direvisi: 17 Mei 2026

Diterima: 21 Mei 2026

**Abstract:** Social media platforms such as Reddit have long served as major forums for discussions among various communities. This study aims to analyze public sentiment to understand community trends and perceptions toward Pokémon TCG. The research applies the RoBERTa (Robustly Optimized BERT Approach) deep learning architecture using the pre-trained model “cardiffnlp/twitter-roberta-base-sentiment” to perform sentiment analysis. The text data were cleaned, tokenized with a maximum limit of 512 tokens, and classified into positive, neutral, and negative sentiments, followed by word length distribution analysis and Top-N Words extraction. The model successfully classified sentiments objectively. The visualization results demonstrate the characteristics of word distribution after outlier handling and identify the top ten keywords representing the main discussion focus within each sentiment category. The findings reveal that the community sentiment is predominantly negative, providing a clear overview of the opinion dynamics within the Pokémon community on Reddit.

**Keywords:** Sentiment Analysis; Information Retrieval; RoBERTa; Reddit; Pokémon Cards.

**Intisari:** Platform media sosial seperti Reddit sudah lama menjadi wadah utama diskusi berbagai komunitas. Penelitian ini bertujuan menganalisis sentimen publik untuk memahami tren dan persepsi komunitas terhadap Pokémon TCG. Penelitian ini menerapkan arsitektur *Deep Learning* RoBERTa (*Robustly Optimized BERT Approach*) melalui model *Pre-Trained* “cardiffnlp/twitter-roberta-base-sentiment” untuk melakukan analisis sentimen. Data teks dibersihkan, ditokenisasi dengan batas 512 token, dan diklasifikasikan ke dalam sentimen positif, netral, dan negatif, dilanjutkan dengan analisis distribusi panjang kata serta ekstraksi *Top-N Words*. Model berhasil mengklasifikasikan sentimen secara objektif. Hasil visualisasi menunjukkan karakteristik distribusi kata setelah penanganan *outliers* serta memetakan sepuluh kata kunci teratas yang merepresentasikan fokus pembahasan pada tiap label sentimen. Hasil penelitian ini menunjukkan bahwa sentimen yang terdapat dalam komunitas didominasi oleh sentimen negatif, memberikan gambaran jelas mengenai dinamika opini komunitas Pokémon di Reddit.

**Kata Kunci:** Analisis Sentimen; *Information Retrieval*; RoBERTa; Reddit; Kartu Pokémon.

## 1. Pendahuluan

Media sosial telah menjadi ruang utama bagi masyarakat untuk menyampaikan opini terhadap berbagai fenomena yang berkembang di lingkungan mereka. Reddit sebagai salah satu platform diskusi daring berbasis komunitas, memiliki sub-komunitas (*subreddit*) yang menjadi tempat khusus bagi pengguna untuk mendiskusikan topik-topik spesifik secara terbuka. Anonimitas yang diberikan Reddit mendorong penggunaannya untuk berbagi secara lebih jujur, menjadikannya sumber data opini yang kaya dan autentik [1]. Dengan menganalisis konten teks dari pos dan komentar

di Reddit, peneliti dapat mengidentifikasi tema-tema yang berulang dan menilai sentimen yang diungkapkan terhadap suatu isu serta diperoleh pemahaman yang lebih mendalam mengenai pandangan dari suatu komunitas [2].

Salah satu fenomena yang ramai diperbincangkan oleh komunitas kolektor di Reddit adalah "brewek" pada kartu Pokémon, yaitu istilah khas komunitas kolektor Indonesia untuk menggambarkan aktivitas membuka kemasan *booster pack* kartu Pokémon. Bagi komunitas kolektor kartu Pokémon, ini adalah sebuah momen yang membuat antusiasme namun juga memunculkan beragam reaksi emosional mulai dari kegembiraan hingga kekecewaan atas hasil yang didapatkan.

Besarnya volume diskusi komunitas global seputar fenomena ini di platform Reddit menghadirkan peluang untuk memahami kecenderungan opini publik secara sistematis melalui pendekatan analisis sentimen. Analisis sentimen merupakan teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, mengekstrak, dan mengkuantifikasi nada emosional dari suatu teks untuk memahami opini publik terhadap suatu topik tertentu [3]. Seiring pertumbuhan data pengguna di berbagai platform media sosial, analisis sentimen telah banyak diterapkan oleh para peneliti untuk mengklasifikasikan opini dari berbagai sumber data daring, termasuk diskusi komunitas di platform seperti Reddit [4].

Untuk menganalisis sentimen secara akurat dari data teks komunitas Reddit, penelitian ini menggunakan arsitektur RoBERTa (*Robustly Optimized BERT Pretraining Approach*). Pemilihan arsitektur ini didasarkan pada beberapa penelitian terdahulu yang telah berhasil membuktikan bahwa penerapan RoBERTa mampu menghasilkan akurasi yang baik. Keunggulan performa ini dicapai oleh optimasi pelatihan yang berhasil memperbaiki kelemahan *undertrained* pada arsitektur BERT. Hal ini dibuktikan pada penelitian Liu et al. (2019), yang memperkenalkan RoBERTa pertama kali sebagai pengembangan dari arsitektur BERT dan mengkaji dampak penyesuaian *hyperparameter* dan peningkatan data latih, dengan hasil bahwa RoBERTa mampu mencapai performa *state-of-the-art* pada *benchmark* pemahaman bahasa global seperti GLUE, RACE, dan SQuAD [5]. Efektivitas dari modifikasi tersebut telah divalidasi pada penelitian perbandingan komprehensif yang dilakukan Fahmi dan Nababan (2025). Penelitian ini mengkonfirmasi keunggulan ini secara empiris, di mana RoBERTa mencapai akurasi tertinggi sebesar 90.45% mengungguli arsitektur BERT (87.60%) [6].

Keunggulan performa ini juga terkonfirmasi dalam penelitian Zhang et al. (2024) yang mengkaji lebih spesifik penerapan RoBERTa pada tugas klasifikasi teks data Reddit untuk mengevaluasi performa pengenalan emosi menggunakan dataset ISEAR. Dalam pengujian tersebut, RoBERTa secara konsisten mengungguli model Transformer lainnya dengan mencatatkan akurasi tertinggi sebesar 0,7431, melampaui performa XLNet (0,7299), BERT (0,7009), dan DistilBERT (0,6693) [7]. Secara keseluruhan, capaian tersebut menegaskan bahwa pilihan desain eksperimental dan durasi pelatihan yang optimal merupakan variabel krusial dalam meningkatkan efikasi model bahasa terhadap tugas-tugas NLP yang kompleks.

Meskipun arsitektur RoBERTa terbukti efektif dalam menganalisis data teks Reddit, penerapannya pada domain diskusi yang spesifik seperti fenomena "brewek" kartu Pokémon masih belum ditemukan. Keterbatasan eksplorasi pada topik tersebut juga dipertegas dengan belum adanya representasi visual yang memadai untuk memetakan distribusi sentimen komunitas pada aktivitas ini.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen pengguna Reddit terhadap fenomena "brewek" kartu Pokémon menggunakan arsitektur RoBERTa. Penelitian difokuskan pada pengolahan data teks hasil *scraping* komentar Reddit untuk mengetahui kecenderungan sentimen positif maupun negatif yang divisualisasikan menggunakan *word cloud*. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi komunitas kolektor Pokémon terhadap fenomena "brewek" serta menunjukkan pemanfaatan model *transformer* dalam analisis sentimen pada media sosial berbasis komunitas.

## 2. Tinjauan Pustaka

Untuk memperkuat rancangan penelitian, bagian ini menelaah literatur pendukung dan kerangka konseptual yang relevan dari berbagai sumber pustaka terpercaya untuk memberikan justifikasi akademis terhadap metode dan metrik yang digunakan dalam memecahkan masalah.

### 2.1. Analisis Sentimen

Analisis sentimen merupakan bidang kajian dalam *Natural Language Processing* (NLP) yang secara khusus bertugas menentukan orientasi opini dari suatu teks ke dalam kategori polaritas seperti positif, negatif, atau netral, dengan memanfaatkan teknik linguistik komputasional dan pembelajaran mesin [8]. Dalam penelitian ini, analisis sentimen menjadi pendekatan yang tepat untuk memungkinkan klasifikasi otomatis terhadap ragam ekspresi emosional pengguna Reddit yang muncul sebagai respons atas hasil yang pengguna dapatkan terhadap fenomena "brewek" kartu Pokémon.

## 2.2. Reddit

Reddit merupakan salah satu platform sosial daring paling menonjol yang terstruktur dalam lebih dari 138.000 komunitas tematik aktif yang disebut *subreddit*, di mana konten dan komentar pengguna memiliki fitur *voting* guna untuk menentukan visibilitas, serta memberikan tingkat anonimitas yang tinggi karena partisipasi bersifat pseudonim tanpa verifikasi identitas nyata [9]. Karakteristik inilah yang menjadikan Reddit relevan sebagai sumber data dalam penelitian ini, dikarenakan komunitas kolektor kartu Pokémon berdiskusi secara terbuka dan anonim di *subreddit* tematik tertentu, sehingga komentar yang terkumpul merepresentasikan opini autentik yang dapat dianalisis sentimennya secara lebih mendalam.

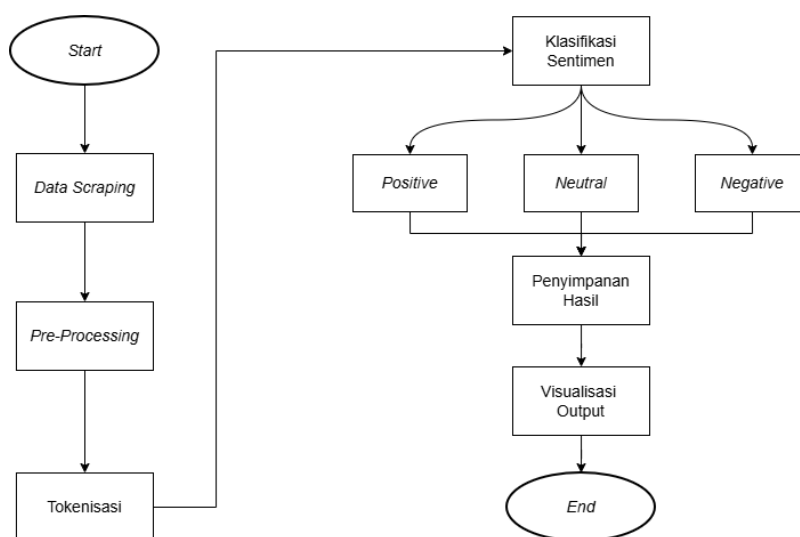
## 2.3. RoBERTa

RoBERTa adalah model bahasa berbasis Transformer yang dikembangkan oleh Liu et al. (2019) sebagai penyempurnaan dari BERT. Menurut Devlin et al. (2019), BERT (*Bidirectional Encoder Representations from Transformers*) adalah arsitektur *deep learning* yang merepresentasikan bahasa secara dua arah dengan mengondisikan setiap token secara bersamaan terhadap konteks kiri dan kanannya [10]. Liu et al. (2019) kemudian menyempurnakan BERT melalui pelatihan lebih lama dengan *batch* lebih besar, penghapusan objektif *Next Sentence Prediction*, serta penerapan *dynamic masking* yang membuat pola *masking* berubah setiap data diumpungkan ke model [5]. Arsitektur RoBERTa yang lebih adaptif terhadap teks informal menjadikannya pilihan yang tepat untuk menganalisis sentimen komentar komunitas Reddit seputar fenomena "brewek" kartu Pokémon.

## 3. Metode dan Pembahasan

Penelitian ini menggunakan model *Deep Learning* berbasis arsitektur *Transformer*, yaitu RoBERTa (*Robustly Optimized BERT Approach*). Secara spesifik, model yang digunakan adalah *cardiffnlp/twitter-roberta-base-sentiment*. Pemilihan model ini didasarkan pada keunggulannya dalam memahami konteks bahasa alami (NLP) pada media sosial secara dua arah (*bidirectional*). Model ini telah dilatih sebelumnya (*pre-trained*) menggunakan jutaan data teks informal, sehingga mampu mengenali pola bahasa, slang, dan nuansa emosi yang kompleks lebih baik daripada algoritma mesin konvensional. Selain itu, penelitian ini memanfaatkan akselerasi perangkat keras menggunakan teknologi CUDA melalui GPU untuk meningkatkan kecepatan komputasi selama proses inferensi model. Implementasi GPU *acceleration* memungkinkan pemrosesan ribuan komentar secara lebih efisien dibandingkan penggunaan CPU biasa.

Dalam penelitian ini, terdapat rancangan tahapan yang dimulai dengan *data crawling* dari Reddit untuk mengumpulkan dataset opini publik. Data mentah tersebut kemudian melalui tahap *pre-processing* dan tokenisasi agar lebih terstruktur dan dapat diproses oleh model. Selanjutnya, dilakukan klasifikasi menggunakan RoBERTa untuk memetakan teks ke dalam sentimen positif, netral, atau negatif. Tahap akhir meliputi penyimpanan dan visualisasi hasil analisis. Hasil rancangan tahapan ini dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

### 3.1. Data Crawling / Scraping

Tahap awal penelitian dimulai dengan akuisisi data dari platform Reddit melalui kombinasi teknik *data crawling* dan *web scraping*. *Web scraping* adalah proses pengumpulan data secara otomatis pada situs web melalui

tindakan program perangkat lunak atau bot [11]. Data yang diambil berfokus pada diskusi komunitas mengenai topik spesifik (misalnya: Pokemon) untuk mendapatkan basis data opini publik. Proses ini dilakukan dengan parameter pencarian tertentu untuk memastikan relevansi data. Atribut utama yang diekstraksi meliputi identitas pengguna (*username*) dan isi komentar atau postingan (*body*). Data mentah tersebut kemudian disimpan dalam format *Comma Separated Values* (.csv) untuk diproses lebih lanjut.

### 3.2. Pre-Processing (Pra-Pemrosesan Data)

*Pre-processing* merupakan tahapan krusial dalam NLP yang bertujuan untuk membersihkan teks mentah untuk mengekstraksi fitur-fitur penting dan meningkatkan relevansi kata sebelum diproses oleh model [12]. Data mentah yang diperoleh sering kali mengandung derau (*noise*) atau informasi yang tidak lengkap, sehingga diperlukan tahap pra-pemrosesan untuk meningkatkan kualitas *input* model. Langkah-langkah yang dilakukan meliputi:

- a. Integrasi Data : Menggabungkan kolom *username* dan *body* menjadi satu variabel *combined\_text*. Hal ini dilakukan untuk mempertahankan konteks identitas dalam analisis sentimen.
- b. Pembersihan Data : Menghapus baris yang mengandung nilai kosong (*null values*) atau *string* kosong agar tidak menyebabkan kesalahan pada saat proses komputasi.
- c. *Stopword Customization* : Melakukan proses penyaringan kata-kata umum yang terlalu dominan namun kurang memberikan makna analitis terhadap sentimen.

### 3.3. Tokenisasi dan Konfigurasi Input

Sebelum proses klasifikasi, data teks harus dikonversi ke dalam bentuk numerik yang dapat dipahami oleh model melalui proses tokenisasi. Tokenisasi merupakan tahapan pemrosesan teks yang bertujuan untuk memecah kalimat panjang menjadi unit-unit terkecil yang disebut token, untuk memfasilitasi ekstraksi fitur dan pengenalan struktur kalimat dalam analisis sentimen [13]. Pada tahap ini, teks dipecah menjadi unit-unit kecil (token) dan diberikan ID unik berdasarkan kosakata model RoBERTa. Konfigurasi *input* mencakup:

- a. *Truncation* : Pemotongan teks secara otomatis jika melebihi batas maksimal.
- b. *Max Length* : Batasan panjang input ditetapkan sebesar 512 token untuk menjaga efisiensi memori tanpa kehilangan substansi informasi yang signifikan.
- c. *Tensor Conversion* : Mengonversi daftar token menjadi bentuk Tensor agar dapat diproses secara paralel menggunakan GPU.

### 3.4. Klasifikasi Sentimen

Proses klasifikasi dilakukan dengan memasukkan token ke dalam model untuk diekstraksi fitur sentimennya. Model akan menghasilkan nilai *logits* untuk setiap kelas. Proses penentuan label akhir mengikuti alur berikut:

- a. *Softmax Function* : Mengonversi nilai *logits* menjadi nilai probabilitas di mana total distribusi peluang adalah 1.0.
- b. *Argmax Selection* : Memilih indeks dengan probabilitas tertinggi sebagai hasil akhir.
- c. *Label Mapping* : Memetakan indeks terpilih ke dalam tiga kategori utama: *Positive*, *Neutral*, dan *Negative*.

### 3.5. Penyimpanan dan Output Analisis

Hasil klasifikasi yang berupa label sentimen kemudian digabungkan kembali ke dalam *dataframe* asli sebagai kolom baru. Tahap akhir adalah melakukan serialisasi data dengan menyimpan seluruh hasil analisis ke dalam *file* baru berformat .csv. *Output* akan digunakan sebagai basis data untuk visualisasi hasil, seperti grafik distribusi sentimen, untuk menarik kesimpulan akhir dari penelitian.

## 4. Hasil Penelitian

Bagian ini memaparkan hasil implementasi dan evaluasi dari metode yang telah dirancang untuk menganalisis opini publik terkait fenomena "brewek". Pemaparan disusun secara sistematis mengikuti alur penelitian, dimulai dari proses *scraping* dataset, hasil *pre-processing* serta hasil tokenisasi model. Kemudian menguraikan hasil pengujian model

dalam klasifikasi sentimen dan visualisasi model data untuk menginterpretasikan kecenderungan opini terhadap fenomena "brewek" kartu Pokémon.

4.1. Tahap Data Crawling / Scraping

Implementasi tahap ini pada platform Reddit menghasilkan himpunan data mentah berjumlah 2.430 baris rekaman diskusi. Seluruh korpus data tersebut direkam berdasarkan periode aktivitas pengguna antara tanggal 19 Maret 2025 hingga 20 Februari 2026. Tampilan data yang telah direkam ini dapat dilihat pada Gambar 2.

	comment_id	createdAt	username	body
0	1	2025-03-19 14:05:07+00:00	jeffthrowonemore	Images:\n\thttps://preview.redd.it/j8t24xb7Inp...
1	2	2025-03-19 14:05:07+00:00	jeffthrowonemore	Images:\n\thttps://preview.redd.it/j8t24xb7Inp...
2	3	2025-03-19 14:09:34+00:00	CheeseGhosty	Guess what would happen if no-one paid over re...
3	4	2025-03-19 14:09:34+00:00	CheeseGhosty	Guess what would happen if no-one paid over re...
4	5	2025-03-19 14:11:35+00:00	Devh1989	post #1259757 on this topic
...	...	...	...	...
2425	2426	2026-02-05 14:20:50+00:00	Calm-Literature7502	Stay away from me. How are you going to clear ...
2426	2427	2026-02-07 03:01:37+00:00	homeDawgSliceDude	Has anyone tried to do the math on \$ packs ope...
2427	2428	2026-02-07 20:35:07+00:00	Suspicious_Lion_5515	jealous
2428	2429	2026-02-15 04:49:40+00:00	jakattakjak19945	I use Google drive and Google sheets. \n\nMake...
2429	2430	2026-02-20 19:35:37+00:00	Acceptable_Bet_4521	scalper???

Gambar 2. Data rekaman diskusi Reddit

Berdasarkan parameter ekstraksi yang ditetapkan, struktur data ini menggunakan kolom terdiri dari atribut *username* dan *body*. Hasil *crawling* sampel data mentah yang telah disimpan dengan format .csv dapat dilihat pada Gambar 3.

	username	body
0	Creative-Data5390	Did me and my girlfriend have an awesome time?...
1	notmushroom23	Everything I've opened since December 24. Roug...
2	thisismyname2129	**I opened 300 packs of 151 so you don't get t...
3	Rekanize90	i always think this too when i see this title ...
4	Alexstatic	Tbf, it is Japanese if that even counts
...	...	...
2425	continentaldrifting	I haven't seen a single pack of prismatic in t...
2426	Shifty_j10	Most average Facebook marketplace "selling" photo
2427	Cold94DFA	Nah it's not, the company could print more, an...
2428	UnderlightIII	It is. I want a Journey Together box for my bi...
2429	Panda_hat	Gotta respect the presentation though

Gambar 3. Hasil *crawling* dataset Reddit

4.2. Tahap Pre-Processing (Pra-Pemrosesan Data)

Tahap ini dimulai dengan proses menggabungkan atribut *username* dan *body* menjadi satu variabel untuk menjaga keterikatan konteks identitas dengan isi opini. Hasil representasi penggabungan dua atribut ini dapat dilihat pada Gambar 4.

	username	body	combined_text
0	Creative-Data5390	Did me and my girlfriend have an awesome time?...	Creative-Data5390Did me and my girlfriend have...
1	notmushroom23	Everything I've opened since December 24. Roug...	notmushroom23Everything I've opened since Dece...
2	thisismyname2129	**I opened 300 packs of 151 so you don't get t...	thisismyname2129**I opened 300 packs of 151 so...
3	Rekanize90	i always think this too when i see this title ...	Rekanize90i always think this too when i see t...
4	Alexstatic	Tbf, it is Japanese if that even counts	AlexstaticTbf, it is Japanese if that even counts
...	...	...	...
2425	continentaldrifting	I haven't seen a single pack of prismatic in t...	continentaldriftingI haven't seen a single pac...
2426	Shifty_j10	Most average Facebook marketplace "selling" photo	Shifty_j10Most average Facebook marketplace "s...
2427	Cold94DFA	Nah it's not, the company could print more, an...	Cold94DFANah it's not, the company could print...
2428	Underlightlll	It is. I want a Journey Together box for my bi...	UnderlightlllIt is. I want a Journey Together ...
2429	Panda_hat	Gotta respect the presentation though	Panda_hatGotta respect the presentation though

Gambar 4. Hasil integrasi data

Kemudian dilakukan proses pembersihan data dengan mengeliminasi baris yang tidak memiliki teks atau bernilai *null* untuk menjamin stabilitas komputasi pada fase klasifikasi. Hasil pembersihan menunjukkan jumlah baris yang tetap konsisten dengan data hasil *crawling*, yang membuktikan bahwa tidak ditemukan data bernilai *null* dalam dataset Reddit ini. Tampilan data yang telah melalui pembersihan data dapat dilihat pada Gambar 5.

	combined_text
0	Creative-Data5390Did me and my girlfriend have...
1	notmushroom23Everything I've opened since Dece...
2	thisismyname2129**I opened 300 packs of 151 so...
3	Rekanize90i always think this too when i see t...
4	AlexstaticTbf, it is Japanese if that even counts
...	...
2425	continentaldriftingI haven't seen a single pac...
2426	Shifty_j10Most average Facebook marketplace "s...
2427	Cold94DFANah it's not, the company could print...
2428	UnderlightlllIt is. I want a Journey Together ...
2429	Panda_hatGotta respect the presentation though

Gambar 5. Hasil pembersihan data

Proses terakhir dilakukan *stopword customization* yaitu penyaringan kata-kata umum dan istilah spesifik komunitas dilakukan secara dinamis untuk mempertajam relevansi kata terhadap nilai sentimen yang dihasilkan. Karena filter ini diterapkan secara langsung saat proses visualisasi, hasilnya tidak disajikan dalam bentuk tabel statis melainkan tercermin pada hasil analisis sentimen akhir.

#### 4.3. Tahap Tokenisasi dan Konfigurasi Input

Tahap ini mengonversi 2.430 baris data Reddit dari teks bahasa alami menjadi format numerik agar dapat dikenali oleh arsitektur RoBERTa. Proses dimulai dengan memecah kalimat pada kolom *combined\_text* menjadi token, lalu memetakannya ke dalam ID unik berdasarkan kosakata model. Melalui proses ini, data teks tidak lagi dipandang sebagai deretan karakter, melainkan sebagai vektor numerik yang merepresentasikan konteks sentimen, sehingga memungkinkan model untuk melakukan ekstraksi fitur secara mendalam terhadap pola komunikasi komunitas kolektor kartu Pokémon.

Untuk menjamin konsistensi data sebelum memasuki tahap klasifikasi, diterapkan konfigurasi input secara otomatis dengan pembatasan *input max length* sebesar 512 token. Komentar yang melebihi batas tersebut dipotong melalui mekanisme *truncation* untuk menjaga efisiensi penggunaan memori GPU. Dan seluruh kumpulan ID token dikonversi menjadi format *tensor* agar dapat diproses secara paralel menggunakan GPU. Hasil tokenisasi dan konfigurasi *input* ini dapat dilihat pada Gambar 6.

	combined_text	tokens	max_length	truncation	tensor_conversion
0	Creative-Data5390Did me and my girlfriend have...	[Creat, ive, -, Data, 5, 390, Did, Ġme, Ġand, ...	198 < 512 (✓)	-	torch.int64 on cuda
1	notmushroom23Everything I've opened since Dece...	[not, m, ush, room, 23, Everything, ĠI, âĢ, Ġ,...	23 < 512 (✓)	-	torch.int64 on cuda
2	thisismyname2129**I opened 300 packs of 151 so...	[this, ism, yn, ame, 2, 129, **, I, Ġopened, Ġ...	24 < 512 (✓)	-	torch.int64 on cuda
3	Rekanize90i always think this too when i see t...	[R, ek, an, ize, 90, i, Ġalways, Ġthink, Ġthis...	41 < 512 (✓)	-	torch.int64 on cuda
4	AlexstaticTbf, it is Japanese if that even counts	[Alex, static, T, bf, ,, Ġit, Ġis, ĠJapanese, ...	12 < 512 (✓)	-	torch.int64 on cuda
...	...	...	...	...	...
2425	continentaldriftingI haven't seen a single pac...	[cont, inent, ald, r, ifting, I, Ġhaven, âĢ, Ġ,...	24 < 512 (✓)	-	torch.int64 on cuda
2426	Shifty_j10Most average Facebook marketplace "s...	[Sh, ifty, _j, 10, Most, Ġaverage, ĠFacebook...	15 < 512 (✓)	-	torch.int64 on cuda
2427	Cold94DFANah it's not, the company could print...	[Cold, 94, DF, AN, ah, Ġit, 's, Ġnot, ,, Ġthe,...	48 < 512 (✓)	-	torch.int64 on cuda
2428	Underlightllllit is. I want a Journey Together ...	[Under, light, ll, ll, It, Ġis, ,, ĠI, Ġwant, Ġa,...	25 < 512 (✓)	-	torch.int64 on cuda
2429	Panda_hatGotta respect the presentation though	[P, anda, _hat, G, otta, Ġrespect, Ġthe, Ġpr...	10 < 512 (✓)	-	torch.int64 on cuda

Gambar 6. Hasil tokenisasi dan konfigurasi input

#### 4.4. Tahap Klasifikasi Sentimen

Data yang telah melewati proses tokenisasi dan konfigurasi input dimasukkan ke dalam arsitektur model RoBERTa untuk dilakukan ekstraksi fitur sentimennya. Proses klasifikasi bekerja dengan menghasilkan nilai logits yang kemudian dikonversi menjadi distribusi probabilitas melalui fungsi Softmax. Melalui mekanisme Argmax Selection, sistem secara otomatis menentukan indeks dengan nilai peluang tertinggi untuk dipetakan ke dalam tiga kategori utama, yaitu Positive, Neutral, dan Negative. Hasil akhir dari implementasi ini adalah pemberian label sentimen secara otomatis terhadap 2.430 baris data. Representasi dari hasil labelisasi data tersebut dapat dilihat pada Gambar 7.

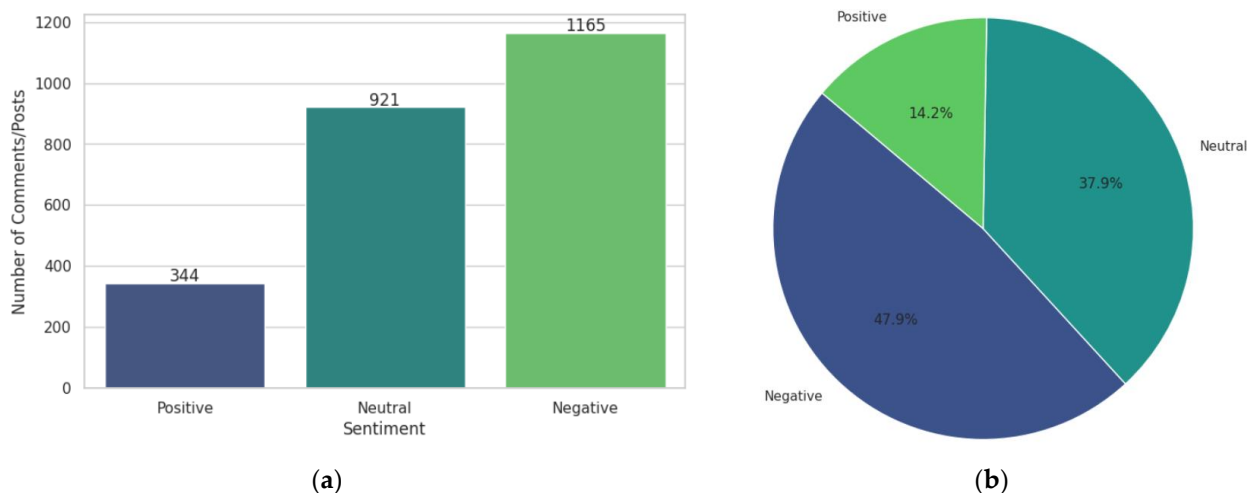
	combined_text	sentiment_label
0	Creative-Data5390Did me and my girlfriend have...	Positive
1	notmushroom23Everything I've opened since Dece...	Neutral
2	thisismyname2129**I opened 300 packs of 151 so...	Neutral
3	Rekanize90i always think this too when i see t...	Neutral
4	AlexstaticTbf, it is Japanese if that even counts	Neutral
...	...	...
2425	continentaldriftingI haven't seen a single pac...	Negative
2426	Shifty_j10Most average Facebook marketplace "s...	Neutral
2427	Cold94DFANah it's not, the company could print...	Neutral
2428	Underlightllllit is. I want a Journey Together ...	Positive
2429	Panda_hatGotta respect the presentation though	Neutral

Gambar 7. Hasil klasifikasi sentimen

#### 4.5. Penyimpanan dan Output Analisis

Tahap akhir dari alur penelitian ini adalah melakukan penyimpanan dataset yang telah klasifikasikan ke dalam format berkas .csv. Kemudian dilakukan ekstraksi informasi melalui berbagai teknik visualisasi untuk memetakan distribusi sentimen dan mendalami pola perilaku serta topik utama yang mendominasi korpus diskusi komunitas kolektor kartu Pokémon.

Hasil visualisasi terhadap distribusi sentimen pada fenomena "brewek" menunjukkan bahwa opini negatif mendominasi ruang diskusi secara mutlak. Sentimen negatif mencatat persentase tertinggi sebesar 47,9% (1.165 rekaman), diikuti oleh sentimen netral sebesar 37,9% (921 rekaman), sedangkan sentimen positif menjadi bagian terkecil dengan perolehan hanya 14,2% (344 rekaman). Tingginya angka polaritas negatif ini merefleksikan adanya keresahan massal dan kekecewaan sistemik di platform Reddit terkait fenomena "brewek" kartu Pokémon pada periode yang diamati dan dapat dilihat pada Gambar 8.



**Gambar 8.** Hasil distribusi klasifikasi sentimen fenomena “brewek” kartu Pokémon

(a) Visualisasi *bar chart* frekuensi jumlah komentar; (b) Visualisasi *pie chart* presentase proporsional distribusi

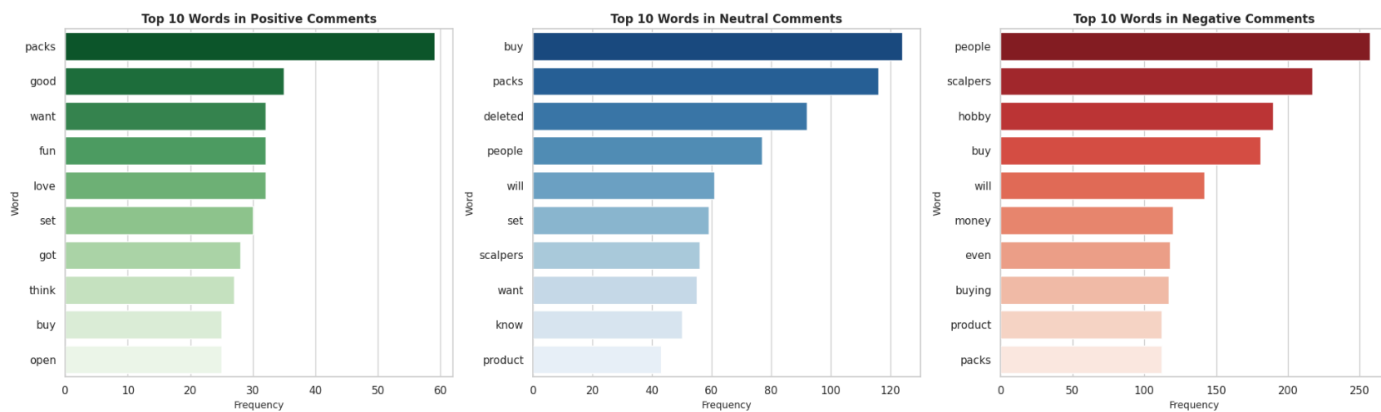
Melalui visualisasi pemetaan kata kunci spesifik kelas menggunakan *word cloud*, terlihat perbedaan fokus emosional yang kontras antara pengalaman positif dan negatif pengguna. Pada kluster positif, interaksi organik komunitas didominasi oleh aspek hobi yang menyenangkan lewat kemunculan kata seperti “pack”, “card”, “fun”, “love”, dan “good”. Sebaliknya, pada kluster negatif, narasi diskusi didominasi secara agresif oleh istilah-istilah kritis seperti “scalper”, “people”, “hobby”, dan “buying” yang mencerminkan kekecewaan mendalam terhadap manipulasi pasar seperti pada Gambar 9.



**Gambar 9.** Hasil perbandingan *word cloud* pengalaman komunitas terhadap fenomena “brewek” kartu Pokémon

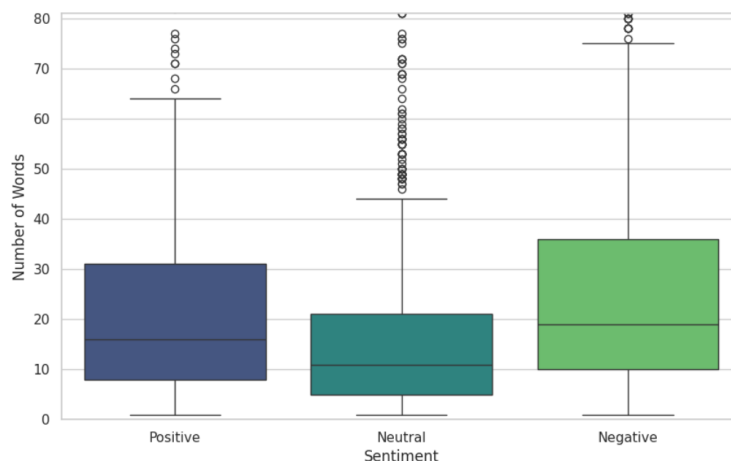
(a) Visualisasi *word cloud* pengalaman yang positif; (b) Visualisasi *word cloud* pengalaman yang negatif

Visualisasi terkait analisis frekuensi kata kunci dipertajam melalui grafik batang top 10 untuk memetakan istilah yang paling sering muncul di tiap-tiap klasifikasi sentimen secara presisi. Kluster positif berpusat pada kepuasan hobi dengan kata kunci tertinggi “packs”, sedangkan kluster netral didominasi oleh kosakata transaksional normatif dan logistik seperti “buy” dan “packs” serta tindakan moderasi teks berupa kata “deleted”. Sementara itu, kluster negatif menempatkan kata “people” dan “scalpers” di peringkat teratas sebagai indikator utama kemarahan komunitas terhadap perilaku fenomena “brewek”. Hasil visualisasi ini dapat dilihat pada Gambar 10.



Gambar 10. Visualisasi tren frekuensi 10 kata kunci tertinggi tiap klasifikasi sentimen

Selanjutnya dilakukan analisis visualisasi karakteristik struktural penulisan menggunakan *box plot* untuk melihat sebaran jumlah kata berdasarkan kategori sentimen. Grafik menunjukkan adanya lingkaran-lingkaran kecil (o) yang berada di atas garis *whisker* yang diidentifikasi secara statistik sebagai *outliers* melalui metode IQR (*Interquartile Range*) pada Seaborn. Kemunculan titik-titik ekstrem ini menandakan adanya beberapa komentar pengguna di Reddit yang menuliskan narasi atau argumen dengan *word count* yang sangat panjang dan ekspresif di luar rata-rata mayoritas data lainnya. Tampilan terkait analisis ini dapat dilihat pada Gambar 11.



Gambar 11. Visualisasi *box plot* distribusi *word count* tiap klasifikasi sentimen

Sebagai kesimpulan visualisasi, keseluruhan istilah dominan dari total 2.430 baris data diskusi digabungkan ke dalam satu pemetaan *word cloud*. Ukuran font yang sangat masif pada kata-kata kunci utama seperti "*scalper*", "*buy*", "*people*", "*hobby*", dan "*price*" memberikan bukti final bahwa dinamika komunikasi di platform Reddit telah bergeser dari sekadar pembahasan teknis permainan menjadi isu kelangkaan produk akibat spekulasi pasar sekunder dan dapat dilihat pada Gambar 12.



- [10] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019, doi: 10.18653/V1/N19-1423.
- [11] G. Machado, "Web Scraping: Data search and retrieval methodology in information science," *Academic Education Navigating the Path of Knowledge*, Dec. 2023, doi: 10.56238/SEVENED2023.008-002.
- [12] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, and Girinoto, "Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding," *AIP Conf. Proc.*, vol. 2680, no. 1, Dec. 2023, doi: 10.1063/5.0126116.
- [13] M. Aryanda, I. Arfyanti, and S. Widya Cipta Dharma, "Application Of K-Nearest Neighbor Algoritma for Customer Review Sentiment Analysis at Ngeboel Vapestore Shop," *J-INTECH*, vol. 13, no. 01, pp. 86–95, Jun. 2025, doi: 10.32664/J-INTECH.V13I01.1893.