

Analisis Sentimen Komunitas Counter-Strike 2 (CS2) Menggunakan Support Vector Machine (SVM)

Saiful Faris Riyadi *, Yulison Herry Chrisnanto, Gunawan Abdillah

Fakultas Sains dan Informatika, Program Teknik Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

Email: ^{2,*}yhc@if.unjani.ac.id, ¹saifulfaris21@if.unjani.ac.id, ³gunawanabdillah03@gmail.com

Email Penulis Korespondensi: yhc@if.unjani.ac.id

Submitted 17-05-2025; Accepted 13-06-2025; Published 30-06-2025

Abstrak

Counter-Strike 2 (CS2) adalah game yang telah mendapatkan banyak antusiasme dari komunitas gaming sejak dirilis. Ulasan pengguna di platform Steam menjadi sumber utama untuk memahami sentimen komunitas terhadap game ini. Penelitian ini bertujuan untuk menganalisis sentimen terhadap ulasan CS2 menggunakan metode Support Vector Machine (SVM). Data dikumpulkan melalui platform Apify, kemudian dibersihkan melalui proses seperti tokenisasi, penghapusan kata stop, dan lemmatisasi. Fitur teks dikonversi menjadi nilai numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) untuk digunakan dalam model SVM. Model SVM digunakan untuk mengklasifikasikan sentimen ulasan ke dalam tiga kategori: positif, netral, dan negatif. Evaluasi dilakukan dengan mengukur akurasi, confusion matrix, dan laporan klasifikasi. Hasil evaluasi menunjukkan bahwa model SVM dengan pendekatan One-vs-Rest (OVR) tanpa SMOTE menghasilkan akurasi sebesar 81,95%. Setelah menerapkan teknik Synthetic Minority Over-sampling (SMOTE) pada data pelatihan untuk menyeimbangkan distribusi antara kelas, akurasi model meningkat sedikit menjadi 82,18%. Penelitian ini memberikan wawasan berharga bagi pengembang game dalam memahami pendapat pemain tentang CS2. Selain itu, penelitian ini menunjukkan potensi SVM dalam analisis sentimen berbasis teks pada platform ulasan pengguna.

Kata Kunci: Analisis Sentimen; Counter-Strike 2; Support Vector Machine; TF-IDF; Ulasan Pengguna

Abstract

Counter-Strike 2 (CS2) is a game that has received a lot of enthusiasm from the gaming community since its release. User reviews on the Steam platform are the main source for understanding community sentiment towards this game. This study aims to analyze sentiment towards CS2 reviews using the Support Vector Machine (SVM) method. Data was collected through the Apify platform, then cleaned through processes such as tokenization, stopword removal, and lemmatization. Text features were converted into numerical values using Term Frequency-Inverse Document Frequency (TF-IDF) to be used in the SVM model. The SVM model was used to classify review sentiment into three categories: positive, neutral, and negative. Evaluation was conducted by measuring accuracy, confusion matrix, and classification reports. In the evaluation results, the SVM model using the One-vs-Rest (OVR) approach showed that the model without SMOTE produced an accuracy of 81.95%. After applying the Synthetic Minority Over-sampling (SMOTE) technique to the training data to balance the distribution between classes, the model accuracy increased slightly to 82.18%. This study provides valuable insights for game developers in understanding players' opinions about CS2. Additionally, this study demonstrates the potential of SVM in text-based sentiment analysis on user review platforms.

Keywords: Sentiment Analysis; Counter-Strike 2; Support Vector Machine; TF-IDF; User Reviews

1. PENDAHULUAN

Perkembangan industri game digital telah mengalami peningkatan yang signifikan, baik dari segi teknologi maupun jumlah pemainnya. Salah satu bentuk interaksi yang berharga antara pemain dan pengembang game adalah melalui ulasan pengguna. Ulasan ini merupakan sumber informasi yang penting dalam menilai kualitas sebuah game karena merefleksikan pengalaman dan persepsi langsung dari pemain. Namun, volume ulasan yang terus meningkat di berbagai platform seperti Steam menyulitkan pengembang untuk menganalisis dan memahami sentimen pemain secara manual. Selain itu, sebagian besar ulasan hanya dinilai secara umum tanpa mempertimbangkan aspek-aspek tertentu seperti gameplay, visual, performa, dan sebagainya. Tantangan lainnya adalah terbatasnya jumlah data yang digunakan pada penelitian-penelitian sebelumnya dan pengelompokan sentimen yang masih terbatas pada dua kelas saja (positif dan negatif), padahal ada juga pendapat yang bersifat netral. Sebagai solusi dari permasalahan tersebut, penelitian ini akan melakukan analisis sentimen berbasis aspek pada ulasan pengguna game Counter-Strike 2 (CS2) yang diambil dari platform Steam, dengan menggunakan metode klasifikasi Support Vector Machine (SVM). Penelitian ini berfokus pada data ulasan berbahasa Inggris untuk meningkatkan jangkauan analisis dan menyertakan tiga label sentimen: positif, negatif, dan netral untuk lebih mencerminkan keragaman pendapat pengguna. Selain itu, penelitian ini menggunakan jumlah data yang lebih besar.

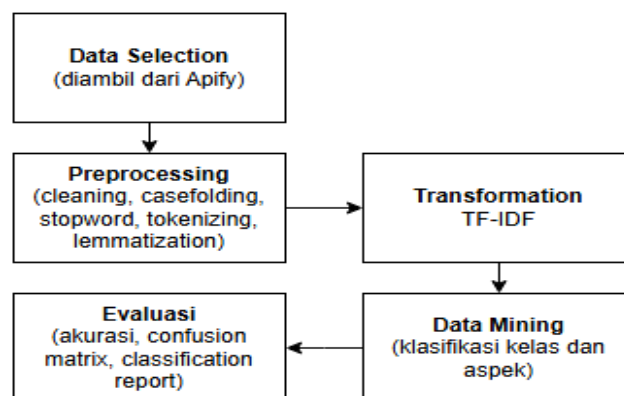
Beberapa penelitian sebelumnya telah meneliti analisis sentimen pada ulasan game. Penelitian [1] menggunakan metode SVM untuk analisis sentimen berbasis aspek pada ulasan game Mobile Legends: Bang-Bang dengan kategori seperti gameplay, visualisasi, performa, dan pemain. Namun, penelitian tersebut hanya menggunakan 2295 ulasan dari 5000 hasil scraping, dan semuanya berbahasa Indonesia, sehingga cakupan datanya masih terbatas. Penelitian pada [2] menunjukkan bahwa SVM tanpa teknik seleksi fitur menghasilkan akurasi sebesar 61.47% dengan menggunakan TF-IDF sebagai metode ekstraksi fitur. Pada penelitian [3], SVM memberikan hasil akurasi hingga 90% pada pembagian data 90:10, sedangkan Naïve Bayes hanya mendapatkan 72.4% pada konfigurasi yang sama. Penelitian lain membandingkan performa antara BERT dan SVM, dimana BERT menghasilkan akurasi sebesar 98%, sedangkan SVM hanya mencapai 81% [4]. Namun, pemilihan SVM pada penelitian ini masih relevan karena memiliki performa yang cukup tinggi dibandingkan dengan metode klasik seperti Naïve Bayes dan lebih efisien secara komputasi dibandingkan dengan deep

learning, seperti BERT. Secara umum, analisis sentimen adalah proses mengklasifikasikan opini dalam teks ke dalam kategori positif, negatif, atau netral [5]. Tujuan utamanya adalah untuk memahami sikap atau perasaan seseorang terhadap suatu topik tertentu, termasuk terhadap produk digital seperti video game. Menurut [6], analisis sentimen membantu menggambarkan opini publik melalui evaluasi konten digital dari berbagai sumber, seperti media sosial dan platform distribusi game digital. Penelitian ini akan memanfaatkan ulasan pengguna dari platform Steam, yang merupakan salah satu layanan distribusi game digital terbesar saat ini. Selain sebagai pusat distribusi game, Steam juga menjadi ruang komunitas tempat para pemain berdiskusi dan secara terbuka menyampaikan keluhan atau apresiasinya [7]. Dinamika opini pengguna di platform ini dapat mencerminkan tren persepsi dan ekspektasi terhadap performa sebuah game, termasuk pada game kompetitif seperti CS2 [8]. Berdasarkan data dari SteamDB, tercatat bahwa satu bulan setelah perilisan CS2, sebanyak 1.392.488 pemain bermain secara bersamaan di platform ini [9]. Fakta ini menunjukkan betapa besarnya antusiasme komunitas pemain terhadap game CS2 dan pentingnya menganalisis opini mereka. Menurut [10] Term Frequency - Inverse Document Frequency (TF-IDF) adalah metrik empiris yang menunjukkan seberapa penting suatu istilah dalam sebuah dokumen di dalam kumpulan dokumen. TF mengukur frekuensi kemunculan kata dalam dokumen, sedangkan IDF menentukan seberapa penting suatu istilah dengan mengevaluasi seberapa sering istilah tersebut muncul di seluruh dokumen. TF-IDF digunakan untuk mengekstraksi fitur dari data teks, yang membantu dalam menilai pentingnya setiap kata dalam komentar pengguna [11]. Untuk konteks Indonesia, leksikon lokal seperti InSet Lexicon telah dikembangkan untuk mendukung kebutuhan analisis sentimen dalam bahasa Indonesia [12]. Untuk bahasa Inggris, pendekatan serupa dapat diterapkan dengan menggunakan kamus seperti VADER, yang secara khusus didesain untuk format teks pendek seperti microblog [13]. Proses pengumpulan data ulasan dilakukan dengan bantuan Apify, sebuah platform web scraping yang memungkinkan pengambilan data dari situs web secara otomatis [14]. Dengan menggunakan aktor yang dapat disesuaikan, pengambilan data dari halaman ulasan di Steam dapat dilakukan dengan cepat dan efisien. Cukup dengan menyalin tautan ulasan ke dalam Apify, sistem akan mengekstrak data yang dibutuhkan secara otomatis [15].

Tujuan utama dari penelitian ini adalah untuk membangun model analisis sentimen berbasis aspek menggunakan SVM pada ulasan pengguna game CS2 berbahasa Inggris dengan klasifikasi sentimen positif, negatif, dan netral. Diharapkan hasil dari penelitian ini dapat memberikan gambaran yang lebih akurat mengenai opini pengguna terhadap berbagai aspek dalam game, dan menjadi alat yang berguna bagi pengembang game dalam meningkatkan kualitas produk. Selain itu, penelitian ini diharapkan dapat memperkaya literatur analisis sentimen berbasis aspek, khususnya dalam domain video game, dengan pendekatan yang efisien namun tetap memberikan hasil yang kompetitif.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) yang terdiri dari lima tahap utama: Data Selection, Preprocessing, Transformation, Data Mining, dan Evaluation. Metode ini digunakan untuk mengekstraksi pengetahuan dari data teks ulasan pengguna game Counter-Strike 2 (CS2) yang diambil dari platform Steam, dapat dilihat pada gambar 1.



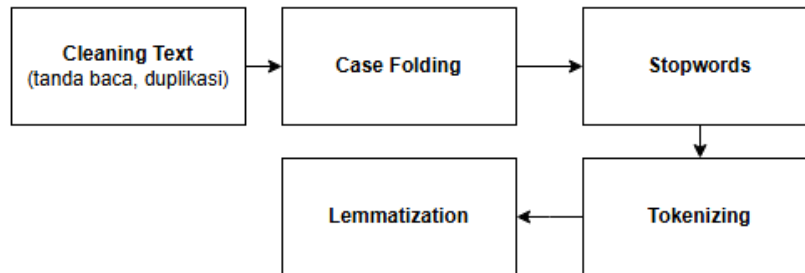
Gambar 1. Alur Penelitian

2.1 Data Selection

Data dalam penelitian ini dikumpulkan dari halaman ulasan pengguna di platform Steam dengan memanfaatkan Apify, sebuah layanan web scraping yang memungkinkan pengambilan data secara otomatis dan efisien dari situs web. Ulasan yang dikumpulkan berupa teks berbahasa Inggris yang mencerminkan opini dan persepsi pengguna terhadap game Counter-Strike 2 (CS2) setelah game tersebut dirilis secara resmi. Proses pengumpulan data dilakukan antara 28 September 2023 hingga 3 Februari 2025, untuk mendapatkan representasi opini pengguna yang lebih luas dan terkini. Ulasan yang dipilih adalah ulasan yang secara eksplisit membahas setidaknya satu aspek dari game dan dapat dikategorikan ke dalam salah satu dari tiga label sentimen, yaitu positif, negatif, atau netral, yang memungkinkan analisis sentimen berbasis aspek yang lebih mendalam.

2.2 Preprocessing

Sebelum digunakan dalam proses klasifikasi, data tinjauan melalui tahap preprocessing untuk memastikan kualitas data yang optimal. Proses prapemrosesan ini meliputi beberapa proses seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Alur tahap *Preprocessing*

2.2.1 Cleaning Text

Proses membersihkan suatu teks atau kalimat dari tanda baca yang tidak penting. Seperti simbol, angka, emoji, dan link.

2.2.2 Case Folding

Case folding merupakan salah satu tahapan penting dalam text preprocessing yang bertujuan untuk menyamakan bentuk huruf, yaitu dengan cara mengubah semua huruf dalam teks menjadi huruf kecil. Contoh “Shame Off VAC System VAC is Valve allow Cheating” menjadi “shame off vac system vac is valve allow cheating”.

2.2.3 Stopwords

Stopwords adalah kata-kata umum yang tidak memiliki arti dalam analisis. Pada tahap ini, kata-kata tersebut akan dihilangkan untuk memfokuskan pada kata-kata yang penting. Contoh dari “shame off vac system vac is valve allow cheating” menjadi “shame vac system vac valve allow cheating”

2.2.4 Tokenizing

Merupakan proses memecah beberapa kata atau kalimat menjadi unit-unit yang lebih kecil, biasanya disebut token. Contoh dari “shame vac system vac valve allow cheating” menjadi “shame”, “vac”, “system”, “vac”, “valve”, “allow”, “cheating”.

2.2.5 Lemmatization

Proses mengembalikan kata ke bentuk dasarnya atau disebut juga lemma berdasarkan makna dan struktur gramatikal. Contoh dari “cheating” menjadi “cheat”.

2.3 Transformation

Menurut [16], TF-IDF (Term Frequency – Inverse Document Frequency) adalah metode pembobotan kata yang digunakan dalam ekstraksi fitur untuk analisis sentimen. TF (Term Frequency) adalah jumlah kemunculan sebuah kata dalam dokumen tertentu, yang berarti semakin sering kata muncul, semakin besar nilai TF. Sedangkan IDF (Inverse Document Frequency) adalah ukuran seberapa jarang sebuah kata muncul di seluruh koleksi dokumen. Nilai TF-IDF merupakan hasil perkalian antara TF dan IDF, yang mencerminkan relevansi kata tersebut terhadap dokumen yang sedang dianalisis [17] dapat dilihat pada persamaan 1.

$$tfidf_{t,d} = tft_{t,d} \times idfd \quad (1)$$

Keterangan

$tfidf_{t,d}$: nilai tf-idf untuk *term* pada dokumen

$tft_{t,d}$: *term frequency*, seberapa sering *term* muncul dalam dokumen

$idfd$: *inverse document frequency*, ukuran keunikan *term* di seluruh dokumen dalam korpus

2.4 Data Mining

SVM adalah salah satu metode supervised learning yang digunakan untuk proses klasifikasi. Dalam model klasifikasi, SVM memiliki konsep yang lebih kuat dan jelas secara matematis. SVM digunakan untuk mencari hyperplane yang optimal dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk memisahkan kelas-kelas [18]. Secara umum, rumus SVM melibatkan fungsi keputusan linear atau kernel yang memetakan data ke ruang fitur yang lebih tinggi. Rumus dasar SVM untuk pengambilan keputusan dapat dilihat pada persamaan 2 [19].

$$f(x) = \text{sign}(\sum_{i=1}^n a_i y_i K(x_i, x) + b) \quad (2)$$

Keterangan

x adalah data uji,

x_i adalah data pelatihan,

y_i adalah label kelas $\{-1, 1\}$,

a_i adalah koefisien Lagrangian yang diperoleh dari proses optimasi,

$K(x_i, x)$ adalah kernel function, yang dalam penelitian ini digunakan kernel polynomial dan RBF (Radial Basis Function),

b adalah bias yang dihitung selama pelatihan.

2.5 Evaluation

Penelitian ini menghasilkan beberapa output analitis, seperti visualisasi awan kata untuk menggambarkan kata-kata yang dominan dalam setiap kategori sentimen, analisis distribusi label sentimen untuk mengidentifikasi ketidakseimbangan data, dan visualisasi confusion matrix untuk mengevaluasi kinerja model. Selanjutnya menurut [20], dalam evaluasi model klasifikasi seperti SVM, kinerja model umumnya diukur menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Metrik-metrik ini biasanya disajikan dalam bentuk laporan klasifikasi untuk menilai efektivitas keseluruhan model. Untuk lebih jelas dapat dilihat pada tabel 1.

Tabel 1. Classification Report

Kelas	Precision	Recall	F1-score	Support
Negatif	0.0	0.0	0.0	562
Netral	0.0	0.0	0.0	107
Positif	0.0	0.0	0.0	805
Accuracy			0.0	1474
Macro Avg	0.0	0.0	0.0	1474
Weighted Avg	0.0	0.0	0.0	1474

3. HASIL DAN PEMBAHASAN

Hasil penelitian ini telah menganalisis sentimen pemain terhadap permainan Counter-Strike 2 (CS2) dengan mengklasifikasikan ulasan pengguna yang diperoleh dari platform Steam. Metode yang digunakan adalah Support Vector Machine (SVM) dengan pendekatan multi-kelas One-vs-Rest (OvR), yang memungkinkan klasifikasi ke dalam tiga kategori sentimen: positif, netral, dan negatif. Sebelum proses klasifikasi, data ulasan melalui tahap preprocessing seperti pembersihan teks, penghapusan stopwords, tokenisasi, dan transformasi ke bentuk numerik menggunakan TF-IDF. Untuk mengatasi ketidakseimbangan data, metode SMOTE digunakan sebagai teknik oversampling. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, yang disajikan dalam laporan klasifikasi. Selain itu, hasil penelitian divisualisasikan melalui word cloud untuk melihat kata-kata dominan, distribusi label sentimen, dan confusion matrix untuk menganalisis akurasi prediksi per kelas.

3.1 Data Collection

Data set yang digunakan dalam penelitian ini terdiri dari total 10.530 ulasan pengguna terkait game Counter-Strike 2 (CS2). Data ulasan dikumpulkan menggunakan platform Apify, yang secara otomatis dan terstruktur mengekstrak data dari ulasan pengguna di platform Steam. Dari total ulasan, 3.163 ulasan tidak membahas aspek-aspek yang menjadi fokus penelitian dan oleh karena itu dihilangkan dari analisis lebih lanjut. Setelah penyaringan, tersisa 7.367 ulasan yang membahas setidaknya satu aspek yang dianalisis. Distribusi sentimen data yang disaring menunjukkan bahwa ulasan positif mendominasi dengan 4.022 ulasan, diikuti oleh ulasan negatif dengan 2.810 ulasan, dan ulasan netral dengan 535 ulasan. Penelitian sebelumnya, jumlah data ulasan yang digunakan sebanyak 2.295 ulasan dari hasil scraping Google Play Store selama periode Januari hingga Maret 2022 [1]. Perbedaan jumlah dan sumber data ini dapat memengaruhi hasil analisis serta akurasi model karena perbedaan gaya bahasa dan variasi opini pengguna. Dataset dapat dilihat pada tabel 2.

Tabel 2. Hasil pembagian kelas

Kelas	Jumlah Ulasan
Positif	4022
Netral	535
Negatif	2810

3.2 Preprocessing

Tahap pra-pemrosesan dilakukan untuk membersihkan dan mempersiapkan data teks sebelum analisis lebih lanjut. Proses ini meliputi penghapusan duplikat, teks kosong, dan karakter atau simbol yang tidak relevan. Selanjutnya, dilakukan normalisasi teks, seperti mengubah huruf menjadi huruf kecil (lowercasing), menghapus stopwords, tokenisasi, dan lemmatisasi. Tahap ini bertujuan untuk memastikan bahwa data konsisten dan siap untuk ekstraksi fitur dan pelatihan

model. Tahap ini penting agar data lebih mudah dipahami oleh algoritma yang digunakan. Lihat pada tabel 3 untuk contoh hasilnya.

Tabel 3. Hasil preprocessing data

Alur	No	Sebelum	Sesudah
Cleaning	1	It's good game. but RIP Aim, everytime i see my enemy either i'm shooting air or a wall?	Its good game but RIP Aim everytime i see my enemy either im shooting air or a wall
	2	ARE YOU HAPPY WITH LETTING CHEATER ROAM IN YOUR GAME? IT'S REALLY BAD FOR A BIG COMPANY BUT IT HAS A BAD SECURITY SYSTEM COMPARED TO VALORANT. OR DOES CS STAND FOR NOT COUNTER-STRIKE BUT CHEATER AND SMURF? JUST CLOSE THIS GAME IF THE CHEATER IS STILL THERE	ARE YOU HAPPY WITH LETTING CHEATER ROAM IN YOUR GAME ITS REALLY BAD FOR A BIG COMPANY BUT IT HAS A BAD SECURITY SYSTEM COMPARED TO VALORANT OR DOES CS STAND FOR NOT COUNTERSTRIKE BUT CHEATER AND SMURF JUST CLOSE THIS GAME IF THE CHEATER IS STILL THERE
Case Folding	1	Its good game but RIP Aim everytime i see my enemy either im shooting air or a wall	its good game but rip aim everytime i see my enemy either im shooting air or a wall
	2	ARE YOU HAPPY WITH LETTING CHEATER ROAM IN YOUR GAME ITS REALLY BAD FOR A BIG COMPANY BUT IT HAS A BAD SECURITY SYSTEM COMPARED TO VALORANT OR DOES CS STAND FOR NOT COUNTERSTRIKE BUT CHEATER AND SMURF JUST CLOSE THIS GAME IF THE CHEATER IS STILL THERE	are you happy with letting cheater roam in your game its really bad for a big company but it has a bad security system compared to valorant or does cs stand for not counterstrike but cheater and smurf just close this game if the cheater is still there
Stopwords	1	its good game but rip aim everytime i see my enemy either im shooting air or a wall	good game rip aim everytime see enemy either shooting air wall
	2	are you happy with letting cheater roam in your game its really bad for a big company but it has a bad security system compared to valorant or does cs stand for not counterstrike but cheater and smurf just close this game if the cheater is still there	happy letting cheater roam game bad big company bad security system compared valorant cs stand counterstrike cheater smurf close game cheater
Tokenizing	1	good game rip aim everytime see enemy either shooting air wall	['good', 'game', 'rip', 'aim', 'everytime', 'see', 'enemy', 'either', 'shooting', 'air', 'wall']
	2	happy letting cheater roam game bad big company bad security system compared valorant cs stand counterstrike cheater smurf close game cheater	['happy', 'letting', 'cheater', 'roam', 'game', 'bad', 'big', 'company', 'bad', 'security', 'system', 'compared', 'valorant', 'cs', 'stand', 'counterstrike', 'cheater', 'smurf', 'close', 'game', 'cheater']
Lemmatization	1	['good', 'game', 'rip', 'aim', 'everytime', 'see', 'enemy', 'either', 'shooting', 'air', 'wall']	['good', 'game', 'rip', 'aim', 'everytime', 'see', 'enemy', 'either', 'shoot', 'air', 'wall']
	2	['happy', 'letting', 'cheater', 'roam', 'game', 'bad', 'big', 'company', 'bad', 'security', 'system', 'compared', 'valorant', 'cs', 'stand', 'counterstrike', 'cheater', 'smurf', 'close', 'game', 'cheat']	['happy', 'let', 'cheat', 'roam', 'game', 'bad', 'big', 'company', 'bad', 'security', 'system', 'compare', 'valorant', 'cs', 'stand', 'counterstrike', 'cheater', 'smurf', 'close', 'game', 'cheat']

3.3 Transformation

Frekuensi kata adalah ukuran sederhana yang menghitung seberapa sering sebuah kata muncul dalam kumpulan teks atau dokumen. Kata-kata yang muncul lebih sering biasa Analisis frekuensi kata membantu dalam memahami pola penggunaan kata dalam data teks.nya dianggap lebih penting atau relevan dengan isi teks, seperti yang ditunjukkan dalam tabel 4 untuk melihat frekuensi kata.

Tabel 4. Frekuensi kata

Kata	Frekuensi
Good	2158
Cheater	1673
Player	938

Time	878
Fun	828
Well	802
Valve	706
Every	701
Still	697
Bad	686

Setelah mengamati frekuensi kemunculan kata (TF), penting untuk dicatat bahwa kata-kata umum yang sering muncul dalam banyak ulasan tidak memberikan informasi yang signifikan. Oleh karena itu, Inverse Document Frequency (IDF) digunakan untuk mengukur keunikan suatu kata berdasarkan jumlah ulasan yang mengandung kata tersebut. Dengan menggabungkan TF dan IDF menjadi TF-IDF, analisis dapat menyoroti kata-kata paling relevan dalam setiap ulasan. Metode ini membantu meningkatkan akurasi model Support Vector Machine (SVM) dalam mengklasifikasikan sentimen komunitas Counter-Strike 2. Seperti yang ditunjukkan dalam tabel 5.

Tabel 5. Frekuensi kata tf-idf

Kata	TF	IDF	TD-IDF
Good	0.018249	2.389012	0.043597
Cheater	0.014148	2.758917	0.039032
Player	0.007932	3.451277	0.027376
Time	0.007425	3.363872	0.024976
Fun	0.007002	3.334019	0.023345
Well	0.006782	3.411148	0.023135
Valve	0.005970	3.594983	0.021463
Every	0.005928	3.557512	0.021089
Still	0.005894	3.529877	0.020806
Bad	0.005801	3.585933	0.020802

3.4 Data Mining

Data sentimen dalam penelitian ini dibagi menjadi empat aspek utama, yaitu gameplay, performa, visualisasi, dan pemain, dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Untuk gameplay, data pelatihan terdiri dari 1.500 positif, 1.140 negatif, dan 311 netral, sementara data pengujian terdiri dari 420 positif, 254 negatif, dan 64 netral. Dalam hal kinerja, data pelatihan terdiri dari 313 positif, 200 negatif, dan 44 netral, sedangkan data pengujian terdiri dari 85 positif, 47 negatif, dan 8 netral. Visualisasi memiliki 248 positif, 92 negatif, dan 23 netral untuk pelatihan, serta 56 positif, 30 negatif, dan 5 netral untuk pengujian. Aspek pemain terdiri dari 531 positif, 469 negatif, dan 68 netral dalam pelatihan, serta 142 positif, 108 negatif, dan 18 netral dalam pengujian. Selain itu, untuk mengatasi ketidakseimbangan kelas, teknik oversampling SMOTE (Synthetic Minority Over-sampling Technique) digunakan untuk membuat distribusi data lebih seimbang. Lihat tabel 6 pembagian dataset.

Tabel 6. Pembagian dataset per aspek

Aspek	Kelas	80%	Total Training	20%	Total Test
Gameplay	Positif	1500	2951	420	738
	Netral	1140		254	
	Negatif	311		64	
Performa	Positif	313	557	85	140
	Netral	200		47	
	Negatif	44		8	
Visualisasi	Positif	248	363	56	91
	Netral	92		30	
	Negatif	23		5	
Pemain	Positif	531	1068	142	268
	Netral	469		108	
	Negatif	68		18	

3.5 Evaluation

Pada tahap evaluasi ini, model yang telah dilatih diuji menggunakan data pengujian untuk mengukur performa klasifikasi sentimen berdasarkan aspek dengan metrik akurasi, presisi, recall, dan F1-score. Matrix konfusi juga dianalisis untuk mengetahui distribusi prediksi yang benar dan salah pada setiap kelas sentimen. Teknik SMOTE digunakan pada data latih untuk mengatasi ketidakseimbangan kelas sehingga model Support Vector Machine (SVM) mampu memberikan hasil yang lebih seimbang pada kelas minoritas. Selain evaluasi model, dilakukan analisis word cloud untuk mengidentifikasi kata-kata yang paling dominan pada setiap sentimen dan aspek, membantu memvisualisasikan pola

utama dan fokus komentar pengguna. Dengan kombinasi evaluasi kuantitatif dan visualisasi word cloud, penelitian ini memberikan gambaran menyeluruh tentang performa model dan karakteristik data sentimen yang dianalisis.

3.5.1 Classification Report

Pada Hasil evaluasi model SVM menggunakan pendekatan One-vs-Rest (OVR) menunjukkan bahwa model tanpa SMOTE menghasilkan akurasi sebesar 81,95% dengan nilai f1-score tertinggi pada kelas positif sebesar 0,87, diikuti oleh kelas negatif sebesar 0,81, dan nilai terendah pada kelas netral sebesar 0,36. Nilai recall untuk kelas netral juga sangat rendah, yaitu 0,23, menunjukkan bahwa banyak data netral tidak diklasifikasikan dengan benar. Setelah menerapkan Teknik Over-sampling Minoritas Sintetis (SMOTE) pada data pelatihan untuk menyeimbangkan distribusi antar kelas, akurasi model meningkat sedikit menjadi 82,18%. Kinerja pada kelas netral juga membaik dengan recall meningkat menjadi 0,55 dan f1-score menjadi 0,50. Kelas positif dan negatif terus menunjukkan hasil yang konsisten dengan nilai presisi dan recall yang tinggi. Skor macro dan weighted rata-rata juga meningkat, menunjukkan bahwa penggunaan SMOTE membantu model mengenali kelas minoritas seperti netral secara lebih efektif. Dengan demikian, penerapan SMOTE berhasil meningkatkan keseimbangan kinerja model tanpa mengorbankan akurasi keseluruhan. Perbandingan hasil lengkap model sebelum dan setelah SMOTE dapat dilihat pada tabel 7 dan tabel 8.

Tabel 7. Classification Report SVM OVR

Kelas	Precision	Recall	F1-score	Support
Negatif	0.81	0.80	0.81	562
Netral	0.76	0.23	0.36	107
Positif	0.83	0.91	0.87	805
Accuracy			0.82	1474
Macro Avg	0.80	0.65	0.68	1474
Weighted Avg	0.82	0.82	0.81	1474

Tabel 8. Classification Report SVM OVR SMOTE

Kelas	Precision	Recall	F1-score	Support
Negatif	0.84	0.79	0.82	466
Netral	0.45	0.55	0.50	82
Positif	0.87	0.88	0.87	631
Accuracy			0.82	1179
Macro Avg	0.72	0.74	0.73	1179
Weighted Avg	0.83	0.82	0.82	1179

3.5.2 Confusion Matrix

Hasil confusion matrix menunjukkan bahwa model SVM OVR tanpa SMOTE cenderung bias ke kelas positif, dengan 733 data positif yang diklasifikasikan dengan benar, tetapi kinerja untuk kelas netral masih rendah (hanya 25 dari 107 yang benar). Setelah menerapkan teknik SMOTE, terjadi peningkatan pada kelas netral (menjadi 45 data benar dari 82), dan distribusi prediksi menjadi lebih seimbang di semua tiga kelas. Hal ini menunjukkan bahwa SMOTE secara efektif meningkatkan akurasi model untuk kelas minoritas. Hasil lengkap dapat dilihat pada tabel 9 dan 10. Selain itu, jumlah kesalahan prediksi pada kelas negatif juga berkurang setelah menerapkan SMOTE, menunjukkan peningkatan kinerja model secara keseluruhan. Perubahan ini memperkuat pentingnya menangani data yang tidak seimbang dalam analisis sentimen.

Tabel 9. Confusion Matrix OVR

Kelas Aktual	Kelas Prediksi		
	Positif	Netral	Negatif
Positif	450	4	108
Netral	38	25	44
Negatif	68	4	733

Tabel 10. Classification Report SVM OVR SMOTE

Kelas Aktual	Kelas Prediksi		
	Positif	Netral	Negatif
Positif	369	23	74
Netral	25	45	12
Negatif	45	31	555

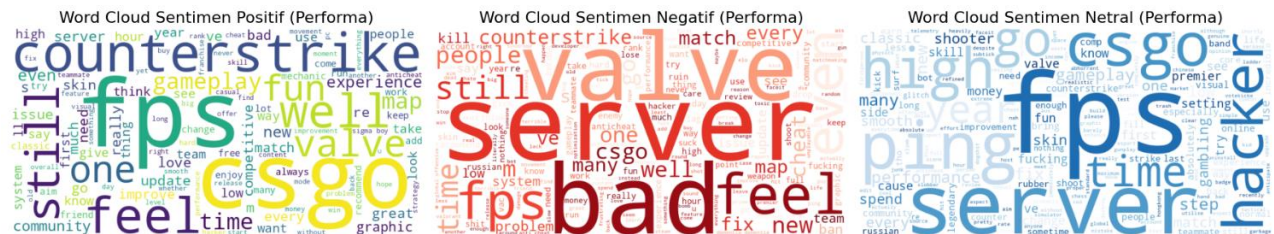
3.5.3 Word Cloud

Visualisasi awan kata digunakan untuk menampilkan kata-kata yang paling sering muncul dalam setiap kategori sentimen berdasarkan empat aspek utama, yaitu gameplay, performa, visualisasi, dan pemain. Dalam hal gameplay, kategori positif

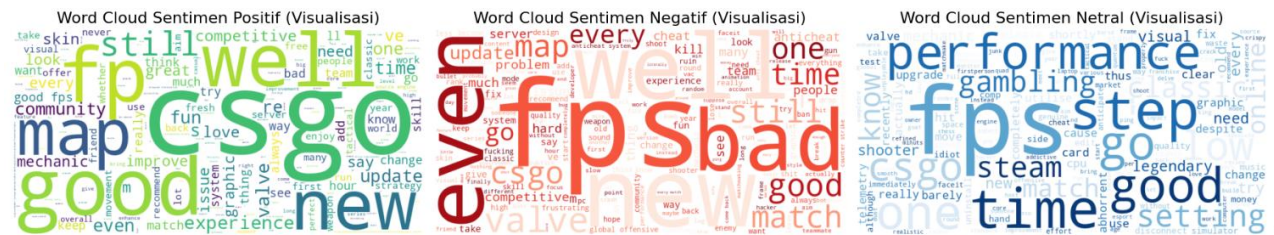
didominasi oleh kata-kata “fun” dan “great”, kategori negatif menampilkan kata-kata “still” dan “bad”, sedangkan kategori netral menampilkan kata-kata seperti “fix” dan “match”, lihat gambar 3 untuk hasilnya. Untuk aspek performa, ulasan positif menampilkan kata-kata “well” dan “fps”, ulasan negatif didominasi oleh “bad” dan “server”, dan ulasan netral menampilkan kata-kata “server” dan “time”, lihat gambar 4 untuk hasilnya. Dalam aspek visualisasi, sentimen positif menampilkan kata-kata “new” dan “csgo”, sentimen negatif menampilkan “fps” dan “bad”, dan sentimen netral menampilkan “performance” dan “gambling”, lihat gambar 5 untuk hasilnya. Akhirnya, aspek “pemain” dalam kategori positif menampilkan kata-kata seperti “fun” dan “community,” negatif oleh kata-kata seperti “cheat” dan “every,” dan netral oleh “every” dan “teammate”, lihat Gambar 6 untuk hasilnya.



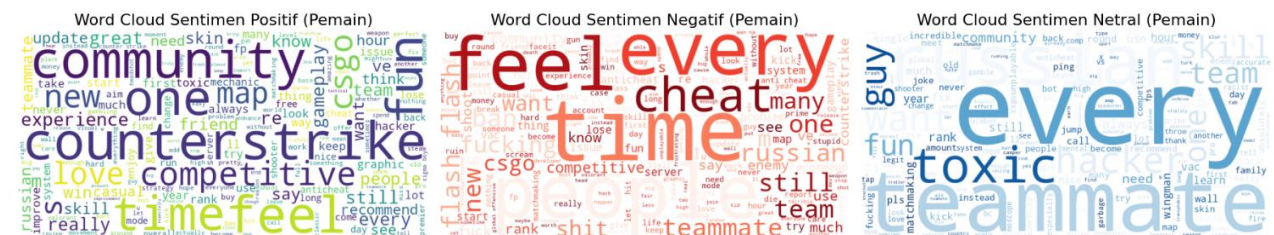
Gambar 3. Word Cloud Aspek *Gameplay*



Gambar 4. Word Cloud Aspek *Perfoma*



Gambar 5. Word Cloud Aspek *Visualisasi*



Gambar 6. Word Cloud Aspek *Pemain*

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Support Vector Machine (SVM) dengan teknik oversampling SMOTE untuk melakukan analisis sentimen pada ulasan game Counter-Strike 2 (CS2). Proses penelitian meliputi pengumpulan data, prapemrosesan teks, transformasi menggunakan TF-IDF, serta pemodelan dan evaluasi, yang menghasilkan akurasi yang baik pada data uji. Penggunaan teknik SMOTE berhasil mengatasi ketidakseimbangan kelas, memungkinkan model SVM mengenali sentimen positif, negatif, dan netral dengan lebih baik. Visualisasi data menggunakan word clouds juga memberikan gambaran umum tentang kata-kata dominan yang muncul dalam setiap aspek yang dianalisis, seperti gameplay, performa, visual, dan pemain. Hasil evaluasi menunjukkan bahwa model dengan SMOTE memiliki kinerja yang lebih stabil pada metrik presisi, recall, dan f1-score dibandingkan dengan model tanpa SMOTE. Namun, penelitian ini terbatas pada analisis sentimen menggunakan data berbahasa Inggris saja. Oleh karena itu, penelitian lebih lanjut diharapkan menggunakan data multibahasa untuk memperluas hasil analisis dan mencakup variasi bahasa yang lebih beragam. Selain itu, model yang dikembangkan belum diuji pada data real-time, sehingga pengujian lebih lanjut dengan

data streaming atau ulasan langsung dapat menjadi langkah penting untuk meningkatkan penerapan model dalam skenario dunia nyata. Dengan demikian, penelitian ini memberikan kontribusi awal dalam pemodelan analisis sentimen untuk komunitas gaming CS2 sambil membuka peluang untuk pengembangan yang lebih komprehensif di masa depan.

REFERENCES

- [1] M. A. A. T. Utami, P. Silvianti, and M. Masjkur, "Algoritme Support Vector Machine untuk Analisis Sentimen Berbasis Aspek Ulasan Game Online Mobile Legends: Bang-Bang," *Xplore: Journal of Statistics*, vol. 12, no. 1, pp. 63–77, Jan. 2023, doi: 10.29244/xplore.v12i1.1064.
- [2] R. Vincent, I. Maulana, and O. Komarudin, "PERBANDINGAN KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN DENGAN MULTICLASS DI TWITTER," 2023.
- [3] A. F. Alkindi and N. Nasution, "ANALISIS SENTIMEN ULASAN PENGGUNA PADA GAME ROBLOX DENGAN METODE SUPPORT VECTOR MACHINE DAN NAIVE BAYES," *J-Com (Journal of Computer)*, vol. 4, no. 2, pp. 164–177, Jul. 2024, doi: 10.33330/j-com.v4i2.3319.
- [4] MAWADDAH, H. (n.d.). ANALISIS DAN PERBANDINGAN KLASIFIKASI HATE SPEECH MENGGUNAKAN METODE FITUR MANUAL SVM DAN BERT.
- [5] A. Pangestu, Y. T. Arifin, and R. A. Safitri, "ANALISIS SENTIMEN REVIEW PUBLIK PENGGUNA GAME ONLINE PADA PLATFORM STEAM MENGGUNAKAN ALGORITMA NAÏVE BAYES," 2023.
- [6] E. Lutfina, W. Andriana, S. Quamila, P. Wiratmaja, and E. Febrianti, "Science, Technology and Management Journal Metode dan Algoritma Dalam Sentimen Analisis: Systematic Literature Review Info Artikel," 2024, doi: 10.53416/stmj.v4i2.
- [7] Y.-H. Chen, "XGBoost Hyperberd Model Using Steam Platform XGBoost Hyperberd Model Using Steam Platform XGBoost Hyperberd Model Using Steam Platform Gaming Data." [Online]. Available: <http://library.ucf.edu>
- [8] M. Y. Febrianta et al., "Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation."
- [9] SteamDB, "Counter - Strike 2 Steam Charts," <https://steamdb.info/app/730/charts/#1y>.
- [10] I. M. Karo Karo, R. Romia, S. Dewi, and P. M. Fadilah, "Hoax Detection on Indonesian Tweets using Naïve Bayes Classifier with TF-IDF," *Journal of Information System Research (JOSH)*, vol. 4, no. 3, pp. 914–919, Apr. 2023, doi: 10.47065/josh.v4i3.3317.
- [11] R. MuizK, R. Fajri Ishar, A. Febrianto, M. Nur Akbar, and K. Author, "ANALISIS SENTIMEN KOMENTAR PENGGUNA TERHADAP GAME MOBA LOKAPALA DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," vol. 4, no. 2, 2024.
- [12] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," 2022. [Online]. Available: <https://github.com/evanmartua34/>
- [13] M. Fernanda and N. Fathoni, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Jurnal Informatika dan Sains Teknologi*, vol. 1, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v1i3.112.
- [14] M. F. A. M. Muluk, "ANALISIS SENTIMEN PENGGUNA TIKTOK TERHADAP PRODUK FASHION DENGAN METODE NAÏVE BAYES," 2024.
- [15] F. Saifullah Fattah and C. Indah Ratnasari, "Sentiment Analysis of Indonesian Presidential Candidate," Fardhan Saifullah Fattah, 2023.
- [16] M. Hafizh Mahendra, D. Triantoro Mardiansyah, and K. Muslim Lhaksmana, "Dike : Jurnal Ilmu Multidisiplin Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," 2023.
- [17] K. Tri Putra, S. Anggraini, L. Sutriani, A. Imprun, and J. Informatika, "Analisis Sentimen Masyarakat Kalimantan Tengah Terhadap Perkebunan Kelapa Sawit Menggunakan TF-IDF dan Support Vector Machine," 2025.
- [18] S. F. H. D. D. I. A. Q. R. A. A. G. F. Rizki Wijayatun Pratiwi, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *Journal of Informatics, Information System, Software Engineering and Applications*, Nov. 21AD.
- [19] H. Apriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-ita/index>
- [20] S. Muawanah, U. Muzayanah, M. G. R. Pandin, M. D. S. Alam, and J. P. N. Trisnangtyas, "Stress and Coping Strategies of Madrasah's Teachers on Applying Distance Learning During COVID-19 Pandemic in Indonesia," *Qubahan Academic Journal*, vol. 3, no. 4, pp. 206–218, Nov. 2023, doi: 10.48161/Issn.2709-8206.