

Analisis Sentimen Pemain terhadap Sistem Gacha Genshin Impact di Media Sosial X Menggunakan SVM

¹Muhammad Taufiqurrahman, ²Ali Alamsyah Kusumadinata, ³Hilmy Aliy Andra Putra

¹Program Studi Ilmu Komputer, Universitas Djuanda, Bogor

²Program Studi Ilmu Sosial dan Ilmu Politik, Universitas Djuanda, Bogor

³Program Studi Ilmu Komputer, Universitas Djuanda, Bogor

E-mail: ¹i.2211117@unida.ac.id, ²ali.alamsyah@unida.ac.id, ³hilmy.aliy@unida.ac.id

ABSTRAK

Sistem *gacha* pada game *Genshin Impact* memungkinkan pemain memperoleh karakter atau senjata langka secara acak, sehingga memicu respons emosional yang beragam di media sosial X. Penelitian ini bertujuan menganalisis kecenderungan sentimen pemain terhadap sistem *gacha* serta menguji kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen tersebut. Data penelitian berupa 690 *tweet* berbahasa Indonesia yang dikumpulkan melalui *scraping* menggunakan *library Tweet Harvest*, kemudian melalui tahap *preprocessing* (*cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*) dan diberi label sentimen positif, negatif, dan netral. Fitur teks diekstraksi menggunakan TF-IDF menghasilkan matriks berdimensi 690×3440, lalu diklasifikasikan menggunakan SVM kernel linear dengan parameter *class_weight='balanced'* dan pembagian data 80:20. Hasil pelabelan menunjukkan dominasi sentimen netral (65,94%), diikuti positif (18,41%) dan negatif (15,65%). Model SVM mencapai akurasi 74% pada data uji, namun evaluasi lanjutan menunjukkan indikasi *overfitting* yang signifikan, ditandai oleh kesenjangan *F1-Score Macro Average* antara data latih (99,35%) dan data uji (54,97%) sebesar 44,37%. Kelas netral menunjukkan performa terbaik (*Recall* 97%), sedangkan kelas negatif dan positif menunjukkan *Recall* rendah (23% dan 36%) akibat tumpang tindih kosakata informal dengan kelas netral. Temuan ini mengindikasikan bahwa performa SVM pada data *microblogging* berbahasa Indonesia yang informal lebih rentan terhadap *overfitting* dibandingkan pada data ulasan yang lebih terstruktur.

Kata kunci: Analisis Sentimen, Gacha, Genshin Impact, Support Vector Machine, TF-IDF

ABSTRACT

The *gacha* system in *Genshin Impact* allows players to randomly obtain rare characters or weapons, triggering diverse emotional responses on social media platform X. This study aims to analyze player sentiment trends toward the *gacha* system and evaluate the performance of the *Support Vector Machine* (SVM) algorithm in classifying such sentiment. The research data consisted of 690 Indonesian-language tweets collected through web scraping using the *Tweet Harvest* library, which were then processed through several preprocessing stages (*cleaning*, *case folding*, *tokenizing*, *stopword removal*, and *stemming*) and labeled into positive, negative, and neutral sentiment classes. Text features were extracted using TF-IDF, producing a 690×3440 weighted matrix, and classified using a linear-kernel SVM with a *class_weight='balanced'* parameter and an 80:20 data split. Labeling results showed a dominance of neutral sentiment (65.94%), followed by positive (18.41%) and negative (15.65%). The SVM model achieved 74% accuracy on the test data; however, further evaluation revealed significant overfitting, indicated by a 44.37% gap in Macro Average F1-Score between training data (99.35%) and testing data (54.97%). The neutral class exhibited the best performance (97% Recall), while the negative and positive classes showed low Recall (23% and 36%, respectively) due to vocabulary overlap between informal expressions and the neutral class. These findings suggest that SVM performance on informal Indonesian-language microblogging data is more susceptible to overfitting compared to more structured review data.

Keyword: Sentiment Analysis, Gacha, Genshin Impact, Support Vector Machine, TF-IDF

PENDAHULUAN

Industri *game* digital dalam beberapa tahun terakhir mencapai skala ekonomi yang besar, didorong oleh kemajuan teknologi dan penetrasi internet. Dalam model bisnis *game* modern, sistem *gacha* telah menjadi mekanisme monetisasi yang dominan, memungkinkan pemain memperoleh item, karakter, atau aset virtual secara acak melalui undian berbayar. Meskipun memberikan elemen kejutan dan *user engagement*, sistem ini menimbulkan kontroversi karena dianggap mirip praktik perjudian digital (Mursidah et al., 2025). Fenomena ini diperkirakan menyumbang pendapatan hingga miliaran dollar, menjadikannya topik mendesak untuk diteliti dari sudut pandang sentimen pemain (Anindya & Kaesmetan, 2025).

Genshin Impact, dikembangkan oleh HoYoverse, adalah salah satu *game* yang sukses menerapkan model ini melalui sistem *gacha* "Wish". Tingkat probabilitas rendah untuk memperoleh karakter *limited* atau senjata bintang lima memicu respons emosional pemain, mulai dari kepuasan hingga frustrasi. Respons tersebut tampak jelas dari unggahan pemain di platform X (Twitter), yang sifatnya *microblogging* memungkinkan penangkapan opini spontan dan dalam jumlah besar, menjadikannya sumber data primer yang efektif untuk analisis sentimen publik (Atma Jaya & Zeniarja, 2024).

Penelitian terdahulu oleh Muiz K et al. (2024) menerapkan *Support Vector Machine* (SVM) untuk menganalisis sentimen ulasan *game* Lokapala di Google Play Store dan mencapai akurasi 90%, menunjukkan efektivitas SVM dalam mengklasifikasikan data teks berdimensi tinggi seperti TF-IDF. Namun, penelitian tersebut menggunakan data ulasan terstruktur, berbeda dengan karakteristik bahasa informal dan dinamis pada data *microblogging* X. Penelitian ini bertujuan menjembatani gap tersebut dengan menguji kinerja SVM kernel linear pada data *tweet* yang lebih kompleks dan spesifik terhadap isu *gacha Genshin Impact*, sebuah domain yang lebih kontroversial dan belum banyak dikaji secara khusus.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk (1) menganalisis kecenderungan sentimen pemain terhadap sistem *gacha Genshin Impact* yang diungkapkan melalui media sosial X, dan (2) menguji kinerja metode SVM dalam mengklasifikasikan sentimen tersebut. Hasil penelitian diharapkan memberikan kontribusi teoritis dalam bidang *Natural Language Processing* (NLP) serta wawasan praktis mengenai persepsi pemain terhadap sistem *gacha*.

LANDASAN TEORI

Sistem Gacha dan Genshin Impact

Genshin Impact adalah *game open-world Action Role-Playing Game* (ARPG) dengan model *Game-as-a-Service* (GaaS) dan sistem *Free-to-Play* (F2P), yang memperoleh pendapatan melalui *microtransaction* dengan sistem *gacha* sebagai sumber pendapatan utama (Blom, 2023). Sistem *gacha* beroperasi pada mekanisme probabilitas yang memungkinkan pemain memperoleh karakter atau

item langka menggunakan mata uang virtual yang dibeli dengan uang nyata (Lakić et al., 2023), merupakan pengembangan dari *loot box system* (Montiel et al., 2022) yang terus memicu perdebatan etika karena kedekatannya dengan praktik perjudian.

Analisis Sentimen

Analisis sentimen atau *opinion mining* adalah cabang ilmu komputer yang menggunakan *Natural Language Processing* (NLP) untuk menentukan polaritas emosi (positif, negatif, atau netral) yang terkandung dalam teks (Arias et al., 2022). Pendekatan analisis sentimen berkembang dari metode *lexicon-based* menuju *machine learning* (*supervised learning*), hingga *deep learning* untuk konteks yang lebih kompleks. Meskipun demikian, metode *machine learning* klasik seperti SVM masih terbukti efisien untuk klasifikasi teks berdimensi tinggi (Anindya & Kaesmetan, 2025).

Text Preprocessing dan TF-IDF

Text preprocessing merupakan komponen fundamental dalam NLP yang mencakup *case folding*, *tokenizing*, *stopword removal*, dan *stemming* untuk menyiapkan teks sebagai masukan proses analisis lanjutan (Braga et al., 2025). *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk memberi bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen relatif terhadap keseluruhan korpus (Al-obaydy et al., 2022; Lestari & Hutagalung, 2025), dengan formulasi:

$$tf_i = n_i / N \quad (1)$$

$$idf_i = \log(D / D_i) \quad (2)$$

$$TF-IDF = tf_i \times idf_i \quad (3)$$

Support Vector Machine (SVM)

SVM merupakan algoritma *supervised learning* yang mencari *hyperplane* optimal untuk memisahkan data ke dalam dua atau lebih kelas dengan margin maksimal (Cortes & Vapnik, 1995; Novianti et al., 2024), dinyatakan sebagai:

$$f(x) = w^T x + b \quad (4)$$

dengan w sebagai vektor bobot, x sebagai vektor data masukan, dan b sebagai nilai bias. Penelitian ini menggunakan kernel linear, $K(x_i, x_j) = x_i^T x_j$, karena jenis kernel ini efisien dan paling banyak digunakan dalam pemrosesan teks berdimensi tinggi (Maulana et al., 2025; Kowalczyk, 2014). Performa SVM dipengaruhi oleh *hyperparameter* seperti parameter regularisasi C , yang mengontrol keseimbangan antara margin dan kesalahan klasifikasi; nilai C kecil menghasilkan margin lebih besar, sedangkan nilai C besar membuat model lebih ketat terhadap kesalahan. SVM juga memiliki keterbatasan dalam *tuning* parameter serta sensitivitas terhadap data tidak seimbang (Andriyani et al., 2025).

Metrik Evaluasi

Evaluasi kinerja model klasifikasi menggunakan *confusion matrix* yang terdiri atas *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang menjadi dasar perhitungan akurasi, presisi,

recall, dan *F1-Score* (Normawati & Prayogi, 2021; Indriyani et al., 2023).

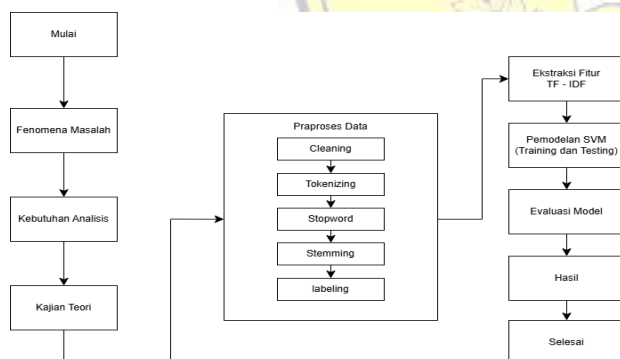
Penelitian Terkait

Beberapa penelitian terdahulu telah mengkaji penerapan SVM pada analisis sentimen *game*. Muiz K et al. (2024) menerapkan SVM pada ulasan *game* Lokapala di Google Play Store dengan akurasi 90%. Mursidah et al. (2025) menerapkan SVM pada ulasan *Honkai Star Rail* dengan akurasi 80%. Anindya dan Kaesmetan (2025) membandingkan BERT dan SVM pada data *Genshin Impact* di X, dengan BERT unggul (85,49%) dibandingkan SVM. Atma Jaya dan Zeniarja (2024) menguji TF-IDF dan SVM pada data *Genshin Impact* di X dan memperoleh akurasi 69% dengan bias *recall* antara kelas positif (0,87) dan negatif (0,56). Novianti et al. (2024) membandingkan SVM dan *Naïve Bayes* pada topik Metaverse, dengan SVM unggul (90,32%). Studi-studi ini menunjukkan bahwa performa SVM pada data *microblogging* cenderung lebih rendah dan lebih rentan bias dibandingkan pada data ulasan platform aplikasi yang lebih terstruktur — pola yang konsisten dengan temuan penelitian ini.

METODOLOGI PENELITIAN

Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan analisis sentimen berbasis *machine learning*. Tahapan penelitian meliputi pengumpulan data, praproses data, pelabelan sentimen, ekstraksi fitur menggunakan TF-IDF, pembagian dataset, pelatihan model SVM, pengujian, dan evaluasi performa, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan penelitian analisis sentimen

Dataset

Dataset penelitian berupa 690 *tweet* publik berbahasa Indonesia yang membahas pengalaman pemain terkait sistem *gacha Genshin Impact*, diperoleh melalui *scraping* dengan kata kunci relevan seperti “*gacha*”, “*wish*”, “*pull*”, “*Genshin Impact*”, “*banner*”, dan “*rate up*”. Data yang diambil merupakan *tweet* asli (bukan *retweet*) dan tidak memuat informasi pribadi seperti nama lengkap atau lokasi detail. Atribut metadata utama dataset ditunjukkan pada Tabel 1.

Tabel 1. Deskripsi atribut metadata dataset

Atribut	Tipe	Keterangan
full_text	String	Teks lengkap tweet
created_at	Datetime	Waktu unggah
lang	String	Kode bahasa

Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik *web scraping* otomatis dengan *library Tweet Harvest* berbasis *Python*, yang memanfaatkan otomasi peramban untuk mengunduh *tweet* publik berdasarkan *keyword* dan rentang waktu tertentu. Proses *scraping* dijalankan melalui lingkungan *Google Colab* dengan parameter bahasa Indonesia (*lang:id*) untuk memastikan relevansi data.

Praproses Data

Data mentah hasil *scraping* melalui lima tahap praproses secara berurutan. *Cleaning* menghapus tanda baca, angka, simbol, URL, tanda *retweet*, *hashtag*, dan *mention*. *Case folding* menyeragamkan seluruh huruf menjadi huruf kecil. *Tokenizing* memecah teks menjadi token kata individu. *Stopword removal* menghilangkan kata umum seperti “yang”, “dan”, “di” yang tidak memberikan bobot informatif. *Stemming* mengembalikan kata berimbuhan ke bentuk dasarnya menggunakan algoritma Sastrawi untuk bahasa Indonesia.

Pelabelan dan Ekstraksi Fitur

Setiap *tweet* diberi label sentimen (positif, negatif, netral) menggunakan pendekatan kamus kata kontekstual (*lexicon-based*) yang dioptimasi. Teks yang telah dilabeli selanjutnya diubah menjadi representasi numerik menggunakan TF-IDF sesuai persamaan (1)–(3), menghasilkan matriks berdimensi 690×3440 fitur kata unik.

Pemodelan SVM

Dataset dibagi menjadi data latih (80%, 552 dokumen) dan data uji (20%, 138 dokumen) menggunakan *stratified sampling* melalui fungsi *train_test_split*, guna menjaga proporsi sebaran kelas sentimen pada kedua subset. Model SVM dilatih menggunakan *library Scikit-Learn* dengan kernel linear, dipilih berdasarkan karakteristik data berdimensi tinggi (3.440 fitur) yang secara teoretis cenderung dapat dipisahkan secara linear dalam ruang vektor. Untuk mengatasi ketimpangan jumlah data antarkelas (kelas netral mendominasi), parameter *class_weight='balanced'* disertakan dalam konfigurasi model.

Evaluasi Model

Performa model dievaluasi menggunakan *confusion matrix* serta metrik akurasi, presisi, *recall*, dan *F1-Score* pada data uji. Selain itu, dilakukan analisis kelayakan model dengan membandingkan *F1-Score* pada data latih dan data uji untuk mendeteksi indikasi *underfitting* maupun *overfitting*.

HASIL DAN PEMBAHASAN

Hasil Praproses dan Pelabelan Data

Tahapan praproses berhasil mentransformasikan 690 *tweet* mentah yang bersifat informal dan tidak terstruktur menjadi data bersih melalui lima tahap berurutan: *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* menggunakan algoritma Sastrawi.

	full_text	step_1_cleaned
0	parah bjir mana currency buat gacha bukan dape...	parah bjir mana currency buat gacha bukan dape...
1	Guys izin nanyaa banner heartol itu bakal ada ...	guys izin nanyaa banner hearto itu bakal ada f...
2	@myufless log in 5 hari fix dapet 1600 jades (b...	log in hari fix dapet jades buka flip of for...
3	@pomptomfless block stat dapetin dadunya susah.....	block stat dapetin dadunya susah sisanya gacha....
4	hearto genshin hsr dll di kasih gratisan gni k...	hearto genshin hsr dll di kasih gratisan gni k...

Gambar 2. Hasil proses *cleaning*

	full_text	step_2_tokenized
0	parah bjir mana currency buat gacha bukan dape...	[parah, bjir, mana, currency, buat, gacha, buk...
1	Guys izin nanyaa banner heartol itu bakal ada ...	[guys, izin, nanyaa, banner, hearto, itu, baka...
2	@myufless log in 5 hari fix dapet 1600 jades (b...	[log, in, hari, fix, dapet, jades, buka, flip, of, f...
3	@pomptomfless block stat dapetin dadunya susah.....	[block, stat, dapetin, dadunya, susah, sisanya...
4	hearto genshin hsr dll di kasih gratisan gni k...	[hearto, genshin, hsr, dll, di, kasih, gratisa...

Gambar 3. Hasil proses *tokenizing*

	full_text	step_3_stopwords
0	parah bjir mana currency buat gacha bukan dape...	[parah, bjir, currency, gacha, dapet, daily, k...
1	Guys izin nanyaa banner heartol itu bakal ada ...	[guys, izin, nanyaa, banner, hearto, rerun, ka...
2	@myufless log in 5 hari fix dapet 1600 jades (b...	[log, in, fix, dapet, jades, buka, flip, of, f...
3	@pomptomfless block stat dapetin dadunya susah.....	[block, stat, dapetin, dadunya, susah, sisanya...
4	hearto genshin hsr dll di kasih gratisan gni k...	[hearto, genshin, hsr, dll, kasih, gratisan, g...

Gambar 4. Hasil proses *stopword removal*

	full_text	step_4_stemmed
0	parah bjir mana currency buat gacha bukan dape...	[parah, bjir, currency, gacha, dapet, daily, k...
1	Guys izin nanyaa banner heartol itu bakal ada ...	[guys, izin, nanyaa, banner, hearto, rerun, ka...
2	@myufless log in 5 hari fix dapet 1600 jades (b...	[log, in, fix, dapet, jades, buka, flip, of, f...
3	@pomptomfless block stat dapetin dadunya susah.....	[block, stat, dapetin, dadu, susah, sisa, gach...
4	hearto genshin hsr dll di kasih gratisan gni k...	[hearto, genshin, hsr, dll, kasih, gratis, gni...

Gambar 5. Hasil proses *stemming*

Hasil pelabelan sentimen terhadap keseluruhan dataset ditunjukkan pada Tabel 2.

Tabel 2. Hasil pelabelan sentimen

Kelas Sentimen	Jumlah Tweet	Persentase
Positif	127	18,41%
Negatif	108	15,65%
Netral	455	65,94%

Kelas netral mendominasi dataset dengan 455 *tweet* (65,94%), menunjukkan bahwa mayoritas pengguna X menggunakan istilah terkait *gacha Genshin Impact* untuk diskusi objektif, transaksi jual-beli akun, atau pertanyaan teknis seputar mekanika permainan tanpa muatan emosi yang kuat. Sentimen positif tercatat 18,41% (127 *tweet*), merepresentasikan kepuasan atau keberuntungan saat melakukan *pulling*, sedangkan sentimen negatif sebesar

15,65% (108 *tweet*) mencerminkan keluhan dan kekecewaan terhadap probabilitas sistem *gacha*.

Ekstraksi Fitur TF-IDF

Ekstraksi fitur terhadap 690 dokumen *tweet* menghasilkan matriks berdimensi 690×3440, menunjukkan 3.440 fitur kata unik teridentifikasi dari keseluruhan dataset.

```

--- HASIL EKSTRAKSI FITUR TF-IDF ---
Total Dokumen (Baris Data): 690
Total Fitur Kata Unik (Dimensi Kolom Matriks): 3440

=====
Menampilkan Sampel Pembobotan Nilai Kata pada Matriks TF-IDF:

```

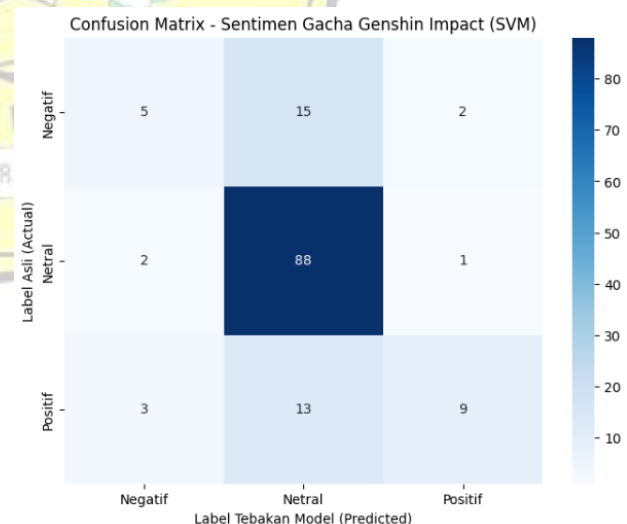
	aaaa	aaaaa	aaaaaaa	aaaaaaaa	amiin	aamon	abai	abang	abis	abistu
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 6. Hasil ekstraksi fitur TF-IDF

Sebagai ilustrasi, pada dokumen sampel d1: “parah bjir *currency gacha* dapet *daily genshin*”, kata “*currency*” yang hanya muncul pada 15 dari 690 dokumen (DF=15) memperoleh bobot TF-IDF sebesar 1,7975, jauh lebih tinggi dibandingkan kata “*genshin*” yang muncul pada 679 dari 690 dokumen (DF=679) dengan bobot 0,0070. Perbedaan ini menegaskan fungsi IDF dalam menurunkan bobot kata yang terlalu umum agar tidak mendominasi proses klasifikasi.

Performa Klasifikasi SVM

Model SVM kernel linear dilatih menggunakan 552 dokumen data latih dan diuji pada 138 dokumen data uji. Secara akumulatif, model berhasil menebak benar 102 dari 138 dokumen uji (5 negatif, 88 netral, 9 positif). Hasil pemetaan tebakan model dibandingkan dengan kunci jawaban asli disajikan pada Gambar 7.



Gambar 7. *Confusion matrix* hasil prediksi sentimen SVM

Hasil pengujian performa klasifikasi secara terperinci ditunjukkan pada Tabel 3.

Tabel 3. Hasil pengujian performa klasifikasi model SVM pada data uji

Kelas	Prec.	Recall	F1	N
Negatif	0.50	0.23	0.31	22
Netral	0.76	0.97	0.85	91
Positif	0.75	0.36	0.49	25

Kelas	<i>Prec.</i>	<i>Recall</i>	F1	N
Rata-rata	0.72	0.74	0.70	138
Akurasi	-	-	0.74	138

Model SVM menghasilkan akurasi global 74% dengan nilai *precision*, *recall*, dan *F1-Score* rata-rata masing-masing sebesar 0,72, 0,74, dan 0,70 pada evaluasi data uji. Kelas netral menunjukkan performa sangat baik dengan *recall* 97% dan *F1-Score* 0,85 — dari 91 data riil, 88 berhasil dikenali dengan tepat. Sebaliknya, kelas minoritas menunjukkan performa lemah: kelas negatif memiliki *recall* terendah (23%, hanya 5 dari 22 data terjaring) dan kelas positif memiliki *recall* 36% (9 dari 25 data terjaring). Kondisi ini mengindikasikan *boundary overlap*, di mana ekspresi kekecewaan maupun kegembiraan pemain sering dituliskan dengan kosakata bercampur slang atau singkatan tidak baku, sehingga pembobotan TF-IDF menghasilkan nilai fitur yang tipis jaraknya dengan kluster kata netral. Meskipun telah dibantu parameter *class_weight='balanced'*, model tetap cenderung mengklasifikasikan sebagian data positif dan negatif ke dalam kelas netral.

Evaluasi Kelayakan Model

Analisis lebih lanjut terhadap kelayakan model dilakukan dengan membandingkan *F1-Score* pada data latih dan data uji, sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan performa *F1-Score* model SVM pada data latih dan data uji

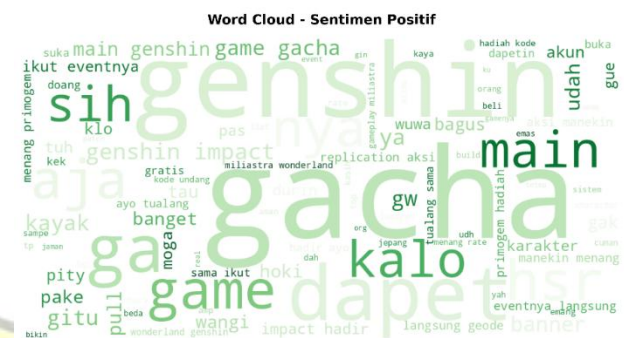
Metrik F1-Score	Training	Testing	Gap
Macro Average	99,35%	54,97%	44,37%
Weighted Average	99,46%	69,86%	29,60%

Tabel 4 menunjukkan kecenderungan *overfitting* yang signifikan pada model SVM. Performa model pada data latih hampir sempurna, dengan *F1-Score Macro Average* mencapai 99,35% dan *F1-Score Weighted Average* sebesar 99,46%. Namun, ketika diuji pada data uji, performa mengalami penurunan tajam menjadi 54,97% (*Macro Average*) dan 69,86% (*Weighted Average*), menghasilkan selisih (*gap*) *F1-Score Macro* sebesar 44,37% — jauh melampaui batas toleransi ideal selisih performa model yang umumnya berada pada rentang 10–15%.

Kondisi *overfitting* ini dipicu oleh kombinasi kompleksitas variasi kata, ukuran dataset yang terbatas, serta karakteristik pembobotan fitur teks berdimensi tinggi. Kondisi ini berbeda dengan penelitian acuan oleh Muiz et al. (2024) yang mengklasifikasikan 300 data ulasan Google Play Store bertema spesifik (*game* Lokapala) dengan akurasi 90%. Pada data ulasan media sosial X, variasi bahasa informal, *typo*, dan istilah slang yang jauh lebih kompleks menyebabkan algoritma SVM cenderung “menghafal” pola kata unik pada data latih, alih-alih mempelajari pola umum yang dapat digeneralisasi.

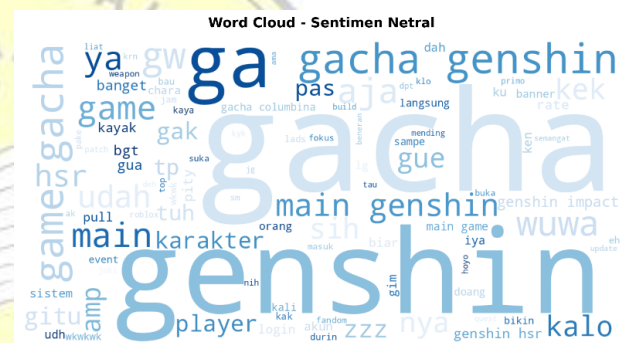
Analisis Kata Kunci Dominan

Visualisasi *word cloud* pada masing-masing kelas sentimen mengungkap pola linguistik yang berbeda. Pada sentimen positif, kata-kata dominan meliputi “hoki”, “wangi”, “bagus”, “menang *rate*”, dan “*primogem*”, merepresentasikan istilah khas komunitas *game* untuk menggambarkan keberuntungan tinggi saat memperoleh karakter atau item impian.



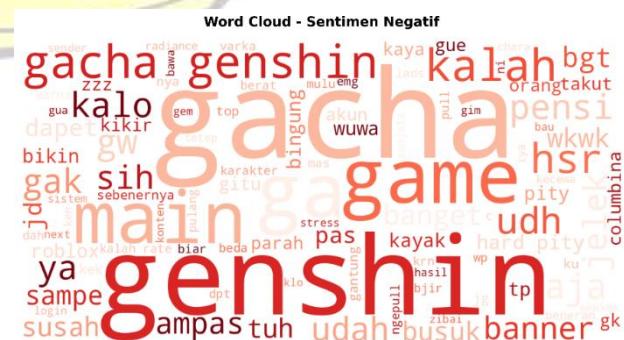
Gambar 8. Word cloud teks sentimen positif

Pada sentimen netral, kata dominan berkaitan dengan diskusi strategi tabungan *primogems* dan rekapitulasi *banner*.



Gambar 9. Word cloud teks sentimen netral

Pada sentimen negatif, istilah seperti “kalah”, “ampas”, “*hard pity*”, “*stress*”, dan “pensi” menandakan tingkat frustrasi dan kekecewaan mendalam terhadap probabilitas *qacha* yang dinilai sulit.



Gambar 10. Word cloud teks sentimen negatif

KESIMPULAN

Berdasarkan hasil penelitian, kecenderungan sentimen pemain *Genshin Impact* terhadap sistem *gacha* didominasi secara mayoritas oleh sentimen netral (65,94%), diikuti sentimen positif (18,41%) yang mencerminkan kepuasan dan keberuntungan saat *pulling*, serta sentimen negatif (15,65%) yang menggambarkan

kekecewaan terhadap kegagalan mekanika *gacha*. Pengujian kinerja model SVM dalam mengklasifikasikan sentimen pemain menghasilkan akurasi 74% pada data uji dengan *F1-Score Macro Average* sebesar 54,97% dan *F1-Score Weighted Average* sebesar 69,86%. Evaluasi kelayakan model menunjukkan indikasi *overfitting* yang signifikan, ditandai oleh *F1-Score Macro Average* pada data latih yang mencapai 99,35% — menghasilkan selisih performa sebesar 44,37% terhadap data uji. Kondisi ini dipicu oleh kompleksitas kosakata informal dan istilah slang *gaming* pada dataset berdimensi tinggi yang berukuran relatif terbatas. Untuk penelitian selanjutnya, disarankan penambahan volume sampel data, penyetelan *hyperparameter* C secara lebih intensif melalui *GridSearchCV*, penerapan teknik penanganan data tidak seimbang seperti SMOTE, serta perluasan kamus normalisasi kata slang guna meningkatkan kemampuan generalisasi model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Djuanda Bogor, serta dosen pembimbing atas bimbingan dan arahan selama proses penelitian ini.

DAFTAR PUSTAKA

- Al-obaydy, W. N. I., Hashim, H. A., Najm, Y. A., & Jalal, A. A. (2022). Document classification using term frequency-inverse document frequency and K-means clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(3), 1517–1524. <https://doi.org/10.11591/ijeecs.v27.i3.pp1517-1524>
- Andriyani, D., Faqih, A., & Permana, S. E. (2025). The Effect of SMOTE Application on Support Vector Machine Performance in Sentiment Classification on Imbalanced Datasets. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 752–757.
- Anindya, F. S., & Kaesmetan, Y. R. (2025). Implementasi Metode Bert Dan Svm Pada Analisis Sentimen Game Genshin Impact. *JMI: Jurnal Manajemen Informatika* Jayakarta, 5(1), 52–60.
- Arias, F., Zambrano Núñez, M., Guerra-Adames, A., Tejedor-Flores, N., & Vargas-Lombardo, M. (2022). Sentiment Analysis of Public Social Media as a Tool for Health-Related Topics. *IEEE Access*, 10, 74850–74872.
- Atma Jaya, S. I., & Zeniarja, J. (2024). Sentiment Analysis of Genshin Impact on X: Mental Health Implications Using TF-IDF and Support Vector Machine. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 8(3), 1589–1599.
- Blom, J. (2023). The Genshin Impact Media Mix: Free-to-Play Monetization from East Asia. *Mechademia*, 16(1), 144–166.
- Braga, M., Milanese, G. C., & Pasi, G. (2025). Investigating Large Language Models' Linguistic Abilities for Text Preprocessing. Preprint (ArXiv). <https://doi.org/10.48550/arXiv.2510.11482>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297.
- Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, 10(2), 176–184.
- Kowalczyk, A. (2014). Linear Kernel: Why is it recommended for text classification? SVM Tutorial.
- Lakić, N., Bernik, A., & Čep, A. (2023). Addiction and Spending in Gacha Games. *Information*, 14(7), 399.
- Lestari, V. B., & Hutagalung, C. A. (2025). Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. *J-KOMA: Jurnal Ilmu Komputer dan Aplikasi*, 8(1), 36–44.
- Maulana, N., Handayani, M. R., Yuniarti, W. D., & Umam, K. (2025). Penerapan Metode Support Vector Machine Dalam Mengklasifikasi Berita Politik: Fakta vs Hoaks. *JUPI*, 10(3), 2679–2688.
- Montiel, I., Basterra-Gonzalez, A., Machimbarrena, J. M., Ortega-Baron, J., & Gonzalez-Cabrera, J. (2022). Loot box engagement: A Scoping Review of Primary Studies on Prevalence and Association with Problematic Gaming and Gambling. *PLOS One*, 17(1), 1–23.
- Muiz K, R., Ishar, R. F., Febrianto, A., & Akbar, M. N. (2024). Analisis Sentimen Komentar Pengguna Terhadap Game Moba Lokapala Di Google Play Store Menggunakan Algoritma Support Vector Machine. *AGENTS: Journal of Artificial Intelligence and Data Science*, 4(2), 1–9.
- Mursidah, I., Sanjaya, R., Yulianto, B., Sweetania, D., & Sularsih, P. (2025). Analisis Sentimen Pengguna pada Ulasan Game Honkai Star Rail Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Minfo Polgan*, 14(1), 1178–1188.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Novianti, D. N., Shiddieq, D. F., Roji, F. F., & Susilwati, W. (2024). Comparison of Support Vector Machine and Naïve Bayes Algorithms for Sentiment Analysis of the Metaverse. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 231–239.