Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Analisis Sentimen Ulasan EA FC 26 di Steam: Perbandingan SVM dan Multinomial Naïve Bayes

Rafi Adha Azhar, Wahyu Widodo*

[1] Prodi Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer El Rahma Yogyakarta, Yogyakarta, Indonesia
Email: [1]rafiadha69@gmail.com, [2*]wahyu@stmikelrahma.ac.id
Email Penulis Korespondensi: wahyu@stmikelrahma.ac.id

Abstrak− Ulasan pengguna pada platform *Steam* menyimpan informasi penting mengenai persepsi pemain terhadap *EA FC 26*, tetapi jumlah data yang besar dan bentuk teks yang tidak terstruktur menyulitkan analisis manual. Penelitian ini bertujuan mengetahui kecenderungan sentimen pengguna serta membandingkan kinerja *Support Vector Machine* (SVM) dan *Multinomial Naïve Bayes* pada ulasan *EA FC 26* di *Steam*. Penelitian menggunakan tahapan *Knowledge Discovery in Database* (KDD), meliputi *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Sebanyak 5.000 ulasan terbaru berbahasa Inggris dikumpulkan melalui *Steam Review API*. Setelah pengolahan data, diperoleh 4.988 ulasan valid yang terdiri dari 2.823 ulasan negatif dan 2.165 ulasan positif berdasarkan status rekomendasi *voted_up*. Teks ulasan direpresentasikan menggunakan TF-IDF *N-Gram* berupa unigram dan bigram, kemudian diklasifikasikan menggunakan SVM dan *Multinomial Naïve Bayes*. Hasil evaluasi menunjukkan bahwa SVM memperoleh *accuracy* 82,26% dan *weighted F1-score* 0,8230, sedangkan *Multinomial Naïve Bayes* memperoleh *accuracy* 77,15% dan *weighted F1-score* 0,7598. Uji McNemar menunjukkan perbedaan performa yang signifikan dengan nilai *p-value* 0,000421. Kontribusi penelitian ini terletak pada evaluasi komparatif berbasis metrik per kelas, *confusion matrix*, analisis kesalahan prediksi, dan uji signifikansi, sehingga perbandingan model tidak hanya dilihat dari akurasi, tetapi juga dari keseimbangan dan signifikansi perbedaan performanya.

Kata Kunci: Analisis Sentimen, EA FC 26, Steam, Support Vector Machine, Multinomial Naïve Bayes, TF-IDF N-Gram

Abstract− User reviews on the *Steam* platform contain important information about player perceptions of *EA FC 26*. However, the large volume of reviews and their unstructured textual form make manual analysis difficult and prone to inconsistency. This study aims to identify user sentiment tendencies and compare the performance of *Support Vector Machine* (SVM) and *Multinomial Naïve Bayes* in classifying *EA FC 26* reviews on *Steam*. The research follows the *Knowledge Discovery in Database* (KDD) stages, consisting of *selection*, *preprocessing*, *transformation*, *data mining*, and *evaluation*. A total of 5,000 recent English-language reviews were collected through the *Steam Review API*. After data processing, 4,988 valid reviews were obtained, consisting of 2,823 negative reviews and 2,165 positive reviews based on the *voted_up* recommendation status. The review texts were represented using TF-IDF *N-Gram* features consisting of unigrams and bigrams, then classified using SVM and *Multinomial Naïve Bayes*. The evaluation results show that SVM achieved an *accuracy* of 82.26% and a *weighted F1-score* of 0.8230, while *Multinomial Naïve Bayes* achieved an *accuracy* of 77.15% and a *weighted F1-score* of 0.7598. The McNemar test produced a *p-value* of 0.000421, indicating a statistically significant difference between the two models. This study contributes by providing a comparative evaluation based not only on general accuracy, but also on per-class performance, *confusion matrix*, prediction error patterns, and statistical significance.

Keywords: Sentiment Analysis, EA FC 26, Steam, Support Vector Machine, Multinomial Naive Bayes, TF-IDF N-Gram

1. PENDAHULUAN

Perkembangan industri *game* digital membuat ulasan pengguna menjadi salah satu sumber informasi penting untuk menilai kualitas dan penerimaan sebuah produk. Melalui platform distribusi digital seperti *Steam*, pemain dapat memberikan komentar, kritik, dan rekomendasi berdasarkan pengalaman bermain. Ulasan tersebut dapat menggambarkan persepsi pemain terhadap *gameplay*, performa, fitur, harga, pembaruan, maupun kendala teknis yang dialami. Penelitian pada ulasan *indie video game* lokal di *Steam* menunjukkan bahwa ulasan pengguna dapat dimanfaatkan untuk mengetahui kecenderungan sentimen, dengan sentimen positif sebesar 69,8% dan akurasi klasifikasi 75,45% [1]. Penelitian lain pada ulasan *game online* di *Steam* juga memperoleh akurasi 80,97% menggunakan algoritma *Naïve Bayes* [2].

Ulasan pengguna pada platform digital umumnya berbentuk teks tidak terstruktur, sehingga sulit dianalisis secara manual apabila jumlah datanya besar. Pengguna dapat menulis ulasan dengan bahasa bebas, kalimat singkat, istilah teknis, simbol, ekspresi subjektif, bahkan pujian dan keluhan dalam satu teks yang sama. Dalam konteks *Steam*, ulasan tidak hanya berisi teks, tetapi juga status rekomendasi seperti *Recommended* dan *Not Recommended*. Status tersebut dapat digunakan sebagai indikator sentimen, tetapi tetap perlu ditafsirkan secara hati-hati karena label rekomendasi tidak selalu sepenuhnya mewakili isi opini pengguna. Studi terhadap 35.983.481 ulasan *Steam* menunjukkan bahwa karakteristik ulasan positif dan negatif dapat berbeda dari sisi panjang ulasan, waktu penulisan, serta variasi emosi antar-genre *game* [3]. Selain itu, pemanfaatan NLP pada ulasan pemain di *Steam* menunjukkan bahwa ulasan pengguna dapat dianalisis secara skalabel untuk memahami pengalaman dan respons pemain pada platform digital [4].Hal ini menunjukkan bahwa analisis sentimen ulasan *game* tidak cukup hanya melihat jumlah sentimen positif dan negatif, tetapi juga perlu memperhatikan karakter data, konteks platform, dan pola kesalahan model.

Beberapa penelitian terdahulu menunjukkan bahwa metode klasifikasi teks masih banyak digunakan dalam analisis sentimen ulasan *game* dan aplikasi. Pendekatan *Naïve Bayes* dan TF-IDF pada ulasan aplikasi *Steam* memperoleh akurasi rata-rata 84% [5], sedangkan *Multinomial Naïve Bayes* berbasis TF-IDF pada ulasan *Stardew Valley* memperoleh akurasi 0,8151 pada data *Steam* dan 0,8382 pada data Google Play [6]. Penelitian pada ulasan *League of Legends: Wild Rift* juga menunjukkan bahwa pemilihan fitur dan tahapan *preprocessing* memengaruhi performa *Naïve Bayes* [7]. sementara

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

penelitian pada ulasan *Free Fire* menekankan pentingnya penggunaan beberapa metrik evaluasi ketika data memiliki distribusi kelas yang tidak seimbang [8]. Selain itu, SVM banyak digunakan karena mampu bekerja pada data teks berdimensi tinggi. Penelitian pada ulasan *Genshin Impact*, *TheoTown*, *Counter-Strike 2*, dan *GTA V* menunjukkan bahwa SVM dengan TF-IDF dapat menghasilkan performa kompetitif pada berbagai objek ulasan *game* [9],[10],[11],[12]. Selain itu, penerapan metode *Knowledge Discovery in Database* (KDD), TF-IDF *N-Gram*, dan SVM pada klasifikasi teks memperoleh hasil terbaik berupa akurasi 70%, presisi 75%, *recall* 69%, dan *F-measure* 71% [13].

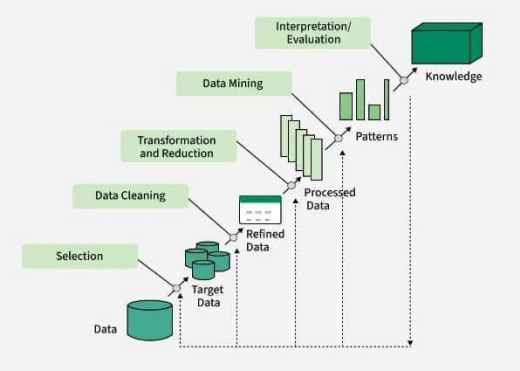
Perbandingan antara SVM dan *Naïve Bayes* telah dilakukan pada beberapa penelitian. Pada ulasan *PUBG Mobile*, SVM memperoleh akurasi 70,95%, sedangkan *Naïve Bayes* memperoleh akurasi 69,83% [14]. Penelitian pada sentimen pengguna *metaverse* juga menunjukkan SVM lebih unggul dibandingkan *Naïve Bayes* berdasarkan akurasi dan *Macro-F1* [15]. Penelitian internasional lain menunjukkan bahwa kedua algoritma masih relevan untuk klasifikasi teks berskala besar [16]. Pada ulasan *game* dari Metacritic dan *Steam*, SVC menjadi model terbaik dibandingkan Logistic Regression, *Multinomial Naïve Bayes*, MLP, dan XGBoost [17]. Hasil tersebut menunjukkan bahwa SVM cenderung kompetitif, tetapi performa model tetap dipengaruhi oleh objek, platform, skema label, fitur, dan proses pengolahan data .

Pada domain *game* sepak bola digital, penelitian analisis sentimen masih relatif terbatas. Penelitian terhadap 1.500 ulasan *eFootball 2024* menggunakan *Naïve Bayes* memperoleh akurasi 85%, presisi 85%, *recall* 86%, dan *F1-score* 85% [18]. Namun, penelitian yang secara khusus membahas ulasan *EA FC 26* di *Steam* masih terbatas, padahal objek ini memiliki karakter ulasan yang khas karena pengguna tidak hanya menilai kepuasan bermain, tetapi juga menyoroti mekanik permainan, performa teknis, pengalaman pertandingan, waktu pemuatan, pembaruan, serta perbandingan dengan seri FIFA atau *EA FC* sebelumnya. Celah penelitian ini juga terletak pada cara model dievaluasi pada ulasan *Steam* yang menggunakan label rekomendasi sebagai proksi sentimen. Label voted_up dapat membentuk kelas positif dan negatif, tetapi bukan anotasi langsung terhadap makna setiap kalimat, sehingga ulasan yang direkomendasikan tetap dapat memuat kritik dan ulasan tidak direkomendasikan masih mungkin mengandung aspek positif. Kondisi ini membuat evaluasi berbasis *accuracy* saja kurang memadai karena dapat menutupi kecenderungan model yang lebih kuat mengenali salah satu kelas. Selain itu, TF-IDF *N-Gram* digunakan karena ulasan *game* sering singkat dan informal, sehingga unigram dan bigram diperlukan untuk menangkap kata utama serta frasa negasi seperti *not good*, *not worth*, atau *not recommended* yang dapat mengubah arah sentimen [19].

Berdasarkan uraian tersebut, penelitian ini menggunakan tahapan *Knowledge Discovery in Database* (KDD) sebagai kerangka pengolahan data yang sistematis, mulai dari *selection*, *preprocessing*, *transformation*, *data mining*, hingga *evaluation* [20]. Data yang digunakan berupa ulasan pengguna *EA FC 26* berbahasa Inggris pada platform *Steam*, dengan label sentimen yang ditentukan berdasarkan status rekomendasi voted_up. Teks ulasan direpresentasikan menggunakan TF-IDF *N-Gram* berupa unigram dan bigram, kemudian diklasifikasikan dengan membandingkan SVM dan *Multinomial Naïve Bayes* pada data serta fitur yang sama. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, analisis kesalahan prediksi, dan uji McNemar agar perbandingan model tidak hanya bertumpu pada akurasi, tetapi juga mempertimbangkan performa per kelas, pola kesalahan, dan signifikansi perbedaan hasil prediksi. Dengan demikian, kebaruan penelitian ini terletak pada evaluasi komparatif sentimen ulasan *EA FC 26* di *Steam* yang menilai stabilitas model secara lebih menyeluruh dalam menentukan algoritma yang lebih sesuai untuk klasifikasi sentimen pengguna.

2. METODOLOGI PENELITIAN

Bagian ini menjelaskan tahapan penelitian yang dilakukan untuk menganalisis sentimen ulasan pengguna *EA FC 26* pada platform *Steam*. Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD) karena alurnya sesuai untuk proses pengolahan data teks, mulai dari pemilihan data, pembersihan data, transformasi, pemodelan, hingga evaluasi. Alur KDD yang digunakan dalam penelitian ini mengacu pada tahapan KDD yang terdiri dari proses pemilihan data, pembersihan data, transformasi, *data mining*, dan evaluasi [20]. Tahapan penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1. Tahapan *Knowledge Discovery in Database* (KDD)

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Berdasarkan Gambar 1, tahapan penelitian dimulai dari pengumpulan ulasan pengguna *EA FC 26* pada *Steam*, kemudian dilanjutkan dengan *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Metode KDD dipilih karena mampu menyusun proses analisis data secara sistematis, terutama pada penelitian klasifikasi teks yang membutuhkan tahapan pembersihan data, pembobotan kata, dan pemodelan algoritma [13].

2.1 Dataset

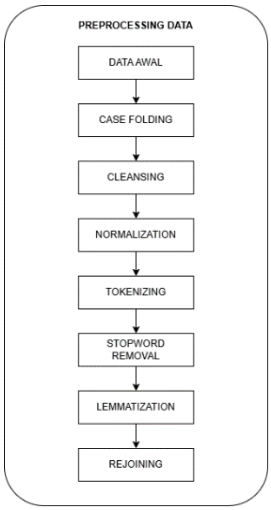
Dataset penelitian terdiri dari 5.000 ulasan pengguna *EA FC 26* yang dikumpulkan melalui *Steam Review API* menggunakan Python pada Google Colab. Jumlah tersebut ditetapkan sebagai batas operasional agar ruang lingkup analisis terkontrol dan kedua model diuji menggunakan data yang sama. Parameter *language=english* digunakan untuk memperoleh ulasan berbahasa Inggris, sedangkan *filter=recent* digunakan untuk mengambil ulasan terbaru. Berdasarkan atribut *timestamp_created*, data berada pada rentang 11 Maret-14 Juni 2026. Atribut yang digunakan meliputi teks ulasan dan status rekomendasi *voted_up*, dengan nilai *true* sebagai sentimen positif dan *false* sebagai sentimen negatif. Pemanfaatan ulasan *Steam* sebagai sumber data juga diterapkan pada penelitian Counter-Strike 2 [11], GTA V [12], dan aplikasi *Steam* [5].

2.2 Data Selection

Tahap *selection* dilakukan terhadap 5.000 ulasan *EA FC 26* hasil *scraping* dari *Steam*. Atribut yang digunakan adalah teks ulasan dan status rekomendasi *voted_up*, dengan nilai *true* sebagai sentimen positif dan *false* sebagai sentimen negatif. Data duplikat, ulasan tanpa teks, serta data yang tidak menghasilkan teks bersih setelah pengolahan tidak digunakan. Hasil seleksi menghasilkan 4.988 ulasan valid yang terdiri dari 2.823 ulasan negatif dan 2.165 ulasan positif.

2.3 Preprocessing Data

Tahap *preprocessing* dilakukan menggunakan Python dengan pustaka re untuk pembersihan pola teks serta NLTK untuk *stopword removal* dan *lemmatization* menggunakan WordNetLemmatizer. Prosesnya meliputi *case folding*, *normalization*, *cleansing*, *tokenizing*, *stopword removal*, *lemmatization*, dan *rejoining* hingga terbentuk kolom *clean review*. Kata negasi seperti *not*, *no*, dan *never* tetap dipertahankan karena dapat mengubah makna sentimen dan mendukung pembentukan fitur bigram pada tahap TF-IDF *N-Gram*. Tahapan serupa juga digunakan pada penelitian analisis sentimen ulasan *game* berbasis TF-IDF [11], [12]. Alur tahapan *preprocessing* data ditunjukkan pada Gambar 2.



Gambar 2. Tahapan *Preprocessing* Data

2.4 Ekstraksi Fitur TF-IDF N-Gram

Tahap *transformation* dilakukan dengan mengubah teks hasil *preprocessing* menjadi fitur numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Ekstraksi fitur menggunakan TfidfVectorizer dari Scikit-learn dengan konfigurasi ngram_range=(1,2), max_features=5000, min_df=2, max_df=0.90, dan sublinear_tf=True. Unigram merepresentasikan kata tunggal, sedangkan bigram memungkinkan model menangkap pasangan kata, termasuk frasa negasi seperti *not good* dan *not worth*. Penggunaan TF-IDF *N-Gram* juga relevan dengan penelitian klasifikasi teks berbasis SVM [11]. Bentuk umum pembobotan TF-IDF ditunjukkan pada Persamaan (1) dan Persamaan (2.

$$TF - IDF(t,d) = TF(td) \times IDF(t) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

TF(t,d) menyatakan frekuensi kata (t) pada dokumen (d), (N) merupakan jumlah seluruh dokumen, dan df(t) adalah jumlah dokumen yang mengandung kata (t). Persamaan tersebut digunakan untuk menjelaskan konsep umum TF-IDF. Dalam implementasinya, parameter sublinear_tf=True menerapkan transformasi logaritmik terhadap frekuensi kata, sedangkan perhitungan IDF mengikuti mekanisme bawaan TfidfVectorizer. Hasil tahap ini berupa matriks fitur numerik yang digunakan sebagai masukan bagi SVM dan *Multinomial Naïve Bayes*.

2.5 Pemodelan Klasifikasi

Tahap *data mining* dilakukan dengan membandingkan dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Multinomial Naïve Bayes*. SVM digunakan karena mampu bekerja baik pada data teks berdimensi tinggi yang dihasilkan dari pembobotan TF-IDF. Pada penelitian ini, SVM diimplementasikan menggunakan *LinearSVC* dari pustaka *Scikit-learn* dengan pendekatan linear, parameter C=1.0, dan *random_state*=42. Model *Multinomial Naïve Bayes* diimplementasikan menggunakan MultinomialNB dengan parameter alpha=1.0 sebagai nilai *smoothing* bawaan. Parameter tersebut digunakan tanpa proses *hyperparameter tuning* karena penelitian ini berfokus pada perbandingan performa dasar kedua algoritma terhadap fitur TF-IDF *N-Gram*. Secara umum, SVM bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas sentimen positif dan negatif. Fungsi keputusan SVM ditunjukkan pada Persamaan (3).

$$f(x) = sign(w^T x + b) \tag{3}$$

Pada Persamaan (3), x merupakan vektor fitur, w merupakan bobot, dan b merupakan bias. Sementara itu, *Multinomial Naïve Bayes* digunakan sebagai pembanding karena sederhana, efisien, dan umum digunakan dalam klasifikasi teks. Perbandingan SVM dan *Naïve Bayes* pada penelitian sebelumnya menunjukkan bahwa performa kedua algoritma dapat berbeda tergantung karakteristik *dataset*, proses *preprocessing*, dan fitur yang digunakan [14], [15], [16]. Rumus probabilitas *Multinomial Naïve Bayes* ditunjukkan pada Persamaan (4).

$$P(c|d) \propto P(c) \times \prod_{i=1}^{n} P(t_i|c)^{x_i} \tag{4}$$

Pada Persamaan (4), P(c|d) merupakan probabilitas dokumen d terhadap kelas c, P(c) merupakan probabilitas awal kelas, $P(t_i|c)$ merupakan probabilitas kata atau fitur ke-i pada kelas c, dan x_i merupakan bobot fitur dalam dokumen. Kelas dengan probabilitas tertinggi dipilih sebagai hasil prediksi sentimen. Kedua model dilatih menggunakan representasi fitur dan pembagian data yang sama agar perbandingan performa dilakukan secara setara. SVM diimplementasikan menggunakan LinearSVC dengan parameter C=1.0 dan random_state=42, sedangkan *Multinomial Naïve Bayes* menggunakan MultinomialNB dengan parameter alpha=1.0.

2.6 Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan SVM dan *Multinomial Naïve Bayes* dalam mengklasifikasikan ulasan positif dan negatif. Data dibagi dengan rasio 80:20 menggunakan train_test_split dengan test_size=0.20, random_state=42, dan stratify=y agar proporsi kelas tetap terjaga pada data latih dan data uji. Pembentukan kosakata TF-IDF dilakukan hanya pada data latih melalui fit_transform, sedangkan data uji diproses menggunakan transform untuk mencegah kebocoran informasi. Performa model dievaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*, kemudian diperkuat dengan analisis kesalahan prediksi serta uji McNemar untuk menilai signifikansi perbedaan prediksi pada data uji yang sama [21]. Uji McNemar digunakan karena kedua model diuji pada data yang sama, sehingga perbedaan hasil prediksi dapat dinilai secara statistik dengan batas signifikansi 0,05. Penggunaan beberapa metrik diperlukan agar performa model tidak hanya dilihat dari jumlah prediksi benar secara keseluruhan, tetapi juga dari kemampuannya mengenali masing-masing kelas sentimen [5], [14], [12], [8]. Rumus evaluasi ditunjukkan pada Persamaan (5) sampai Persamaan (8).

$$Accuracy = \frac{(TP) + (TN)}{(TP + TN + FP + FN)} \tag{5}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{7}$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{8}$$

Pada persamaan evaluasi, TP dan TN menunjukkan prediksi benar pada kelas positif dan negatif, sedangkan FP dan FN menunjukkan kesalahan prediksi pada masing-masing kelas. Hasil evaluasi kedua model kemudian dibandingkan untuk menentukan performa SVM dan *Multinomial Naïve Bayes*. Untuk mengetahui apakah perbedaan prediksi kedua model signifikan secara statistik, digunakan uji McNemar dengan koreksi kontinuitas sebagaimana ditunjukkan pada Persamaan (9).

$$x^2 = \times \frac{(|b - c| - 1)^2}{b + c} \tag{9}$$

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Nilai b menunjukkan prediksi benar oleh SVM tetapi salah oleh *Multinomial Naïve Bayes*, sedangkan c menunjukkan prediksi salah oleh SVM tetapi benar oleh *Multinomial Naïve Bayes*. Hasil dinyatakan signifikan jika *p-value* < 0,05.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Seleksi Data

Data penelitian diperoleh melalui proses scraping ulasan pengguna EA FC 26 pada platform Steam. Jumlah data awal yang berhasil dikumpulkan sebanyak 5.000 ulasan. Data tersebut kemudian melalui proses seleksi untuk memastikan bahwa ulasan yang digunakan sesuai dengan kebutuhan penelitian. Data yang tidak memenuhi kriteria, seperti data duplikat, tidak berbahasa Inggris, atau tidak memiliki teks bersih setelah proses pengolahan, tidak digunakan pada tahap berikutnya. Hasil seleksi data ditampilkan pada Tabel 1.

Tabel 1. Hasil seleksi Data

No	Tahapan Data	Jumlah Data
1	Data awal hasil scraping	5000
2	Data tidak digunakan setelah proses pengolahan	12
3	Data valid akhir	4988

Berdasarkan Tabel 1, dari 5.000 ulasan awal diperoleh 4.988 data valid, sedangkan 12 ulasan tidak digunakan. Sebagian besar data hasil *scraping* dapat dipertahankan, sehingga data yang diperoleh melalui *Steam Review API* telah sesuai untuk proses klasifikasi. Data valid tersebut kemudian dikelompokkan menjadi sentimen positif dan negatif berdasarkan status rekomendasi pengguna, seperti ditampilkan pada Tabel 2.

Tabel 2. Distribusi Sentimen Data Valid

Sentimen	Jumlah Data	Persentase
Negatif	2823	56,60%
Positif	2165	43,40%
Total	4988	100%

Berdasarkan Tabel 2, terdapat 2.823 ulasan berstatus tidak direkomendasikan atau 56,60% dan 2.165 ulasan berstatus direkomendasikan atau 43,40%. Distribusi tersebut menunjukkan bahwa kelas negatif sedikit lebih dominan, meskipun perbedaannya tidak tergolong ekstrem. Kondisi ini perlu diperhatikan karena dapat memengaruhi kecenderungan model dalam mengenali masing-masing kelas. Oleh karena itu, evaluasi model tidak cukup hanya dilihat dari *accuracy*, tetapi juga perlu dianalisis melalui *precision*, *recall*, *F1-score*, dan *confusion matrix*.

3.2 Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk mengubah teks ulasan mentah menjadi data yang lebih bersih dan siap diproses oleh algoritma klasifikasi. Proses ini meliputi *case folding*, *normalization*, *cleansing*, *tokenizing*, *stopword removal*, *lemmatization*, dan *rejoining*. Hasil akhir dari proses ini berupa teks bersih pada kolom clean review. Contoh perubahan teks sebelum dan sesudah preprocessing ditampilkan pada Tabel 3.

Tabel 3. Contoh Hasil *Preprocessing* Data

No	Ulasan Awal	Hasil Preprocessing	Sentimen
1	The game keeps on stuttering on what is a highly specifified gaming laptop, i have adjusted the graphic settings with no joy.	game keep stuttering highly specifified gaming laptop adjusted graphic setting no joy	Negatif
2	Best football game I've ever played in my entire 30 years of life.	best football game ever played entire year life	Positif
3	Game keeps closing and crashing	game keep closing crashing	Negatif
4	The Graphics is good! But the game mechanics could have been better	graphic good game mechanic could better	Positif

Berdasarkan Tabel 3, tahap *preprocessing* mampu mengubah ulasan mentah menjadi teks yang lebih ringkas dan terstruktur dengan mengurangi huruf kapital, tanda baca, simbol, serta kata umum yang kurang relevan tanpa menghilangkan informasi penting. Kata negasi seperti *not*, *no*, dan *never* tetap dipertahankan karena dapat mengubah makna sentimen, misalnya *good* menjadi *not good*. Keputusan ini membantu model menangkap konteks ulasan dengan lebih baik, terutama melalui fitur bigram pada TF-IDF *N-Gram*. Hasil *preprocessing* selanjutnya divisualisasikan dalam bentuk *wordcloud* pada Gambar 3.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Gambar 3. *Wordcloud* Hasil *Preprocessing* Ulasan *EA FC 26*

3.3 Hasil Pembagian Data dan Ekstraksi TF-IDF N-Gram

Setelah data valid diperoleh, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi data yang belum pernah dipelajari sebelumnya. Hasil pembagian data ditampilkan pada Tabel 4.

Tabel 4. Hasil Pembagian Data Latih dan Data Uji

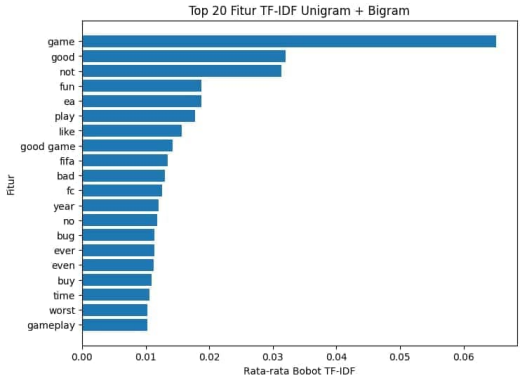
Jenis Data	Jumlah Data	Persentase
Data Latih	3990	79,99%
Data Uji	998	20,01%
Total	4988	100%

Berdasarkan Tabel 4, data dibagi menjadi 3.990 data latih dan 998 data uji dengan rasio 80:20. Pembagian ini digunakan agar model memiliki data yang cukup untuk proses pelatihan, tetapi tetap dapat diuji pada data yang belum pernah dipelajari sebelumnya. Pada data uji, jumlah ulasan negatif sebanyak 565 dan ulasan positif sebanyak 433, sehingga distribusi kelas pada data uji masih mengikuti proporsi data keseluruhan. Konfigurasi ekstraksi fitur yang digunakan ditampilkan pada Tabel 5.

Tabel 5. Konfigurasi Ekstraksi Fitur TF-IDF *N-Gram*

No	Komponen	Keterangan
1	Sumber fitur	Kolom clean review
2	Metode Pembobotan	TF-IDF
3	Pendekatan fitur	N-Gram
4	Jenis N-Gram	Unigram dan Bigram
5	Jumlah fitur maksimum	5000 fitur
6	Output	Matriks fitur numerik

Berdasarkan Tabel 5, ekstraksi fitur menggunakan TF-IDF *N-Gram* dengan kombinasi unigram dan bigram serta batas maksimum 5.000 fitur. Unigram digunakan untuk merepresentasikan kata tunggal, sedangkan bigram membantu menangkap pasangan kata yang memiliki makna sentimen lebih jelas. Pendekatan ini sesuai untuk ulasan *Steam* yang cenderung singkat dan informal, sehingga representasi teks menjadi lebih informatif sebelum diproses menggunakan SVM dan *Multinomial Naïve Bayes*. Visualisasi fitur TF-IDF dominan ditampilkan pada Gambar 4.



Gambar 4. Fitur TF-IDF N-Gram Dominan

Berdasarkan Gambar 4, fitur TF-IDF *N-Gram* dominan menunjukkan kata dan pasangan kata dengan bobot tinggi dalam merepresentasikan ulasan. Unigram menangkap kata utama, sedangkan bigram memberikan konteks yang lebih spesifik,

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

terutama pada frasa penilaian atau negasi seperti *not good* dan *not worth*. Dengan demikian, TF-IDF *N-Gram* menghasilkan representasi teks yang lebih informatif sebelum diproses menggunakan SVM dan *Multinomial Naïve Bayes*.

3.4 Hasil Evaluasi Model SVM dan Multinomial Naïve Bayes

Setelah fitur numerik diperoleh, data latih digunakan untuk membangun model SVM dan Multinomial Naïve Bayes. Kedua model kemudian diuji menggunakan data uji sebanyak 998 ulasan. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score. Ringkasan hasil evaluasi kedua model ditampilkan pada Tabel 6.

Tabel 6. Perbandingan Hasil Evaluasi Model

Model	Accuracy	Precision Weighted	Recall Weighted	F1-Score Weighted
Support Vector Machine	0,8226	0,8237	0,8226	0,8230
Multinomial Naïve Bayes	0,7715	0,7945	0,7715	0,7598

Berdasarkan Tabel 6, SVM memperoleh *accuracy* sebesar 0,8226 atau 82,26%, sedangkan *Multinomial Naïve Bayes* memperoleh *accuracy* sebesar 0,7715 atau 77,15%. Selisih akurasi sebesar 5,11% menunjukkan bahwa SVM lebih mampu memisahkan pola sentimen pada data ulasan *EA FC 26* dengan fitur TF-IDF *N-Gram*. Selain itu, nilai *F1-score weighted* SVM sebesar 0,8230 juga lebih tinggi dibandingkan *Multinomial Naïve Bayes* sebesar 0,7598. Hal ini menunjukkan bahwa keunggulan SVM tidak hanya terlihat dari akurasi, tetapi juga dari keseimbangan performa secara keseluruhan. Untuk melihat performa model secara lebih detail, hasil evaluasi per kelas ditampilkan pada Tabel 7.

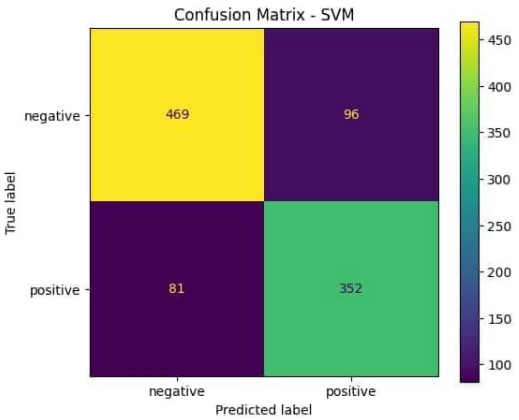
Tabel 7. Hasil Evaluasi Per Kelas

Model	Kelas	Precision	Recall	F1-Score	Support
Support Vector Machine	Negatif	0,8527	0,8301	0,8413	565
Support Vector Machine	Positif	0,7857	0,8129	0,7991	433
Multinomial Naïve Bayes	Negatif	0,7324	0,9398	0,8233	565
Multinomial Naïve Bayes	Positif	0,8755	0,5520	0,6771	433

Berdasarkan Tabel 7, SVM menunjukkan performa yang lebih seimbang pada kedua kelas dengan *F1-score* sebesar 0,8413 untuk sentimen negatif dan 0,7991 untuk sentimen positif. Sebaliknya, *Multinomial Naïve Bayes* memiliki *recall* sangat tinggi pada kelas negatif sebesar 0,9398, tetapi hanya 0,5520 pada kelas positif. Hal ini menunjukkan bahwa model lebih mampu mengenali ulasan negatif, tetapi masih sering mengklasifikasikan ulasan positif sebagai negatif. Kondisi tersebut kemungkinan dipengaruhi oleh dominasi kelas negatif dan adanya ulasan positif yang tetap memuat kata bernuansa kritik. Oleh karena itu, evaluasi per kelas diperlukan karena nilai akurasi saja belum cukup untuk menunjukkan kecenderungan kesalahan model.

3.5 Analisis Confusion Matrix

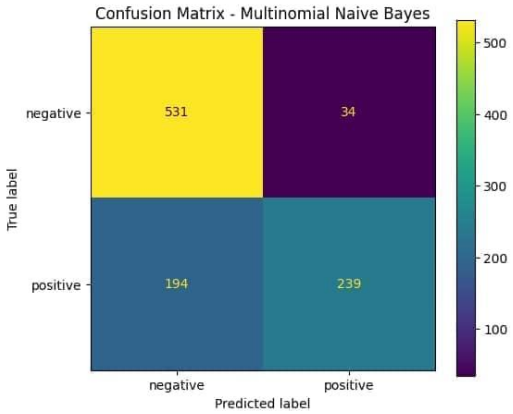
Selain menggunakan metrik evaluasi, performa model juga dianalisis melalui *confusion matrix*. *Confusion matrix* digunakan untuk melihat jumlah prediksi benar dan salah pada setiap kelas. Hasil *confusion matrix* SVM ditampilkan pada Gambar 5.



Gambar 5. *Confusion Matrix* Model SVM

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Berdasarkan Gambar 5, SVM berhasil mengklasifikasikan 469 ulasan negatif dan 352 ulasan positif dengan benar. Kesalahan prediksi pada SVM terdiri dari 96 ulasan negatif yang diprediksi sebagai positif dan 81 ulasan positif yang diprediksi sebagai negatif. Jumlah kesalahan pada kedua kelas relatif seimbang, sehingga memperkuat hasil evaluasi sebelumnya bahwa SVM memiliki kemampuan yang cukup stabil dalam mengenali sentimen negatif maupun positif.
Hasil *confusion matrix* untuk model *Multinomial Naïve Bayes* ditampilkan pada Gambar 6.



Gambar 6. *Confusion Matrix* Model Multinomial Naïve Bayes

Berdasarkan Gambar 6, *Multinomial Naïve Bayes* berhasil mengenali 531 ulasan negatif dengan benar, tetapi hanya mampu mengenali 239 ulasan positif dengan benar. Kesalahan terbesar terjadi pada 194 ulasan positif yang diprediksi sebagai negatif. Pola ini menunjukkan bahwa *Multinomial Naïve Bayes* cenderung lebih kuat dalam mengenali pola ulasan negatif, tetapi belum mampu menangkap karakter ulasan positif secara optimal. Temuan ini sejalan dengan nilai *recall* positif yang rendah pada Tabel 7.

3.6 Pembahasan Perbandingan Model

Berdasarkan hasil evaluasi, SVM menunjukkan performa terbaik karena memperoleh nilai *accuracy*, *precision weighted*, *recall weighted*, dan *F1-score weighted* yang lebih tinggi dibandingkan *Multinomial Naïve Bayes*. Hasil ini menunjukkan bahwa SVM lebih sesuai digunakan pada data ulasan *EA FC 26* yang direpresentasikan menggunakan TF-IDF *N-Gram*, terutama karena mampu bekerja stabil pada fitur teks yang berdimensi tinggi dan bersifat jarang. Sebaliknya, *Multinomial Naïve Bayes* masih kurang seimbang dalam mengenali kedua kelas, meskipun cukup baik pada ulasan negatif. Banyaknya ulasan positif yang diprediksi sebagai negatif kemungkinan dipengaruhi oleh dominasi kelas negatif sebesar 56,60%, karakter ulasan yang mengandung opini campuran, serta asumsi independensi antarfitur pada *Naïve Bayes* yang kurang mampu menangkap hubungan antar kata, terutama pada frasa negasi. Karena penelitian ini belum menerapkan SMOTE atau *class weighting*, pengaruh distribusi kelas masih menjadi keterbatasan. Dengan demikian, SVM lebih unggul pada dataset dan konfigurasi penelitian ini, sedangkan *Multinomial Naïve Bayes* tetap dapat digunakan sebagai model pembanding yang sederhana dan efisien.

3.7 Analisis Kesalahan Prediksi

Analisis kesalahan prediksi dilakukan untuk melihat pola kesalahan yang dihasilkan oleh model SVM dan Multinomial Naïve Bayes pada data uji. Analisis ini penting karena nilai accuracy saja belum cukup untuk menjelaskan kecenderungan model dalam mengenali masing-masing kelas sentimen. Ringkasan pola kesalahan prediksi ditampilkan pada Tabel 8.

Tabel 8. Ringkasan Pola Kesalahan Prediksi

Kategori Analisis	SVM	Multinomial Naïve Bayes	Interpretasi
Prediksi benar	821	770	SVM menghasilkan jumlah prediksi benar lebih tinggi.
Total kesalahan prediksi	177	228	*Multinomial Naïve Bayes* menghasilkan kesalahan lebih banyak.
Ulasan positif diprediksi negatif	81	194	Kesalahan dominan pada *Multinomial Naïve Bayes* terjadi pada kelas positif.
Kesalahan unik model	75	126	*Multinomial Naïve Bayes* memiliki lebih banyak kesalahan yang tidak terjadi pada SVM.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

Tabel 8 menunjukkan bahwa SVM menghasilkan 821 prediksi benar dari 998 data uji, lebih tinggi dibandingkan *Multinomial Naïve Bayes* yang menghasilkan 770 prediksi benar. Sebanyak 695 ulasan diklasifikasikan benar oleh kedua model, sedangkan 102 ulasan sama-sama salah diklasifikasikan. Pola kesalahan SVM relatif lebih seimbang, sementara *Multinomial Naïve Bayes* cenderung mengarahkan ulasan positif ke kelas negatif, terlihat dari 194 ulasan positif yang salah diprediksi sebagai negatif. Kondisi ini dapat dipengaruhi oleh ulasan yang singkat, mengandung opini campuran, serta penggunaan voted_up sebagai label rekomendasi yang tidak selalu merepresentasikan isi teks secara langsung. Selain itu, dominasi kelas negatif sebesar 56,60% dan belum diterapkannya SMOTE atau *class weighting* menjadi keterbatasan yang dapat memengaruhi kecenderungan prediksi model.

3.8 Uji Signifikansi McNemar

Uji McNemar dilakukan untuk mengetahui apakah perbedaan hasil prediksi SVM dan Multinomial Naïve Bayes signifikan secara statistik. Uji ini digunakan karena kedua model diuji pada data uji yang sama, sehingga perbandingan tidak hanya dilihat dari nilai accuracy dan F1-score, tetapi juga dari perbedaan pasangan prediksi benar dan salah. Hasil kontingensi prediksi kedua model ditampilkan pada Tabel 9.

Tabel 9. Tabel Kontingensi Uji McNemar

Kondisi Prediksi	MNB Benar	MNB Salah	Total
SVM benar	695	126	821
SVM Salah	75	102	177
Total	770	228	998

Berdasarkan Tabel 9, terdapat 126 data yang benar diklasifikasikan oleh SVM tetapi salah oleh *Multinomial Naïve Bayes*, sedangkan 75 data salah diklasifikasikan oleh SVM tetapi benar oleh *Multinomial Naïve Bayes*. Hasil uji McNemar dengan koreksi kontinuitas menghasilkan nilai x^2 sebesar 12,44 dengan *p-value* 0,000421. Karena nilai *p-value* lebih kecil dari 0,05, perbedaan performa kedua model dinyatakan signifikan secara statistik. Hasil ini menunjukkan bahwa keunggulan SVM tidak hanya terlihat dari metrik evaluasi umum, tetapi juga didukung oleh perbedaan pola prediksi yang signifikan pada data uji yang sama. Dengan demikian, SVM dapat dinyatakan lebih sesuai dibandingkan *Multinomial Naïve Bayes* untuk klasifikasi sentimen ulasan *EA FC 26* pada konfigurasi penelitian ini.

4. KESIMPULAN

Penelitian ini menganalisis sentimen ulasan pengguna *EA FC 26* di *Steam* menggunakan tahapan KDD, representasi TF-IDF *N-Gram*, serta perbandingan SVM dan *Multinomial Naïve Bayes*. Dari 5.000 ulasan berbahasa Inggris yang dikumpulkan melalui *Steam Review API*, diperoleh 4.988 data valid yang terdiri dari 2.823 ulasan negatif dan 2.165 ulasan positif berdasarkan status rekomendasi voted_up. Hasil evaluasi menunjukkan bahwa SVM memperoleh *accuracy* 82,26% dan *weighted F1-score* 0,8230, lebih tinggi dibandingkan *Multinomial Naïve Bayes* yang memperoleh *accuracy* 77,15% dan *weighted F1-score* 0,7598. Evaluasi per kelas dan *confusion matrix* menunjukkan bahwa SVM memiliki performa lebih seimbang, sedangkan *Multinomial Naïve Bayes* cenderung lebih kuat mengenali ulasan negatif tetapi masih salah mengklasifikasikan 194 ulasan positif sebagai negatif. Uji McNemar menghasilkan *p-value* 0,000421, sehingga perbedaan performa kedua model terbukti signifikan secara statistik. Dengan demikian, SVM lebih sesuai digunakan pada dataset dan konfigurasi penelitian ini karena tidak hanya unggul dari sisi akurasi, tetapi juga lebih stabil berdasarkan performa per kelas, pola kesalahan, dan uji signifikansi. Namun, penelitian ini masih terbatas pada ulasan *Steam* berbahasa Inggris dan penggunaan label voted_up sebagai proksi sentimen, sehingga ulasan dengan opini campuran berpotensi tidak sepenuhnya terwakili oleh labelnya. Selain itu, penelitian ini belum menerapkan *hyperparameter tuning*, *cross-validation*, serta penanganan ketidakseimbangan kelas. Penelitian selanjutnya dapat memperluas sumber data, menerapkan validasi silang, menguji teknik penyeimbangan kelas, serta membandingkan model lain untuk memperoleh hasil klasifikasi yang lebih kuat.

REFERENCES

[1] M. Y. Febrianta, S. Widiyanesti, and S. Robbiansyah, "Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *Journal of Animation & Games Studies*, vol. 7, no. 2, pp. 117–144, 2021, doi: 10.24821/jags.v7i2.5162.

[2] A. Pangestu *et al.*, "ANALISIS SENTIMEN REVIEW PUBLIK PENGGUNA GAME ONLINE PADA PLATFORM STEAM MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3106–3113, 2023, doi: 10.36040/jati.v7i6.8829.

[3] T. Guzsvinecz and J. Szűcs, "Length and sentiment analysis of reviews about top-level video game genres on the steam platform," *Computers in Human Behavior*, vol. 149, no. September, 2023, doi: 10.1016/j.chb.2023.107955.

[4] M. Edlinger, S. E. Huber, and M. Ninaus, "Analyzing player reviews with natural language processing to identify ecogames for education and research," *Computers in Human Behavior Reports*, vol. 20, no. October, p. 100850, 2025, doi: 10.1016/j.chbr.2025.100850.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 238 - 247
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2974
https://journal.fkpt.org/index.php/MIS

[5] A. Kho and F. F. Tampinongkol, "Analisis Sentimen Pengguna Aplikasi STEAM dengan Algoritma Naive Bayes," *Ranah Research : Journal Of Multidisciplinary Research and Development*, vol. 7, no. 6, pp. 4620–4627, 2025, doi: 10.38035/rrj.v7i6.1803.

[6] S. V. Tampubolon and D. A. Sulistyo, "Analisis Sentimen Ulasan Game Stardew Valley pada Steam dan Google Play," *JURNAL FASILKOM*, vol. 16, no. 1, pp. 57–67, 2026, doi: 10.37859/jf.v16i1.11217.

[7] N. Sabilly, Khadijah, and F. A. Nugroho, "SENTIMENT ANALYSIS OF LEAGUE OF LEGENDS: WILD RIFT REVIEWS ON GOOGLE PLAY USING NAÏVE BAYES CLASSIFIER," *Jurnal Ilmiah KURSOR*, vol. 12, no. 1, pp. 23–30, 2023, doi: 10.21107/kursor.

[8] A. Bachtiar, M. J. Hawali, N. C. Simbolon, D. A. Maulana, and R. A. Anggraini, "ANALISIS SENTIMEN ULASAN FREE FIRE MENGGUNAKAN NAIVE BAYES LOGISTIC REGRESSION," *Journal of Information System Management (JOISM*, vol. 7, no. 2, 2026, doi: 10.24076/joism.2026v7i2.2433.

[9] M. F. Fahrudillah and D. A. Sulistyo, "Analisis Sentimen Ulasan Pemain Genshin Impact Menggunakan Kombinasi TF-IDF , Lexicon , dan Support Vector Machine," *JURNAL FASILKOM*, vol. 15, no. 3, pp. 580–593, 2025, doi: 10.37859/jf.v15i3.10553.

[10] S. Aura and D. Novianto, "Comparative Sentiment Analysis of TheoTown Reviews on Steam and Google Play Store Using Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. April, pp. 886–896, 2026, doi: 10.57152/malcom.v6i2.2661.

[11] S. F. Riyadi, Y. H. Chrisnanto, and G. Abdillah, "Analisis Sentimen Komunitas Counter-Strike 2 (CS2) Menggunakan Support Vector Machine (SVM)," *JURIKOM (Jurnal Riset Komputer)*, vol. 12, no. 3, 2025, doi: 10.30865/jurikom.v12i3.8620.

[12] A. K. Saputra, M. R. Handayani, N. Cahyo, H. Wibowo, and K. Umam, "Sentiment Analysis of User Reviews on the Game GTA V Using Support Vector Machine," *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, vol. 14, pp. 284–290, 2025, doi: 10.33884/jif.v13i01.9913.

[13] N. Arifin, U. Enri, and N. Sulistiyowati, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DENGAN TF-IDF N-GRAM UNTUK TEXT CLASSIFICATION," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, 2021, doi: 10.30998/string.v6i2.10133.

[14] P. R. Sari *et al.*, "Comparison of Naive Bayes and SVM Algorithms for Sentiment Analysis of PUBG Mobile on Google Play Store," *Sistemasi: Jurnal Sistem Informasi Volume*, vol. 13, pp. 2767–2779, 2024, doi: 10.32520/stmsi.v13i6.4814.

[15] S. D. Parameswari, M. Lubis, S. Suakanto, Y. Z. Ramadhan, N. Amanah, and R. A. Dila, "Studi Perbandingan Naï ve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Pengguna Metaverse," *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, vol. 4, no. 3, pp. 1059–1065, 2025, doi: 10.55826/jtmit.v4i3.1122.

[16] R. Karsi, M. Zaim, and J. El Alami, "Assessing naive bayes and support vector machine performance in sentiment classification on a big data platform," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, pp. 990–996, 2021, doi: 10.11591/ijai.v10.i4.pp990-996.

[17] J. Y. Tan and A. S. K. Chow, "Sentiment Analysis on Game Reviews: A Comparative Study of Machine Learning Approaches," *International Conference on Digital Transformation and Applications (ICDXA)*, no. October, pp. 209–216, 2021, doi: 10.56453/icdxa.2021.1023.

[18] R. Nur, R. Abdul, and W. J. Pranoto, "Analisis Sentimen Ulasan Game eFootball 2024 Pada Playstore menggunakan Algoritma Naïve Bayes," *JURNAL ILMIAH INFORMATIKA*, vol. 13, no. 15, 2025, doi: 10.33884/jif.v13i01.9913.

[19] Sutriawan, Muljono, Khairunnisa, Z. Alamin, T. A. Lorosae, and S. Ramadhan, "Improving Performance Sentiment Movie Review Classification Using Hybrid Feature TFIDF, N-Gram, Information Gain and Support Vector Machine," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 2, pp. 375–384, 2024, doi: 10.18280/mmep.110209.

[20] Muttaqin *et al.*, *Pengenalan Data Mining*. Yayasan Kita Menulis, 2023.

[21] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.