



ANALISIS SENTIMEN TERHADAP GAME CLASH OF CLANS BERDASARKAN ULASAN PEMAIN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Satria Bayu Agustian¹, Ahmad Dani Tengku Pasyah², Lahenda Vinaro³,
Rame Santoso⁴, Indah Purwandani⁵

¹19210622@bsi.ac.id, ²19210616@bsi.ac.id, ³19210635@bsi.ac.id,

⁴rame.rms@bsi.ac.id, ⁵indah@bsi.ac.id

^{1,2,3,4,5}Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

Abstrak

Popularitas *game Clash of Clans* menghasilkan volume ulasan yang besar pada platform *Google Play Store* hingga saat ini. Studi ini mengevaluasi opini pemain menggunakan pendekatan *Support Vector Machine* (SVM) terhadap 3.287 data ulasan yang dihimpun pada periode April-Mei 2025. Serangkaian tahapan *preprocessing* diterapkan, mulai dari pembersihan data hingga *stemming*. Selanjutnya, ulasan dikategorikan ke dalam label sentimen positif dan negatif. Data tersebut kemudian diproses melalui pembobotan teks TF-IDF untuk selanjutnya diklasifikasikan menggunakan algoritma SVM. Hasil pengujian menunjukkan dominasi sentimen positif dengan tingkat akurasi mencapai 89%. Temuan ini memberikan wawasan bagi pengembang dalam memetakan preferensi serta aspirasi pemain, sekaligus mengonfirmasi keandalan teknik *machine learning* untuk analisis sentimen yang presisi.

Kata kunci: Analisis Sentimen, *Clash of Clans*, TF-IDF, *Support Vector Machine*.

Abstract

The sustained popularity of the *game Clash of Clans* has generated a substantial volume of user reviews on the *Google Play Store*. This study evaluates player opinions using a *Support Vector Machine* (SVM) approach on 3,287 reviews collected between April and May 2025. A series of *preprocessing* stages, ranging from data cleaning to *stemming*, were implemented. Subsequently, reviews were categorized into positive and negative sentiment labels. The data was then processed through TF-IDF text weighting to be further classified using the SVM algorithm. The results indicate a predominance of positive sentiment, with the model achieving an accuracy rate of 89%. These findings provide valuable insights for developers to map player preferences and feedback, while confirming the reliability of machine learning techniques for precise sentiment analysis.

Keywords: Sentiment Analysis, *Clash of Clans*, TF-IDF, *Support Vector Machine*.

1. Pendahuluan

Perkembangan dalam teknologi informasi dan komunikasi berperan besar dalam mempercepat pertumbuhan industri *game*. Peran *game* saat ini telah berkembang dari hiburan semata menjadi media edukasi, sarana simulasi, bahkan bagian dari *e-sport*. Dalam bahasa Indonesia, istilah *game* berarti permainan. Namun seiring kemajuan teknologi, istilah *game* juga mencakup suatu bentuk pemrograman permainan pada perangkat dan memiliki fleksibilitas akses, baik melalui jaringan internet maupun tanpa koneksi [1]. Menurut Azkia, *Game* adalah permainan yang dapat mengisi waktu luang berbagai usia dan menjadi sarana mengekspresikan perasaan, bahkan hingga membuat pemain lupa waktu [2].

Salah satu *game* strategi populer adalah *Clash of Clans* yang dikembangkan oleh *supercell* dari Finlandia. Diluncurkan pertama kali di iOS pada 2 Agustus 2012 dan di Android pada 7 Oktober 2013, *game* ini dibuat oleh Lasse Louhento dan Lassi Leppinen untuk menghadirkan pengalaman strategi yang sederhana [3]. *Game* ini dapat diakses di Android, iOS, dan Windows, dan pada tahun 2025 telah diunduh lebih dari 500 juta kali dengan *rating* di atas 7 serta mendapatkan lebih dari 61,9 juta ulasan. Meskipun popularitasnya tinggi, penelitian yang menganalisis persepsi pemain secara mendalam masih terbatas. Hal ini menunjukkan bahwa meskipun *game Clash of Clans* sangat populer, masih sedikit penelitian yang membahas bagaimana pemain merespons dan merasakan pengalaman bermainnya. Pemahaman ini dapat membantu pengembangan *game* strategi di masa mendatang.

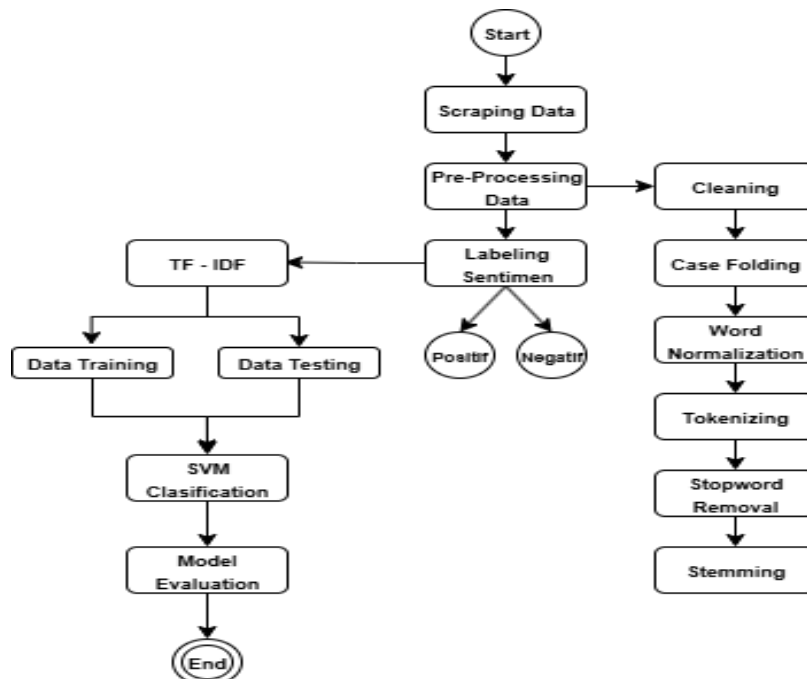
Ulasan pemain yang diperoleh melalui *Google Play Store* dapat dijadikan referensi penting bagi pengembang untuk melakukan peningkatan kualitas *game*. Dengan demikian, analisis sentimen diperlukan agar persepsi pemain dapat dipahami secara keseluruhan. Metode *Support Vector Machine* dapat diterapkan karena memiliki kemampuan yang efektif dalam mengklasifikasikan data teks. Menurut Tanggraeni, Dalam membagi kategori opini menjadi positif dan negatif, SVM bekerja sebagai teknik klasifikasi berbasis pembelajaran mesin yang mengandalkan penggunaan *hyperplane* sebagai garis pemisah [4]. Menurut Hartmann, SVM unggul dalam memproses data non-linear serta dataset berdimensi besar [5].

Beberapa studi terdahulu membuktikan efektivitas SVM efektif dalam analisis sentimen pada berbagai bidang, seperti *game Mobile Legends* [6], *game Lokapala* [7], aplikasi *e-commerce* seperti Shopee [8], media sosial seperti TikTok [9], dan aplikasi Jamsostek *Mobile* [10]. Namun, penelitian khusus untuk *game* strategi seperti *Clash of Clans* masih jarang dilakukan. Hal ini menjadi celah yang akan diisi dalam penelitian ini. Karena ulasan pemain di *Google Play Store* dapat memberikan gambaran mendalam mengenai testimoni dan persepsi pemain *game Clash of Clans*.

Penelitian ini memiliki tujuan untuk mengevaluasi sentimen pemain terhadap *game Clash of Clans* melalui ulasan yang ada di *Google Play Store*. Metode SVM diterapkan untuk membagi opini para pemain ke dalam kelompok sentimen positif dan negatif, serta menilai tingkat keakuratan SVM dalam memproses dan menangani data ulasan tersebut. Studi ini diproyeksikan mampu memperdalam wawasan terkait persepsi para pemain serta menilai efektivitas SVM dalam analisis sentimen pada *game* strategi.

2. Metode

Metode penelitian ini menggunakan alur sistematis untuk menganalisis sentimen pemain *game Clash of Clans*. Tahapan dimulai dari *scraping* data, lalu *preprocessing* yang mencakup *cleaning*, *case folding*, *word normalization*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya, dilakukan pembobotan TF-IDF serta pelabelan sentimen. Data kemudian dibagi dengan rasio 80% (data latih) dan 20% (data uji) untuk proses klasifikasi menggunakan SVM. Evaluasi performa model diukur melalui *accuracy*, *precision*, *recall*, dan *f1-score* berdasarkan *confusion matrix*. Alur penelitian ini secara ringkas disajikan pada Gambar 1:



Gambar 1. Kerangka Penelitian

2.1 Desain Penelitian

Penelitian ini berfokus pada pemanfaatan data *mining* sebagai upaya mengekstraksi informasi berharga dari kumpulan data berukuran besar dengan memanfaatkan metode matematis, statistik, dan kecerdasan buatan [11]. Metode ini digunakan untuk menggali pola data sebagai dasar pengambilan keputusan [12].

Dalam penelitian ini, data *mining* digunakan untuk melakukan analisis sentimen, yaitu proses mengidentifikasi opini atau penilaian pengguna terhadap suatu entitas, baik bersifat positif maupun negatif [13]. Menurut Mahesh, Dalam *machine learning*, analisis sentimen diklasifikasikan sebagai cabang dari NLP atau *text mining*, yang memanfaatkan algoritma serta model statistik untuk membantu sistem komputer mengelola data dalam jumlah besar secara lebih efektif [14].

Pemanfaatan ulasan di *Google Play Store* melalui analisis sentimen bertujuan untuk memetakan sudut pandang pemain *game Clash of Clans*. Dengan pendekatan ini, opini para pemain dapat diolah menjadi informasi yang lebih terstruktur sehingga dapat bermanfaat dalam memahami kecenderungan sentimen pemain terhadap *game* tersebut.

2.2 Scraping Data

Menurut Ernianti, *Scraping* data merupakan metode otomatis untuk mengumpulkan informasi dari sumber tertentu secara terstruktur [15]. Proses *scraping* data ulasan dilakukan di lingkungan *Google Colab* menggunakan *Python* dengan memanfaatkan pustaka *google-play-scraper*. Data yang dikumpulkan meliputi ulasan serta penilaian pengguna terhadap *game Clash of Clans*. Selanjutnya, data tersebut disimpan dalam format CSV atau Excel untuk diproses pada tahap *preprocessing*.

2.3 Preprocessing Data

Dalam pengolahan teks, *preprocessing* menjadi salah satu tahap yang sangat penting guna mentransformasi data kasar ke dalam bentuk yang lebih sistematis. Proses ini membuat data lebih bersih dan konsisten sehingga meningkatkan akurasi analisis sentimen. Tahapan *preprocessing* meliputi *cleaning*, *case folding*, *word normalization*, *tokenizing*, *stopword removal*, dan *stemming* [16].

1. *Cleaning*: Tahap eliminasi elemen non-teks seperti angka, simbol, serta karakter unik yang tidak relevan.
2. *Case Folding*: Penyeragaman seluruh karakter menjadi huruf kecil.
3. *Word Normalization*: Mengonversi istilah non formal menjadi kosakata bahasa Indonesia yang baku.
4. *Tokenizing*: Pemotongan rangkaian kalimat menjadi fragmen kata atau unit tunggal.
5. *Stopword Removal*: Pembersihan kata-kata fungsional yang minim makna misalnya "di", "dan", "untuk" guna mempercepat komputasi.
6. *Stemming*: Proses reduksi kata berimbuhan kembali ke bentuk akar atau kata dasarnya.

2.4 Labeling Sentimen

Penentuan kelas sentimen mencakup kategori positif, negatif, dan netral. Meskipun pelabelan manual memakan waktu, cara ini menghasilkan data yang lebih presisi atau secara otomatis dengan library seperti *TextBlob* yang lebih efisien tetapi kurang presisi [17]. Pada penelitian ini, *labeling* digunakan untuk memudahkan klasifikasi dan analisis sentimen ulasan pemain *game Clash of Clans*.

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan metode yang mengubah teks menjadi data numerik dengan memberi bobot pada setiap kata sesuai tingkat kepentingannya dalam dokumen [18]. Metode ini membantu membedakan kata umum dan kata bermakna khusus, sehingga meningkatkan akurasi analisis sentimen dengan algoritma seperti SVM. Data kemudian dibagi menjadi data latih 80% dan data uji 20% untuk melatih serta mengevaluasi model [19]. Adapun rumus dari metode TF-IDF sebagai berikut [20].

$$TF(t) = \frac{\text{Jumlah Kata (t) dalam komentar}}{\text{Jumlah total kata dalam komentar}}$$

$$IDF = \frac{1 + \text{Total Dokumen Frekuensi (N)}}{1 + \text{Jumlah Total Kata (DF)}} + 1$$

$$TF-IDF = TF \times IDF$$

Keterangan:

- TF : *Term Frequency*
- DF : *Document Frequency*
- IDF : *Inverse Document Frequency*
- (t) : *Term* atau sebuah kata dalam komentar
- N : Jumlah total dokumen

2.6 SVM Clasification

Menurut Lidya, SVM bekerja dengan prinsip *supervised learning* untuk melakukan analisis data dengan mencari *hyperplane* terbaik sebagai pemisah antar kelas, sehingga efektif digunakan dalam klasifikasi sentimen [21]. Dalam penelitian ini, metode SVM diterapkan karena mampu mengklasifikasikan sentimen ulasan pemain game *Clash of Clans* secara lebih akurat berdasarkan pola yang terbentuk dari data teks. Fungsi dasar dalam klasifikasi SVM dapat dirumuskan melalui [22]:

$$f(x) = w \cdot x + b$$

Keterangan:

- w : vektor bobot (*weight*)
- x : vektor input (hasil TF-IDF)
- b : bias
- f(x) : fungsi keputusan

2.7 Evaluasi Model

Kinerja model dievaluasi melalui *confusion matrix* untuk membandingkan hasil klasifikasi dengan label asli pada data uji [23]. Matriks ini berfungsi sebagai standar evaluasi klasifikasi dua kelas untuk melihat akurasi pemisahan data positif dan negatif [24]. Elemen-elemen yang menyusun *confusion matrix* meliputi:

- a. *True Positive (TP)* : Data positif diprediksi positif.
- b. *True Negative (TN)* : Data negatif diprediksi negatif.
- c. *False Positive (FP)* : Data negatif diprediksi positif.
- d. *False Negative (FN)* : Data positif diprediksi negatif.

Tabel 1. *Confusion Matrix*

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

Dari *confusion matrix*, dapat dihitung metrik evaluasi seperti tingkat *accuracy*, *precision*, *recall*, dan *F1-score*. Rumus untuk menghitung setiap metrik tersebut meliputi [25]:

a. *Accuracy*: Sebagai persentase dari semua data yang diprediksi dengan benar.

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

b. *Recall*: Tingkat keberhasilan model dalam mengidentifikasi seluruh data positif yang sebenarnya.

$$Recall \text{ (positif)} = \frac{TP}{TP+FN}$$

$$Recall \text{ (negatif)} = \frac{TN}{TN+FP}$$

c. *Precision*: Rasio ketepatan antara hasil prediksi positif dengan data aktualnya.

$$Precision \text{ (positif)} = \frac{TP}{TP+FP}$$

$$Precision \text{ (negatif)} = \frac{TN}{TN+FN}$$

d. *F1-Score*: Rata-rata harmonik yang menyeimbangkan nilai *Precision* dan *Recall*.

$$F1\text{-Score} \text{ (positif)} = \frac{2x \text{ Precision (positif)} x \text{ Recall (positif)}}{\text{Precision (positif)} + \text{Recall (positif)}}$$

$$F1\text{-Score} \text{ (negatif)} = \frac{2x \text{ Precision (negatif)} x \text{ Recall (negatif)}}{\text{Precision (negatif)} + \text{Recall (negatif)}}$$

Metrik ini menggambarkan efektivitas model dalam membedakan kategori sentimen positif maupun negatif dengan proporsi yang seimbang, sehingga hasil evaluasi dapat menjadi dasar dalam menilai keandalan metode SVM dalam mengolah ulasan pemain.

3. Hasil dan Pembahasan

Penelitian ini menghasilkan temuan dari serangkaian tahapan pengolahan data. Ulasan pemain dikumpulkan melalui proses *scraping* kemudian melalui tahap *preprocessing*, pelabelan sentimen, pembobotan TF-IDF, dan klasifikasi menggunakan SVM. Setiap tahapan menampilkan *output* yang berbeda mengenai kondisi data maupun kinerja model. Hasil tersebut disajikan secara bertahap mulai dari jumlah ulasan yang terkumpul, distribusi sentimen, hingga nilai akurasi dan metrik evaluasi.

3.1 Hasil Scraping Data

Teknik *web scraping* digunakan untuk mengumpulkan data ulasan sebanyak 5.000 ulasan terbaru dari bulan April hingga Mei 2025, dengan pengurutan menggunakan *Sort NEWEST*. Rentang waktu ini dipilih untuk memperoleh gambaran persepsi terkini pemain terhadap *game Clash of Clans*, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Scraping* Data

No	Nama	Rating	Ulasan	Tanggal ulasan
1	Pengguna Google	1	It's good but the storage is getting worse so I can't play it	2025-05-11 23:52:37
2	Pengguna Google	5	game ini bagus saya sangat suka dengan game ini	2025-05-11 23:44:32
3	Pengguna Google	5	Cekurukuk 🍷 🍷 🍷 mantap polll	2025-05-11 23:22:13
⋮	⋮	⋮	⋮	⋮
5000	Pengguna Google	5	ok	2025-04-16 06:16:30

Hasil dari *scraping* data yaitu kumpulan ulasan mentah yang diambil dari *Google Play Store*. Data tersebut masih mengandung bahasa yang tidak baku, simbol, dan tanda baca yang belum tersaring, seperti yang ditunjukkan pada Tabel 2. Kondisi ini menunjukkan bahwa hasil *scraping* masih memerlukan proses pembersihan sebelum dapat digunakan dalam analisis.

Drop duplicates dilakukan untuk menghilangkan data ulasan yang semula 5.000 menjadi 3.529 agar data ulasan menjadi lebih bersih seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Drop Duplicates*

No	Nama	Rating	Ulasan	Tanggal ulasan
1	Pengguna Google	1	It's good but the storage is getting worse so I can't play it	2025-05-11 23:52:37
2	Pengguna Google	5	game ini bagus saya sangat suka dengan game ini	2025-05-11 23:44:32
3	Pengguna Google	5	Cekurukuk 🍷 🍷 🍷 mantap polll	2025-05-11 23:22:13
⋮	⋮	⋮	⋮	⋮
3529	Pengguna Google	1	suka ga jls tiba tiba jaringan leg padahal jaringan bagus	2025-04-22 21:19:13

3.2 Hasil *Preprocessing* Data

Data yang terkumpul melalui tahapan *preprocessing* untuk menjamin kesiapan analisis. Rangkaian proses ini meliputi *cleaning*, *case folding*, *word snormalization*, *tokenizing*, *stopword removal*, dan *stemming*. Dengan tahapan ini, ulasan pemain yang beragam dapat diseragamkan bentuk katanya sehingga lebih mudah dipahami oleh sistem. Selain itu, kata-kata umum yang tidak berpengaruh pada sentimen dibuang agar analisis lebih fokus. Proses ini penting karena ulasan *game* seringkali menggunakan bahasa gaul, singkatan, bahkan emoji yang perlu dibersihkan terlebih dahulu.

Tabel 4. Hasil *Cleaning*

Sebelum	Sesudah
It's good but the storage is getting worse so I can't play it	It s good but the storage is getting worse so I can t play it
game ini bagus saya sangat suka dengan game ini	game ini bagus saya sangat suka dengan game ini
Cekurukuk 🍷 🍷 🍷 mantap polll	Cekurukuk mantap polll
suka ga jls tiba tiba jaringan leg padahal jaringan bagus	suka ga jls tiba tiba jaringan leg padahal jaringan bagus

Tahap *cleaning* menghasilkan data yang bersih dari elemen non-teks seperti angka, tanda baca, serta simbol unik seperti yang ditunjukkan pada Tabel 4.

Tabel 5. *Case Folding*

Sebelum	Sesudah
It s good but the storage is getting worse so I can t play it	it s good but the storage is getting worse so i can t play it
game ini bagus saya sangat suka dengan game ini	game ini bagus saya sangat suka dengan game ini
bagus dan seru	bagus dan seru
Cekurukuk mantap polll	cekurukuk mantap polll
suka ga jls tiba tiba jaringan leg padahal jaringan bagus	suka ga jls tiba tiba jaringan leg padahal jaringan bagus

Hasil dari *Case Folding* yaitu mengubah teks yang semula huruf besar menjadi huruf kecil yang lebih konsisten seperti yang ditunjukkan pada Tabel 5.

Tabel 6. *Word Normalization*

Sebelum	Sesudah
it s good but the storage is getting worse so i can t play it	it s good but the storage is getting worse so i can t play it
game ini bagus saya sangat suka dengan game ini	game ini bagus saya sangat suka dengan game ini
bagus dan seru	bagus dan seru
cekurukuk mantap polll	cekurukuk mantap sekali
suka ga jls tiba tiba jaringan lag padahal jaringan bagus	suka tidak jelas tiba tiba jaringan lag padahal jaringan bagus

Hasil dari *Word Normalization* yaitu Mentransformasi istilah non-formal atau singkatan ke dalam kosakata standar seperti yang tertera pada Tabel 6.

Tabel 7. *Tokenizing*

Sebelum	Sesudah
it s good but the storage is getting worse so i can t play it	[it, s, good, but, the, storage, is, getting, worse, so, i, can, t, play, it]
game ini bagus saya sangat suka dengan game ini	[game, ini, bagus, saya, sangat, suka, dengan, game, ini]
bagus dan seru	[bagus, dan, seru]
cekurukuk mantap sekali	[cekurukuk, mantap, sekali]
suka tidak jelas tiba tiba jaringan lag padahal jaringan bagus	[suka, tidak, jelas, tiba, tiba, jaringan, lag, padahal, jaringan, bagus]

Hasil dari *Tokenizing* yaitu memecah sebuah teks menjadi bagian-bagian kata atau token seperti yang ditunjukkan pada Tabel 7.

Tabel 8. *Stopword Removal*

Sebelum	Sesudah
[it, s, good, but, the, storage, is, getting, worse, so, i, can, t, play, it]	[good, storage, getting, worse, play]
[game, ini, bagus, saya, sangat, suka, dengan, game, ini]	[game, bagus, suka, game]
[bagus, dan, seru]	[bagus, seru]
[cekurukuk, mantap, sekali]	[cekurukuk, mantap, sekali]
[suka, tidak, jelas, tiba, tiba, jaringan, lag, padahal, jaringan, bagus]	[suka, jelas, tiba, tiba, jaringan, lag, jaringan, bagus]

Tahap *stopword removal* mengeliminasi kata-kata fungsional yang minim makna, dengan hasil yang tertera pada Tabel 8.

Tabel 9. *Stemming*

Sebelum	Sesudah
[good, storage, getting, worse, play]	good storage get worse play
[game, bagus, suka, game]	game bagus suka game
[bagus, seru]	bagus seru
[cekurukuk, mantap, sekali]	cekurukuk mantap sekali
[suka, jelas, tiba, tiba, jaringan, lag, jaringan, bagus]	suka jelas tiba tiba jaringan lag jaringan bagus

Proses *stemming* mereduksi kata menjadi bentuk akar dengan mengeliminasi imbuhan, sebagaimana dipaparkan pada Tabel 9.

3.3 Hasil Labeling Sentimen

Tahap pelabelan diterapkan untuk memisahkan ulasan menjadi kategori sentimen positif serta negatif, sehingga mempermudah pemahaman mengenai persepsi pemain terhadap game *Clash of Clans* yang tertera pada Tabel 10.

Tabel 10. Hasil Labeling Sentimen

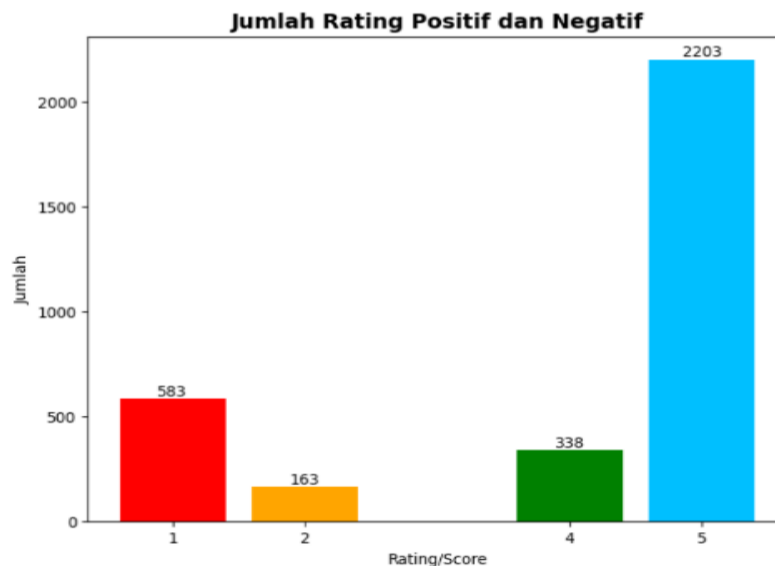
Rating	Ulasan	Sentimen
1	good storage get worse play	Negatif
5	game bagus suka game	Positif
5	bagus seru	Positif
5	cekurukuk mantap sekali	Positif
1	suka jelas tiba tiba jaringan lag jaringan bagus	Negatif

```
Jumlah masing-masing sentimen:
Labeling_Sentimen
positif      2541
negatif      746
Name: count, dtype: int64

Jumlah data setelah filtering: 3287
```

Gambar 2. Jumlah Labeling Sentimen

Dari total 3.287 data, terdapat 2.541 ulasan positif dan 746 ulasan negative, sementara ulasan netral diabaikan untuk menghindari ambiguitas. Hal ini menunjukkan bahwa sentimen positif lebih mendominasi dibanding negatif.



Gambar 3. Jumlah Rating Positif dan Negatif

Gambar 3 menunjukkan jumlah rating pemain game *Clash of Clans* di Google Play Store dengan rating 3 ditiadakan karena membuat ambiguitas. Rating 5 mendominasi (2.203 ulasan), disusul rating 1 (583), rating 4 (338), dan rating 2 (163), yang mencerminkan mayoritas ulasan positif meski masih ada keluhan pemain.

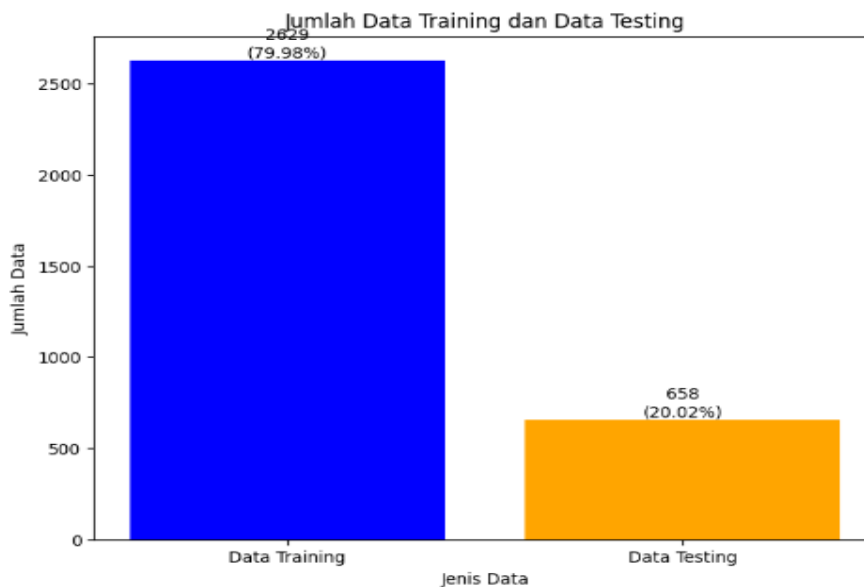
3.4 Hasil *Term Frequency-Inverse Document Frequency* (TF-IDF)

Sebelum memasuki klasifikasi, ulasan diberi label sentimen kemudian diubah ke bentuk numerik dengan TF-IDF. Metode ini memberi bobot lebih tinggi pada kata yang jarang muncul di keseluruhan data namun sering muncul dalam satu ulasan. Hasilnya berupa vektor numerik yang memudahkan model menangkap pola tekstual guna memisahkan kategori sentimen positif dari negatif.

Tabel 11. Hasil Perhitungan TF-IDF

Kata (t)	TF	DF	IDF	TF-IDF
bagus	$1/7 = 0,14$	714	3,983	0,557
game	$1/7 = 0,14$	1049	3,031	0,424
cepat	$1/7 = 0,14$	40	53,02	7,422
serang	$1/7 = 0,14$	159	14,33	2,006
masak	$1/7 = 0,14$	31	67,65	9,471
tentara	$1/7 = 0,14$	14	143	20,02
perang	$1/7 = 0,14$	125	17,92	2,508

Tabel 11 menampilkan hasil salah satu contoh ulasan yaitu “bagus game cepat serangan masak tentara perang”. Setiap kata memiliki nilai TF sebesar $1/7 = 0,14$ karena masing-masing kata muncul sekali dari total tujuh kata. Nilai DF menunjukkan jumlah ulasan dalam himpunan data yang memuat kata tersebut dan hasil perhitungan TF-IDF menunjukkan bahwa kata “tentara” memperoleh bobot tertinggi (20,02) karena jarang muncul di seluruh dokumen sehingga dianggap paling penting, sedangkan kata “game” memiliki bobot rendah (0,424) karena sering muncul sehingga tingkat kekhususannya kecil. Adapun kata “masak” dan “cepat” memiliki bobot cukup tinggi (9,471 dan 7,422) sehingga berperan penting dalam ulasan. Setelah itu, Pembagian data dilakukan dengan proporsi 80% untuk fase pelatihan dan 20% untuk tahap pengujian untuk memastikan model dapat dievaluasi secara objektif.



Gambar 4. Hasil *Splitting* Data

Berdasarkan tampilan Gambar 4, tampak bahwa 2.629 data (79,98%) digunakan untuk data *training*, sedangkan 658 data (20,02%) untuk data *testing*, sehingga evaluasi model menjadi lebih objektif.

3.5 Hasil Klasifikasi SVM

Perhitungan *Support Vector Machine* dilakukan berdasarkan salah satu contoh ulasan komentar yang sebelumnya telah melalui proses TF-IDF, yaitu “bagus *game* cepat serang masak tentara perang”, emudian hasil tersebut ditunjukkan pada Tabel 12.

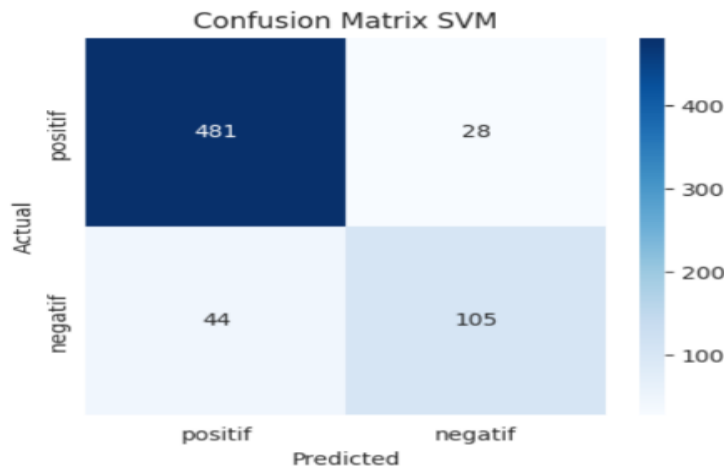
Tabel 12. Hasil Perhitungan SVM

Kata (t)	Bobot (w)	TF-IDF (x)	w * x
bagus	0,978	0,557	0,544
game	0,120	0,424	0,050
cepat	-0,115	7,422	-0,853
serang	-1,141	2,006	-2,288
masak	0,203	9,471	1,922
tentara	0,288	20,02	5,765
perang	-0,389	2,508	-0,975
Bias (b)	-	-	0,608
Total Fungsi Keputusan f(x)	-	-	4,773

Berdasarkan hasil perhitungan fungsi keputusan salah satu ulasan pemain *game Clash of Clans* menggunakan SVM, diperoleh nilai f(x) sebesar 4,773 yang diklasifikasikan sebagai sentimen positif. Nilai ini berasal dari bobot kata TF-IDF guna merepresentasikan bobot relevansi sebuah kata di dalam ulasan. Kata bagus, cepat, perang menjadi indikator positif terhadap pengalaman bermain, menandakan pembaruan *game* dianggap berhasil meningkatkan kualitas permainan. Hasil klasifikasi dapat memberikan masukan bagi pengembang dalam menjaga mutu *game* serta merancang pembaruan sesuai ekspektasi pemain.

3.6 Hasil Evaluasi Model

Pengukuran kinerja model menggunakan *confusion matrix* sebagai basis perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. *Word cloud* ditunjukkan untuk memperlihatkan kata dominan pada tiap sentimen, sehingga memudahkan pemahaman pola dan karakteristik kata dalam analisis.



Gambar 5. *Confusion Matrix*

Confusion matrix menunjukkan hasil prediksi model SVM pada 658 data uji. Sebanyak 481 data positif diprediksi benar (TP), 28 positif salah diprediksi negatif (FN), 105 negatif diprediksi benar (TN), dan 44 negatif salah diprediksi positif (FP).

```

Classification Report:
              precision    recall  f1-score   support

   negatif      0.79      0.70      0.74      149
   positif      0.92      0.94      0.93      509

 accuracy      0.85      0.82      0.89      658
 macro avg     0.85      0.82      0.84      658
 weighted avg  0.89      0.89      0.89      658

Confusion Matrix:
[[481  28]
 [ 44 105]]

Accuracy: 0.8905775075987842
    
```

Gambar 6. Hasil Evaluasi Kinerja Model SVM

Gambar 6 menunjukkan evaluasi model klasifikasi sentimen dua kelas dengan *accuracy* sebesar 89%. Untuk sentimen positif diperoleh *precision* 92%, *recall* 94%, dan *F1-score* 93%, sedangkan sentimen negatif memiliki *precision* 79%, *recall* 70%, dan *F1-score* 74%.

1. Accuracy

Tingkat ketepatan klasifikasi sentimen diukur dengan metrik *accuracy*, yaitu perbandingan antara jumlah prediksi benar dengan total data uji. Rumus *accuracy* ditunjukkan sebagai berikut:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} = \frac{481+105}{481+44+105+28} = \frac{586}{658} = 0,89$$

2. Recall

Untuk memperoleh nilai *recall* pada masing-masing kelas, kemampuan model dalam mengenali data yang benar dari kelas tersebut diukur, digunakan rumus berikut:

$$Recall \text{ (positif)} = \frac{TP}{TP+FN} = \frac{481}{481+28} = \frac{481}{509} = 0,94$$

$$Recall \text{ (negatif)} = \frac{TN}{TN+FP} = \frac{105}{105+44} = \frac{105}{149} = 0,70$$

3. Precision

Precision merepresentasikan rasio keakuratan model dalam memvalidasi jumlah prediksi yang terbukti relevan dengan label asli. Untuk memperoleh nilai *precision* pada masing-masing kelas, digunakan rumus berikut:

$$Precision \text{ (positif)} = \frac{TP}{TP+FP} = \frac{481}{481+40} = 0,92$$

$$Precision \text{ (negatif)} = \frac{TN}{TN+FN} = \frac{105}{105+28} = 0,79$$

4. F1-Score

F1-Score mengevaluasi performa model dengan memperhatikan kedua metrik, *precision* dan *recall*, khususnya pada distribusi kelas yang memiliki ketimpangan jumlah. Untuk memperoleh nilai *F1-Score* pada masing-masing kelas, digunakan rumus berikut:

$$\begin{aligned}
 F1\text{-Score (positif)} &= \frac{2x \text{ Precision (positif)} x \text{ Recall (positif)}}{\text{Precision (positif)} + \text{Recall (positif)}} \\
 &= \frac{2x 0,92 x 0,94}{0,92 + 0,94} = \frac{1,729}{1,86} = 0,93
 \end{aligned}$$

$$\begin{aligned}
 F1\text{-Score (negatif)} &= \frac{2x \text{ Precision (negatif)} x \text{ Recall (negatif)}}{\text{Precision (negatif)} + \text{Recall (negatif)}} \\
 &= \frac{2x 0,79 x 0,70}{0,79 + 0,70} = \frac{1,106}{1,49} = 0,74
 \end{aligned}$$

-
- [4] I. Siti Aisah, B. Irawan, and T. Suprapti, "Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur'an Digital," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 6, pp. 3759–3765, 2024, doi: 10.36040/jati.v7i6.8263.
- [5] Septi Putri, Yohanes Agung Apriyanto, and Andri Wijaya, "Analisis Sentimen Ulasan Aplikasi DeepL Pada Google Play Dengan Metode Support Vector Machine (Svm)," *J. Sist. Inf.,* vol. 4, no. 2, pp. 59–66, 2023, doi: 10.32546/jusin.v4i2.2368.
- [6] M. A. A. T. Utami, P. Silvianti, and M. Masjukur, "Algoritme Support Vector Machine untuk Analisis Sentimen Berbasis Aspek Ulasan Game Online Mobile Legends: Bang-Bang," *Xplore J. Stat.,* vol. 12, no. 1, pp. 63–77, 2023, doi: 10.29244/xplore.v12i1.1064.
- [7] R. F. Ishar and A. Febrianto, "Analisis Sentimen Komentar Pengguna Terhadap Game Moba Lokapala di *Google Play Store* Menggunakan Algoritma *Support Vector Machine* dan Menawan Yang Dapat Dimainkan Secara Bersamaan Oleh Jutaan Pemain di Indonesia Bahkan di Seluruh Dunia," vol. 4, no. 2, pp. 1–9, 2024.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.,* vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] S. Fide, S. Suparti, and S. Sudarno, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi," *J. Gaussian,* vol. 10, no. 3, pp. 346–358, 2021, doi: 10.14710/j.gauss.v10i3.32786.
- [10] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi),* vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [11] M. M. Muttaqin, Wahyu Wijaya Widiyanto, A. W. Green Ferry Mandias, Stenly Richard Pungus, S. A. H. Wiranti Kusuma Hapsari, E. F. B. Aslam Fatkhudin, Pasnur, and N. S. Mochammad Anshori, Suryani, *Pengenalan Data Mining,* no. July. 2023.
- [12] P. Rahayu *et al.,* *Buku Ajar Data Mining,* vol. 1, no. January 2024. 2024.
- [13] Asrumi, *Analisis Sentimen Dan Penggalan Opini.* Purbalingga, Jawa Tengah: EUREKA MEDIA AKSARA, 2023.
- [14] D. Purnamasari, *Pengantar Metode Analisis Sentimen,* Edisi Pert. Depok: Gunadarma Penerbit, 2023.
- [15] E. Dwi, K. Wardani, F. F. Yo, W. N. Meylugita, U. Katolik, and M. Charitas, "Implementasi Algoritma *Naïve Bayes* Untuk Analisis Ulasan *Implementation Of The Naive Bayes Algorithm For User Review,*" vol. 4, no. 1, pp. 13–24, 2025.
- [16] M. Haikal, M. Martanto, and U. Hayati, "Analisis Sentimen Terhadap Penggunaan Aplikasi Game Online Pubg Mobile Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 6, pp. 3275–3281, 2024, doi: 10.36040/jati.v7i6.8174.
- [17] W. Nurfitri and A. Chowanda, "Analisis Sentimen Pada Kasus Positif Covid-19 Berdasarkan Pemberitaan Media Di Indonesia Menggunakan Indobert," *Progresif J. Ilm. Komput.,* vol. 20, no. 1, p. 580, 2024, doi: 10.35889/progresif.v20i1.1897.
- [18] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.,* vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [19] A. F. Azmi and A. Voutama, "KOMPUTA : Jurnal Ilmiah Komputer dan Informatika Prediksi Churn Nasabah Bank Menggunakan Klasifikasi *Random Forest* dan *Decision Tree* Dengan Evaluasi *Confusion Matrix* Komputa : Jurnal Ilmiah Komputer dan Informatika," vol. 13, no. 1, 2024.
- [20] T. Nct, D. Menggunakan, and A. P. Wanda, "Analisis Sentimen di Media Sosial X (Twitter) Terhadap NCT Dream Menggunakan Algoritma Naïve Bayes," 2024.

-
- [21] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.
- [22] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [23] K. Putri *et al.*, "Implementasi Algoritma Support Vector Machine dalam Klasifikasi Deteksi Depresi dari Postingan pada Media Sosial," *J. Nas. Teknol. Inf. dan Apl.*, vol. 2, no. 1, pp. 193–202, 2023.
- [24] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.
- [25] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecom.v5i1.3708.