

Analisis Sentimen Ulasan *Game Stardew Valley* pada *Steam* dan *Google Play*

Surya Viari Tampubolon¹, Danang Arbian Sulisty²

¹Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis Asia Malang

suryaviari22@gmail.com, ²danangarbian@gmail.com *

Abstract

The large number of user reviews on Steam and Google Play platforms makes manual analysis difficult and prone to subjective bias. This study aims to analyze and compare user sentiment toward Stardew Valley game reviews on both platforms using a text mining approach. The data used consist of 25,099 Steam reviews and 25,594 Google Play reviews. The text preprocessing stage includes case folding, cleansing (removal of punctuation and non-alphabetic characters), tokenization, stopword removal, and lemmatization to produce more structured data. Sentiment labeling is performed using the VADER method, followed by feature extraction using TF-IDF and classification using the Multinomial "Naïve Bayes" algorithm. Model evaluation is conducted using 5-Fold Cross Validation with accuracy, precision, recall, and F1-score as evaluation metrics. The results show that most reviews on both platforms have positive sentiment. The classification model achieves an average accuracy of 0.8151 on Steam and 0.8382 on Google Play. In addition, the model obtains an average F1-score (macro average) of 0.55 on Steam and 0.40 on Google Play. These results indicate that the model performs adequately in sentiment classification, although it still has limitations in identifying minority sentiment classes such as negative and neutral.

Keywords: sentiment analysis, stardew valley, TF-IDF, naïve bayes, VADER.

Abstrak

Jumlah ulasan pengguna pada platform *Steam* dan *Google Play* yang sangat besar menyulitkan proses analisis secara manual serta berpotensi menghasilkan penilaian yang tidak objektif. Penelitian ini bertujuan untuk menganalisis dan membandingkan sentimen ulasan pengguna *game Stardew Valley* pada kedua platform tersebut menggunakan pendekatan *text mining*. Data yang digunakan terdiri dari 25.099 ulasan *Steam* dan 25.594 ulasan *Google Play*. Tahap pra-pemrosesan teks meliputi *case folding*, *cleansing* (penghapusan tanda baca dan karakter non-alfabet), tokenisasi, *stopword removal*, dan lemmatisasi untuk menghasilkan data yang lebih terstruktur. Selanjutnya, pelabelan sentimen dilakukan menggunakan metode VADER, ekstraksi fitur menggunakan TF-IDF, serta klasifikasi menggunakan algoritma Multinomial "Naïve Bayes". Evaluasi model dilakukan menggunakan metode *5-Fold Cross Validation* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa mayoritas ulasan pada kedua platform memiliki sentimen positif. Model klasifikasi memperoleh rata-rata akurasi sebesar 0,8151 pada *Steam* dan 0,8382 pada *Google Play*. Selain itu, diperoleh rata-rata *F1-score (macro average)* sebesar 0,55 pada *Steam* dan 0,40 pada *Google Play*. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan sentimen, meskipun masih memiliki keterbatasan dalam mengidentifikasi sentimen minoritas seperti negatif dan netral.

Kata kunci: analisis sentimen, stardew valley, TF-IDF, naïve bayes, VADER.

©This work is licensed under a Creative Commons Attribution -ShareAlike 4.0 International License

1. Pendahuluan

Industri game digital berkembang pesat dalam beberapa tahun terakhir, karena penggunaan perangkat digital semakin meningkat dan platform penjualan daring seperti *Steam* serta *Google Play* semakin populer. Platform tersebut bukan hanya menjadi tempat untuk mendistribusikan *game*, tetapi juga menyediakan cara bagi pengguna dan pengembang untuk berinteraksi melalui fitur ulasan. Ulasan yang diberikan oleh pengguna mencakup berbagai hal seperti pengalaman saat bermain, tingkat kepuasan mereka, serta kritik terhadap sebuah *game*, sehingga dapat digunakan sebagai informasi penting untuk menilai kualitas dan sejauh mana produk tersebut diterima oleh para pengguna.

Meningkatnya banyaknya ulasan membuat analisis manual tidak efektif dan bisa menyebabkan penilaian yang tidak objektif. Maka itu, diperlukan pendekatan

yang dalam penelitian ini, *text mining* dan *natural language processing* (NLP) digunakan untuk secara otomatis mengubah data tulisan yang tidak terstruktur. NLP adalah cabang dari kecerdasan buatan yang berfokus pada kemampuan komputer untuk memahami, memproses, dan menganalisis bahasa manusia, baik dalam bentuk kata maupun lisan. [1], [2]. Penelitian dalam bidang NLP menunjukkan bahwa teknik pemrosesan bahasa alami dapat digunakan untuk mengekstraksi informasi dari data teks dalam jumlah besar serta mendukung berbagai aplikasi pengolahan bahasa seperti klasifikasi teks, penerjemahan mesin, dan analisis sentimen [3], [4].

Banyak studi terdahulu telah mengimplementasikan analisis sentimen pada berbagai jenis data, seperti media sosial, produk digital, hingga permainan daring. Dalam konteks *game*, studi oleh Siregar et al [5]. membuktikan bahwa penggunaan TF-IDF yang

dikombinasikan dengan *Support Vector Machine* (SVM) mampu menghasilkan klasifikasi ulasan pemain yang akurat pada *platform Google Play*. Penelitian lain oleh Febrianta et al. [6] juga mengkaji ulasan *game* pada platform Steam untuk mengevaluasi opini pengguna. Dari sisi pendekatan metode, Khotimah et al [7] mengungkapkan bahwa algoritma Naïve Bayes memiliki kinerja yang efektif dalam mengolah data teks, sedangkan Aprinastya et al. [8] melakukan analisis perbandingan antara Naïve Bayes dan SVM dalam domain game dan menemukan adanya perbedaan performa tergantung pada karakteristik *dataset* yang digunakan. Meskipun demikian, sebagian besar penelitian tersebut masih terbatas pada satu platform dan belum mengkaji perbandingan sentimen lintas platform terhadap *game* yang sama

Salah satu cara yang sering digunakan dalam bidang NLP adalah analisis sentimen, yaitu metode untuk mengenali dan mengelompokkan pendapat dalam teks ke dalam tiga kategori, yaitu positif, negatif, atau netral [9], [10]. Cara ini memungkinkan pengambilan informasi dari sekumpulan data yang sangat banyak dilakukan secara lebih rapi dan dapat diukur.

Dalam proses klasifikasi sentimen, representasi fitur memiliki peran penting dalam menentukan performa model. Teknik TF-IDF banyak digunakan untuk Mengubah teks menjadi bentuk angka dengan melihat seberapa sering dan pentingnya sebuah kata dalam dokumen [11], [12]. Metode ini mampu meningkatkan kualitas representasi teks sehingga mempermudah model klasifikasi dalam mengidentifikasi pola kata yang berkaitan dengan kategori sentimen tertentu.

Penggunaan TF-IDF yang dipadukan dengan algoritma *machine learning* seperti Naïve Bayes banyak diterapkan dalam analisis sentimen karena mampu melakukan pemrosesan data teks secara efisien serta memiliki kinerja komputasi yang baik. [13]. Karena menggunakan pendekatan probabilistik yang sederhana namun tetap efektif, Algoritma Multinomial Naive Bayes berhasil dalam berbagai tugas klasifikasi teks.

Beberapa penelitian menunjukkan bahwa algoritma Naïve Bayes, khususnya varian Multinomial Naïve Bayes, memiliki keunggulan dalam klasifikasi teks karena mampu memodelkan distribusi frekuensi kata dalam dokumen secara efektif. Berbeda dengan Naïve Bayes Gaussian yang digunakan untuk data kontinu, Multinomial Naïve Bayes lebih sesuai untuk data diskrit seperti teks karena mempertimbangkan jumlah kemunculan kata dalam dokumen. Hal ini menjadikan metode ini lebih optimal ketika digunakan pada data yang telah direpresentasikan menggunakan TF-IDF. Studi yang dilakukan oleh Li et al. [14] juga menunjukkan bahwa algoritma berbasis probabilistik seperti Naïve Bayes memiliki performa yang baik dalam tugas klasifikasi teks dengan kompleksitas komputasi yang relatif rendah

Selain memilih algoritma untuk klasifikasi, proses persiapan kalimat sebelum dianalisa juga sangat

penting untuk meningkatkan hasil analisis sentimen. Langkah-langkah seperti tokenisasi, penghapusan kata tidak penting, dan normalisasi kata bertujuan untuk mengurangi gangguan pada data teks agar mendapatkan representasi data yang lebih baik sebelum proses klasifikasi dilakukan. Penelitian dalam bidang NLP juga menunjukkan bahwa teknik pemrosesan bahasa seperti stemming dan pengolahan morfologi bahasa dapat meningkatkan kualitas pemrosesan teks pada berbagai aplikasi pengolahan bahasa alami [15].

Selain itu, perkembangan penelitian NLP juga menunjukkan peningkatan penggunaan teknik pembelajaran mesin dan deep learning dalam berbagai tugas pemrosesan bahasa seperti penerjemahan mesin dan pengembangan korpus bahasa untuk bahasa dengan sumber daya terbatas [16], [17]. Hal ini membuktikan NLP memiliki tugas yang penting dalam mengembangkan berbagai metode analisis teks termasuk analisis sentimen.

Meskipun sudah banyak penelitian tentang analisis sentimen yang dilakukan di bidang media sosial dan kebijakan publik [18], Penelitian pada domain permainan digital masih relatif terbatas, khususnya yang membandingkan sentimen pengguna terhadap *game* yang sama pada platform yang berbeda. Padahal, perbedaan karakteristik pengguna, pola interaksi, serta konteks penggunaan antara platform PC dan *mobile* berpotensi menghasilkan persepsi yang berbeda terhadap suatu *game*. Selain itu, sebagian besar penelitian sebelumnya masih berfokus pada satu sumber data sehingga belum mampu memberikan gambaran komparatif secara menyeluruh. Oleh karena itu, diperlukan penelitian yang dapat menganalisis dan membandingkan sentimen pengguna pada berbagai platform untuk memperoleh pemahaman yang lebih komprehensif.

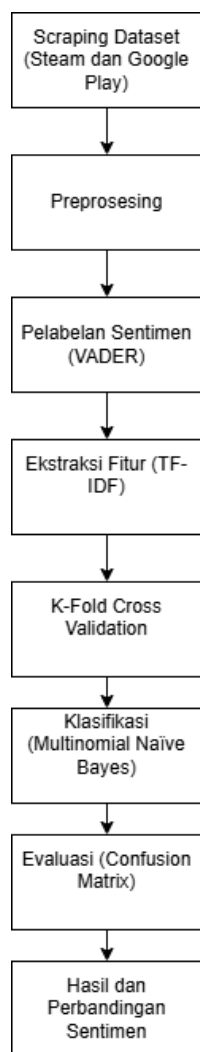
Stardew Valley dipilih sebagai objek penelitian karena merupakan salah satu *game* yang memiliki basis pemain yang besar dan aktif pada berbagai platform. Hal ini tercermin dari tingginya jumlah pemain aktif pada platform *Steam* yang mencapai puluhan ribu pengguna secara bersamaan, serta jumlah unduhan yang telah mencapai jutaan pada *platform Google Play*. Popularitas tersebut menunjukkan bahwa *game* ini memiliki komunitas yang luas dan menghasilkan volume ulasan pengguna yang besar, sehingga menyediakan data yang representatif untuk dianalisis dalam penelitian sentimen. Selain itu, *Stardew Valley* tersedia pada dua platform dengan karakteristik yang berbeda, yaitu PC dan *mobile*, yang memiliki perbedaan dalam mekanisme kontrol, pola penggunaan, serta kecepatan pembaruan fitur. Perbedaan ini berpotensi mempengaruhi pengalaman bermain dan persepsi pengguna, sehingga menjadikan *Stardew Valley* sebagai objek yang relevan untuk dianalisis dalam konteks perbandingan sentimen lintas platform.

Penelitian ini berfokus pada analisis sentimen pengguna terhadap ulasan game *Stardew Valley* pada platform *Steam* dan *Google Play* dengan menerapkan metode Multinomial Naïve Bayes berbasis TF-IDF. Evaluasi dilakukan menggunakan metode *5-Fold Cross Validation* untuk memastikan model memiliki performa yang stabil dan konsisten. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna serta membandingkan distribusi sentimen antara kedua platform, sehingga dapat memberikan gambaran perbedaan kecenderungan opini pengguna pada masing-masing platform.

2. Metode Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan memanfaatkan teknik *text mining* untuk melakukan analisis perasaan pengguna dalam ulasan game *Stardew Valley* di dua platform penjualan digital, yaitu *Steam* dan *Google Play*.

2.1. Alur Penelitian



Gambar 1. Alur Penelitian

Alur penelitian dalam studi ini ditampilkan pada Gambar 1. Tahapan penelitian diawali dengan pengumpulan data ulasan pengguna dari platform

Steam dan *Google Play*, kemudian dilanjutkan dengan proses *preprocessing*, pelabelan sentimen menggunakan metode VADER, serta ekstraksi fitur dengan TF-IDF. Selanjutnya, dilakukan klasifikasi menggunakan algoritma Multinomial Naïve Bayes dengan pendekatan validasi *5-Fold Cross Validation*. Hasil klasifikasi kemudian dievaluasi menggunakan *Confusion Matrix* untuk mengukur kinerja model, dan diakhiri dengan analisis perbandingan sentimen antara kedua platform.

2.2. Dataset

Dataset pada penelitian ini diperoleh melalui teknik web scraping terhadap ulasan pengguna game *Stardew Valley* dari dua platform, yaitu *Steam* dan *Google Play*. Pengambilan data dilakukan menggunakan bahasa pemrograman *Python* dengan memanfaatkan library *requests* untuk platform *Steam* dan *google-play-scraper* untuk platform *Google Play*.

Data yang diambil merupakan ulasan pengguna berbahasa Inggris dengan rentang waktu seluruh periode ulasan yang tersedia hingga saat pengambilan data dilakukan pada tahun 2024. Pada platform *Steam*, data yang diperoleh meliputi teks ulasan (*review*), status rekomendasi (*voted_up*), waktu pembuatan ulasan (*timestamp_created*), identitas pengguna (*steamid*), serta total waktu bermain (*playtime_forever*). Sementara itu, pada platform *Google Play*, data yang dikumpulkan mencakup nama pengguna (*userName*), skor penilaian (*score*), isi ulasan (*content*), dan waktu ulasan dibuat (*at*).

Jumlah data yang diperoleh pada penelitian ini adalah sebanyak 25.099 ulasan dari *Steam* dan 25.594 ulasan dari *Google Play*. Seluruh data kemudian disimpan dalam format CSV untuk mempermudah proses pengolahan dan analisis lebih lanjut.

2.3. Pra-pemrosesan

Pra-pemrosesan dilakukan agar ulasan menjadi lebih rapi dan mengurangi gangguan sebelum analisis sentimen dimulai. Tahapan ini bertujuan membuat representasi teks menjadi lebih sederhana dan jelas, sehingga bisa meningkatkan kualitas dalam proses pengelompokan [19].

Berdasarkan tahapan yang diterapkan, proses *preprocessing* mencakup beberapa langkah, yaitu *case folding*, *cleaning*, *tokenisasi*, *stopword removal*, dan *lemmatisasi*.

Tahap awal adalah *case folding*, yaitu mengubah seluruh teks ulasan menjadi huruf kecil guna menghindari perbedaan akibat penggunaan huruf kapital. Selanjutnya dilakukan proses *cleaning* dengan menghapus URL, angka, simbol, emoji, serta karakter non-alfabet menggunakan ekspresi reguler.

Tahap berikutnya adalah tokenisasi, yaitu proses memecah teks menjadi unit-unit kata. Setelah itu dilakukan *stopword removal* dengan memanfaatkan daftar *stopwords* bahasa Inggris dari pustaka NLTK

untuk menghilangkan kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis sentimen.

Tahap terakhir adalah lemmatisasi menggunakan *WordNetLemmatizer* untuk mengubah kata ke bentuk dasarnya sehingga dapat mengurangi variasi morfologis.

Contoh hasil dari setiap tahapan *preprocessing* akan disajikan pada Bab 3 bagian hasil dan pembahasan

2.4. Label Sentimen VADER

Pada penelitian ini, proses penentuan sentimen dilakukan menggunakan metode VADER. Metode ini termasuk dalam pendekatan berbasis leksikon yang dirancang untuk menganalisis teks berbahasa Inggris, terutama pada teks pendek seperti ulasan pengguna dan media sosial. VADER bekerja dengan memanfaatkan kamus leksikon yang berisi daftar kata beserta bobot sentimennya sehingga dapat menentukan polaritas sentimen suatu teks secara otomatis [20].

Setiap teks ulasan yang telah melalui tahap pra-pemrosesan dianalisis menggunakan fungsi *polarity_scores()* untuk memperoleh skor sentimen. Fungsi tersebut menghasilkan empat nilai skor: skor positif, skor negatif, skor neutral, dan skor campuran. Nilai campuran dinormalisasi menjadi nol hingga satu. Perhitungan nilai *compound score* pada metode VADER direpresentasikan pada rumus 1 sebagai berikut [21]:

$$Compound = \frac{\sum_{i=1}^n s_i}{\sqrt{\sum_{i=1}^n s_i^2 + \alpha}} \quad (1)$$

dengan s_i merupakan skor sentimen dari kata ke- i dalam teks, n menyatakan jumlah kata yang memiliki skor sentimen dalam dokumen, dan α merupakan konstanta normalisasi yang digunakan dalam proses perhitungan skor sentimen. Nilai konstanta α dalam metode VADER umumnya bernilai 15 dan berfungsi untuk menstabilkan hasil perhitungan serta mencegah nilai pembagi menjadi nol.

Pada rumus tersebut, bagian pembilang $\sum_{i=1}^n s_i$ sampai n) merepresentasikan jumlah total skor sentimen dari seluruh kata dalam teks. Nilai ini menggambarkan kecenderungan sentimen awal dari suatu ulasan. Semakin besar nilai penjumlahan skor sentimen, maka semakin kuat kecenderungan sentimen positif atau negatif yang terkandung dalam teks tersebut.

Sementara itu, bagian penyebut $\sqrt{(\sum_{i=1}^n s_i^2 + \alpha)}$ digunakan untuk melakukan proses normalisasi skor. Pada tahap ini setiap skor sentimen terlebih dahulu dikuadratkan untuk menghilangkan pengaruh tanda negatif serta memperhitungkan besarnya kekuatan sentimen dari setiap kata. Selanjutnya seluruh nilai kuadrat tersebut dijumlahkan dan ditambahkan dengan konstanta normalisasi α . Hasil penjumlahan tersebut kemudian diakarkan untuk menghasilkan nilai

pembagi yang berfungsi menstabilkan hasil akhir perhitungan sehingga nilai *compound score* tetap berada pada rentang yang terkontrol.

Pendekatan ini dipilih karena metode VADER mampu melakukan pelabelan sentimen secara otomatis dengan memanfaatkan kamus leksikon yang telah memiliki bobot sentimen pada setiap kata. Metode ini telah banyak digunakan dalam penelitian analisis sentimen teks dan terbukti efektif dalam mengidentifikasi polaritas sentimen, terutama pada data ulasan dan media sosial [21]. Proses pelabelan sentimen menggunakan metode VADER pada penelitian ini ditunjukkan pada pseudocode sebagai berikut :

```

Program Jurnal
Program VADER_Labeling
Input : cleaned_final
Output : sentiment_label

Load dataset
Initialize SentimentIntensityAnalyzer

for each review in cleaned_final do
score = polarity_scores(review)
compound = score["compound"]

if compound >= 0.05 then
sentiment = positive
else if compound <= -0.05 then
sentiment = negative
else
sentiment = neutral
end if

save sentiment
end for
    
```

Berdasarkan *pseudocode* di atas, proses pelabelan sentimen dimulai dengan memuat dataset yang telah melalui tahap *preprocessing*. Selanjutnya, dilakukan inialisasi *SentimentIntensityAnalyzer* dari metode VADER yang digunakan untuk menghitung skor sentimen pada setiap ulasan.

Setiap data ulasan diproses secara iteratif untuk memperoleh nilai *polarity scores*, khususnya nilai *compound* yang merepresentasikan keseluruhan sentimen dari teks. Nilai *compound* tersebut kemudian digunakan sebagai dasar dalam menentukan kategori sentimen.

Jika nilai *compound* lebih besar atau sama dengan 0,05 maka ulasan diklasifikasikan sebagai sentimen positif. Sebaliknya, jika nilai *compound* lebih kecil atau sama dengan -0,05 maka dikategorikan sebagai sentimen negatif. Sementara itu, jika nilai *compound* berada di antara -0,05 hingga 0,05 maka ulasan dikategorikan sebagai sentimen netral.

Hasil dari proses ini berupa label sentimen untuk setiap ulasan yang kemudian digunakan pada tahap klasifikasi selanjutnya

2.5. Ekstraksi Fitur TF-IDF

Selesaiannya tahap pelabelan sentimen, Proses mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi dikenal sebagai ekstraksi fitur. Teknik TF-IDF digunakan sebagai

teknik pembobotan kata dalam dokumen dalam penelitian ini. Metode ini banyak digunakan dalam analisis sentimen karena mampu memberikan bobot pada setiap kata sesuai tingkat kepentingannya dalam suatu dokumen dibandingkan dengan koleksi dokumen secara keseluruhan, sehingga meningkatkan kinerja klasifikasi teks [22] [23].

Nilai Frekuensi Kata (TF) menunjukkan berapa kali suatu kata muncul dalam dokumen. Semakin sering kata muncul, maka nilai TF yang dihasilkan juga lebih tinggi. Nilai TF dapat dihitung dengan Rumus 2.

$$TF(t, d) = \frac{f(t, d)}{\sum f(t, d)} \quad (2)$$

dengan $TF(t, d)$ menunjukkan nilai frekuensi kemunculan kata t pada dokumen d dan $f(t, d)$ menunjukkan jumlah kemunculan kata t pada dokumen d .

Selanjutnya, *Inverse Document Frequency* (IDF) digunakan untuk mengukur tingkat keunikan suatu kata dalam keseluruhan *dataset*. Kata yang muncul pada banyak dokumen akan memiliki nilai IDF yang lebih kecil, sedangkan kata yang jarang muncul akan memiliki nilai yang lebih tinggi. Rumus 3 menunjukkan perhitungan nilai IDF.

$$IDF(t) = \log \frac{N}{df(t)} \quad (3)$$

dengan N menyatakan semua dokumen dan $df(t)$ menunjukkan jumlah dokumen yang mengandung kata t .

Metode TF-IDF menghasilkan bobot akhir dari perkalian nilai TF dan IDF. Ini menghasilkan bobot kata yang menunjukkan tingkat kepentingan kata pada dokumen tertentu. Perhitungan bobot TF-IDF ditunjukkan pada Rumus 4.

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (4)$$

dengan $TFIDF(t, d)$ menyatakan bobot TF-IDF untuk kata t terhadap d . Nilai yang diperoleh selanjutnya dimanfaatkan sebagai representasi fitur dalam bentuk numerik yang digunakan sebagai input pada proses klasifikasi sentimen..

2.6 Klasifikasi Multinomial Naïve Bayes

Langkah berikutnya adalah proses mengklasifikasikan sentimen menggunakan algoritma *Multinomial Naïve Bayes*. Algoritma ini didasarkan pada Teorema Bayes yang mengasumsikan bahwa setiap fitur data bersifat independen satu sama lain. Metode *Multinomial Naïve Bayes* sering digunakan dalam klasifikasi teks karena memiliki kecepatan komputasi yang baik dan mampu menghasilkan kinerja yang cukup baik meskipun jumlah data yang digunakan relatif besar [14].

Probabilitas sebuah dokumen d untuk diklasifikasikan ke dalam kelas sentimen tertentu c dapat dihitung menggunakan Rumus 5 berikut.

$$P(c | d) = \frac{P(d|c) \times P(c)}{P(d)} \quad (5)$$

dengan $P(c|d)$ adalah kemungkinan bahwa dokumen d akan termasuk ke dalam kelas c , $P(d|c)$ adalah kemungkinan bahwa dokumen d akan muncul di kelas c , dan $P(c)$ adalah kemungkinan awal kelas c , sedangkan $P(d)$ menyatakan probabilitas kemunculan dokumen d .

Multinomial Naïve Bayes merupakan varian dari *Naïve Bayes* yang digunakan untuk data diskrit, khususnya data teks. Berbeda dengan *Naïve Bayes Gaussian* yang digunakan untuk data kontinu, *Multinomial Naïve Bayes* menghitung probabilitas berdasarkan frekuensi kemunculan kata dalam dokumen, sehingga lebih sesuai digunakan pada data teks yang telah direpresentasikan menggunakan TF-IDF.

Pada algoritma ini, dokumen akan diklasifikasikan ke dalam kelas yang memiliki nilai probabilitas terbesar. Dalam penelitian ini, kategori sentimen yang digunakan meliputi tiga kelas, yaitu positif, negatif, dan netral.

2.7. K-Fold Cross Validation

Pada penelitian ini digunakan metode *K-Fold Cross Validation* dengan nilai $k = 5$ untuk melakukan validasi terhadap model klasifikasi. Metode ini bertujuan untuk menguji kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah digunakan pada proses pelatihan.

Pada teknik *5-Fold Cross Validation*, *dataset* dibagi menjadi lima bagian dengan ukuran yang relatif sama. Pada setiap iterasi, empat bagian digunakan sebagai data pelatihan (*training data*), sedangkan satu bagian lainnya digunakan sebagai data pengujian (*testing data*). Proses ini dilakukan secara bergantian sebanyak lima kali hingga seluruh bagian data berperan sebagai data uji.

Dengan pendekatan ini, setiap data memiliki kesempatan yang sama untuk menjadi data pelatihan maupun data pengujian, sehingga hasil evaluasi yang diperoleh menjadi lebih stabil dan tidak bergantung pada satu kali pembagian data. Secara umum, pada setiap iterasi digunakan sekitar 80% data sebagai data pelatihan dan 20% sebagai data pengujian.

2.8 Evaluasi Model

Evaluasi performa model klasifikasi pada penelitian ini dilakukan menggunakan *Confusion Matrix*. *Confusion Matrix* digunakan untuk mengukur kinerja model dalam mengklasifikasikan data ke dalam kategori yang benar berdasarkan hasil prediksi dan label sebenarnya.

Berdasarkan *Confusion Matrix*, digunakan beberapa parameter evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* untuk memberikan gambaran menyeluruh terhadap performa model.

Accuracy menunjukkan tingkat ketepatan model dalam mengklasifikasikan data secara keseluruhan dan dihitung menggunakan rumus 6.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision menunjukkan tingkat ketepatan prediksi positif yang dihasilkan oleh model dan dihitung menggunakan rumus 7.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall menunjukkan kemampuan model dalam mengidentifikasi data positif secara benar dan dihitung menggunakan rumus 8.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-score merupakan rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk menyeimbangkan kedua metrik tersebut, dan dihitung menggunakan rumus 9.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Dimana TP (*True Positive*) adalah jumlah data yang diprediksi benar sebagai positif, TN (*True Negative*) adalah jumlah data yang diprediksi benar sebagai negatif, FP (*False Positive*) adalah jumlah data negatif yang diprediksi sebagai positif, dan FN (*False Negative*) adalah jumlah data positif yang diprediksi sebagai negatif.

3. Hasil dan Pembahasan

Pada bagian ini dipaparkan hasil analisis sentimen terhadap ulasan pengguna *game Stardew Valley* yang dikumpulkan dari dua platform, yaitu *Steam* dan *Google Play Store*. Proses analisis dilakukan melalui beberapa tahapan, meliputi *preprocessing* data, pelabelan sentimen menggunakan metode VADER, ekstraksi fitur dengan TF-IDF, serta proses klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*.

Seluruh tahapan tersebut dilaksanakan secara terstruktur untuk menghasilkan model klasifikasi sentimen yang mampu mengidentifikasi polaritas ulasan pengguna. Evaluasi model dilakukan dengan metode *5-Fold Cross Validation*, dengan pengukuran kinerja menggunakan *confusion matrix* yang mencakup metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Hasil dari setiap tahapan analisis kemudian disajikan dan dibahas secara bertahap guna memberikan gambaran yang jelas mengenai performa model dalam mengklasifikasikan sentimen ulasan pengguna pada kedua platform.

3.1. Hasil Preprocessing

Tahapan *preprocessing* dilakukan untuk membersihkan dan menormalisasi data teks sebelum digunakan dalam proses analisis sentimen. Proses ini meliputi *case folding*, *cleansing*, *tokenisasi*, *stopword removal*, dan *lemmatisasi*.

Hasil *preprocessing* menunjukkan bahwa data ulasan yang semula tidak terstruktur berhasil diubah menjadi bentuk teks yang lebih sederhana dan terstandarisasi.

Proses ini menghasilkan data akhir berupa teks yang telah dibersihkan dan dinormalisasi, yang selanjutnya disebut sebagai *cleaned_final*.

Data *cleaned_final* digunakan sebagai input pada tahap pelabelan sentimen menggunakan metode VADER serta pada proses ekstraksi fitur menggunakan TF-IDF. Contoh hasil *preprocessing* pada dataset *Steam* dan *Google Play* ditunjukkan pada Tabel 1.

Tabel 1. Tabel Contoh Pra-pemrosesan *steam* dan *google play*

| Tahap | Steam | Google Play |
|--------------------|---|--|
| Review asli | This game is AMAZING!! I've played for 300 hours. | Great mobile port, controls are a bit tricky at first. |
| Case folding | this game is amazing!!! i've played for 300 hours. | great mobile port, controls are a bit tricky at first |
| Cleansing | this game is amazing ive played for hours | great mobile port controls are a bit tricky at first |
| Tokenizing | 'this','game','is','amazing','ive','played','for','300','hours' | 'great','mobile','port','controls','are','a','bit','tricky','at','first' |
| Stopword removal | 'game','amazing','ive','played','hours' | 'great','mobile','port','controls','bit','tricky','first' |
| Lemmatization/stem | 'game','amazing','ive','play','hour' | 'great','mobile','port','control','bit','tricky','first' |
| cleaned_final | game amazing ive play hour | great mobile port control bit tricky first |

3.2. Hasil Pelabelan Sentimen dan Distribusi Sentimen

Pelabelan sentimen dilakukan menggunakan metode VADER dengan memanfaatkan nilai *compound score* untuk mengelompokkan ulasan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.

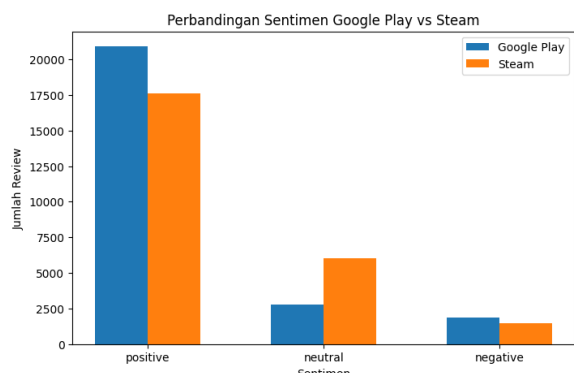
Berdasarkan hasil pelabelan pada dataset *Steam*, diperoleh sebanyak 17.626 data dengan sentimen positif, 6.006 data dengan sentimen netral, dan 1.467 data dengan sentimen negatif. Sementara itu, pada dataset *Google Play* diperoleh sebanyak 20.912 data dengan sentimen positif, 2.788 data dengan sentimen netral, dan 1.894 data dengan sentimen negatif.

Distribusi jumlah data pada masing-masing kategori sentimen untuk kedua platform ditunjukkan pada Tabel 2.

Tabel 2. Hasil Distribusi Sentimen

| Dataset | Positif | Netral | Negatif |
|-------------|---------|--------|---------|
| Steam | 17.626 | 6.006 | 1467 |
| Google Play | 20.912 | 2788 | 1894 |

Selanjutnya, distribusi sentimen tersebut divisualisasikan dalam bentuk grafik untuk mempermudah analisis perbandingan antar *platform*, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Sentimen Steam vs Google Play

Gambar 2 menunjukkan bahwa sebagian besar ulasan pada kedua *platform* memiliki sentimen positif. Pada dataset *Google Play*, jumlah ulasan positif mencapai lebih dari 20 ribu ulasan, sedangkan pada dataset *Steam* jumlah ulasan positif mencapai sekitar 17 ribu ulasan. Sebaliknya, jumlah ulasan netral dan negatif jauh lebih sedikit dibandingkan ulasan positif.

Selain itu, terlihat bahwa dataset *Steam* memiliki jumlah ulasan netral yang lebih tinggi dibandingkan *Google Play*, sedangkan *Google Play* cenderung memiliki proporsi ulasan positif yang lebih dominan.

Distribusi sentimen yang tidak seimbang ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) pada dataset. Kondisi ini dapat mempengaruhi performa model klasifikasi, terutama dalam mengenali sentimen negatif dan netral yang jumlahnya lebih sedikit dibandingkan sentimen positif.

3.3. Hasil Ekstraksi Fitur (TF-IDF)

Pada tahap ini dilakukan ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data teks hasil preprocessing menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi.

Proses ekstraksi fitur dilakukan menggunakan *TfidfVectorizer* dengan jumlah maksimum fitur sebanyak 5000 kata. Setiap dokumen ulasan direpresentasikan dalam bentuk vektor berdasarkan bobot kata yang dihitung menggunakan metode TF-IDF. Nilai bobot yang dihasilkan menunjukkan tingkat kepentingan suatu kata dalam dokumen, dimana semakin tinggi nilai TF-IDF maka semakin besar kontribusi kata tersebut dalam merepresentasikan isi dokumen.

Hasil dari proses ini berupa matriks fitur berbentuk numerik dengan dimensi sejumlah dokumen dan jumlah fitur yang digunakan. Matriks TF-IDF yang dihasilkan bersifat *sparse*, dimana sebagian besar nilainya adalah nol karena tidak semua kata muncul pada setiap dokumen.

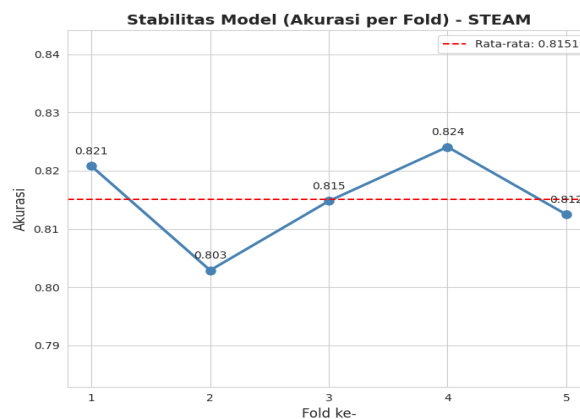
Sebagai contoh, beberapa kata yang memiliki bobot tinggi dalam dataset antara lain berkaitan dengan pengalaman bermain seperti “*game*”, “*farm*”, “*fun*”, dan “*story*”, yang menunjukkan bahwa kata-kata tersebut memiliki peran penting dalam membentuk sentimen ulasan pengguna.

Matriks TF-IDF yang dihasilkan kemudian digunakan sebagai input dalam proses klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes* pada tahap berikutnya.

3.4. Hasil Evaluasi Model pada Dataset Steam

Evaluasi model pada dataset ulasan *Steam* dilakukan menggunakan metode *5-Fold Cross Validation*, dimana dataset dibagi menjadi lima bagian dan pada setiap iterasi digunakan kombinasi data pelatihan dan data pengujian secara bergantian.

Berdasarkan hasil evaluasi, nilai akurasi model pada dataset *Steam* berada pada rentang 0.803 hingga 0.824 dengan rata-rata akurasi sebesar 0.8151. Visualisasi stabilitas model berdasarkan nilai akurasi pada setiap fold ditunjukkan pada Gambar 3.



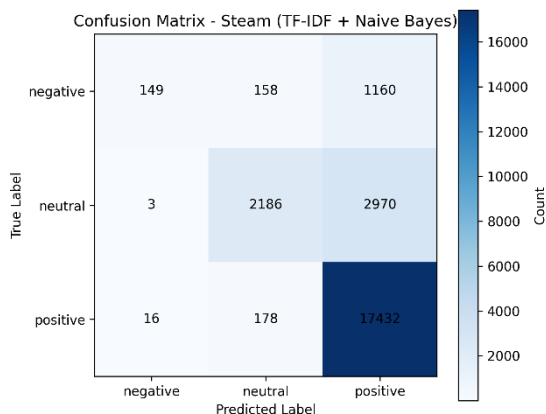
Gambar 3. Stabilitas Model Berdasarkan Akurasi pada Setiap Fold (Steam)

Berdasarkan gambar 3, nilai akurasi pada setiap *fold* menunjukkan adanya variasi. Selisih nilai akurasi antar *fold* sebesar 0.021 menunjukkan bahwa perubahan performa model masih berada dalam rentang yang relatif kecil. Meskipun terdapat fluktuasi nilai akurasi pada setiap *fold*, perbedaan tersebut tidak terlalu signifikan sehingga model tetap menunjukkan performa yang tidak banyak berubah pada setiap pembagian data. Detail nilai akurasi pada setiap *fold* disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model per Fold (Steam)

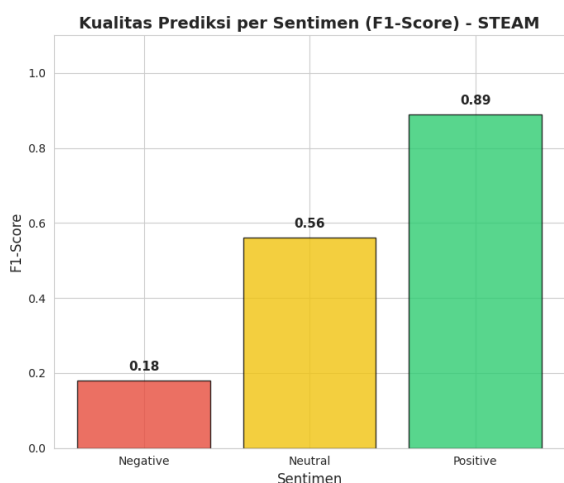
| Fold | Accuracy |
|-----------|----------|
| 1 | 0.82 |
| 2 | 0.80 |
| 3 | 0.81 |
| 4 | 0.82 |
| 5 | 0.81 |
| Rata-Rata | 0.8151 |

Berdasarkan Tabel 3, nilai akurasi pada setiap *fold* berada pada rentang 0.80 hingga 0.82. Nilai akurasi tertinggi diperoleh pada *fold* ke-1 dan ke-4 sebesar 0.82, sedangkan nilai terendah terdapat pada *fold* ke-2 sebesar 0.80. Rata-rata akurasi sebesar 0.8151 menunjukkan bahwa perbedaan performa model antar *fold* relatif kecil pada setiap pembagian data. Distribusi hasil klasifikasi model dapat dilihat melalui confusion matrix pada Gambar 4.



Gambar 4. Confusion Matrix Klasifikasi Sentimen Dataset Steam

Berdasarkan *confusion matrix*, model mampu mengklasifikasikan sebagian besar data dengan benar, terutama pada kategori sentimen positif. Namun, pada kategori sentimen negatif masih terdapat kesalahan klasifikasi yang cukup tinggi, dimana sebagian besar ulasan negatif diklasifikasikan sebagai sentimen positif atau netral. Performa model pada setiap kategori sentimen berdasarkan nilai *F1-score* ditunjukkan pada Gambar 5.



Gambar 5. Nilai *F1-Score* untuk Setiap Kategori Sentimen pada Dataset Steam

Nilai *F1-score* digunakan untuk mengukur keseimbangan antara *precision* dan *recall* pada setiap kategori. Berdasarkan Gambar 5, kategori sentimen positif memiliki nilai *F1-score* tertinggi sebesar 0.89, yang menunjukkan bahwa model mampu mengenali pola sentimen positif dengan baik.

Sebaliknya, kategori sentimen negatif memiliki nilai *F1-score* yang paling rendah, yaitu sebesar 0.18. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi ulasan dengan sentimen negatif secara akurat. Ringkasan performa model berdasarkan nilai *precision*, *recall*, dan *F1-score* disajikan pada Tabel 4.

Tabel 4. Rata-rata Performa Model (Steam)

| Kategori | Precision | Recall | F1-Score |
|----------|-----------|--------|----------|
| Positif | 0.81 | 0.99 | 0.89 |
| Netral | 0.87 | 0.42 | 0.57 |
| Negatif | 0.89 | 0.10 | 0.18 |

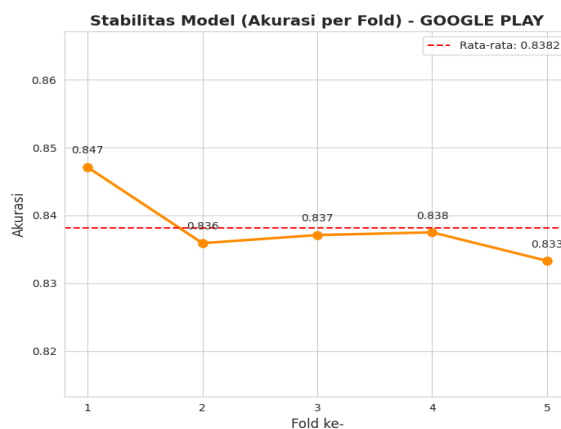
Berdasarkan Tabel 4, kategori sentimen positif memiliki nilai *recall* sebesar 0.99 dan *F1-score* sebesar 0.89, yang menunjukkan bahwa hampir seluruh data sentimen positif berhasil dikenali oleh model dengan baik.

Sebaliknya, pada kategori sentimen netral diperoleh nilai *recall* sebesar 0.42 dengan *F1-score* sebesar 0.57, sedangkan pada kategori negatif nilai *recall* hanya sebesar 0.10 dengan *F1-score* sebesar 0.18. Hal ini menunjukkan bahwa sebagian besar data netral dan negatif belum dapat diklasifikasikan secara akurat oleh model.

Perbedaan performa ini menunjukkan bahwa model cenderung lebih efektif dalam mengenali sentimen positif dibandingkan kategori lainnya. Kondisi ini dipengaruhi oleh ketidakseimbangan distribusi data (*class imbalance*), dimana jumlah data sentimen positif jauh lebih dominan dibandingkan kategori lainnya.

3.5. Hasil Evaluasi Model pada Dataset Google Play

Pengujian model pada *dataset* ulasan *Google Play* juga dilakukan menggunakan metode *5-Fold Cross Validation* sebagai prosedur evaluasi yang sama seperti pada *dataset Steam*.



Gambar 6. Stabilitas Model Berdasarkan Akurasi pada Setiap *Fold* (*Google Play*)

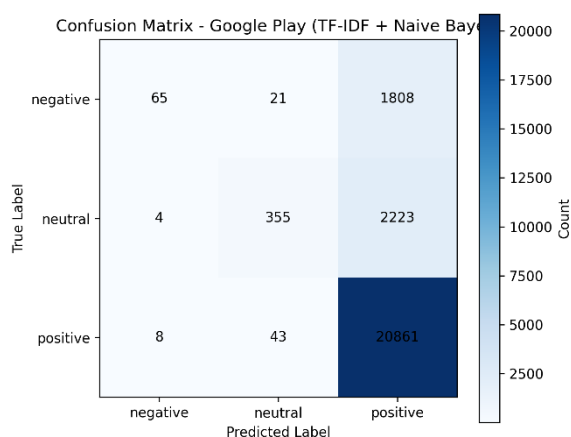
Berdasarkan Gambar 6, nilai akurasi pada setiap *fold* menunjukkan adanya variasi. Selisih nilai akurasi antar

fold sebesar 0.02 menunjukkan bahwa perubahan performa model masih berada dalam rentang yang relatif kecil. Meskipun terdapat fluktuasi, perbedaan tersebut tidak terlalu signifikan sehingga performa model tidak banyak berubah pada setiap pembagian data. Detail nilai akurasi pada setiap *fold* disajikan pada Tabel 5.

Tabel 5. Hasil Evaluasi Model per *Fold* (Google Play)

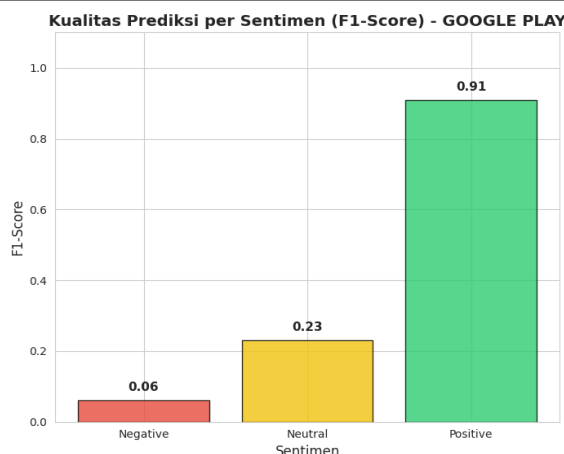
| <i>Fold</i> | Accuracy |
|-------------|----------|
| 1 | 0.85 |
| 2 | 0.84 |
| 3 | 0.84 |
| 4 | 0.84 |
| 5 | 0.83 |
| Rata-Rata | 0.8382 |

Berdasarkan Tabel 5, nilai akurasi pada setiap *fold* berada pada rentang 0.83 hingga 0.85. Nilai akurasi tertinggi diperoleh pada *fold* ke-1 sebesar 0.85, sedangkan nilai terendah terdapat pada *fold* ke-5 sebesar 0.83. Rata-rata akurasi sebesar 0.8382 menunjukkan bahwa perbedaan performa model antar *fold* relatif kecil pada setiap pembagian data. Distribusi hasil klasifikasi model dapat dilihat melalui confusion matrix pada Gambar 7.



Gambar 7. Confusion Matrix Klasifikasi Sentimen Dataset Google Play.

Berdasarkan *confusion matrix*, model memiliki kemampuan yang sangat baik dalam mengklasifikasikan sentimen positif, dengan jumlah prediksi benar yang sangat tinggi. Namun, performa model pada kategori sentimen netral dan negatif masih rendah. Performa model pada setiap kategori sentimen berdasarkan nilai *F1-score* ditunjukkan pada Gambar 8.



Gambar 8. Nilai *F1-Score* untuk Setiap Kategori Sentimen pada Dataset Google Play

Nilai *F1-score* digunakan untuk mengukur keseimbangan antara *precision* dan *recall*. Berdasarkan Gambar 8, kategori sentimen positif memiliki nilai *F1-score* tertinggi sebesar 0.91, yang menunjukkan bahwa model sangat baik dalam mengenali ulasan dengan sentimen positif.

Sebaliknya, kategori sentimen negatif memiliki nilai *F1-score* yang paling rendah, yaitu sebesar 0.06, diikuti oleh kategori netral sebesar 0.24. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi sentimen negatif dan netral secara akurat. Ringkasan performa model berdasarkan nilai *precision*, *recall*, dan *F1-score* disajikan pada Tabel 6.

Tabel 6. Rata-Rata Performa Model (Google Play)

| Kategori | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
|----------|------------------|---------------|-----------------|
| Positif | 0.84 | 1.00 | 0.91 |
| Netral | 0.84 | 0.14 | 0.24 |
| Negatif | 0.85 | 0.04 | 0.06 |

Nilai *precision* dan *recall* yang sangat tinggi pada Berdasarkan Tabel 6, kategori sentimen positif memiliki nilai *precision* sebesar 0.84, *recall* sebesar 1.00, dan *F1-Score* sebesar 0.91. Nilai *recall* yang mencapai 1.00 menunjukkan bahwa seluruh data sentimen positif berhasil dikenali oleh model dengan sangat baik.

Sebaliknya, pada kategori sentimen netral diperoleh nilai *recall* sebesar 0.14 dan *F1-score* sebesar 0.24, sedangkan pada kategori negatif nilai *recall* hanya sebesar 0.04 dengan *F1-score* sebesar 0.06. Nilai tersebut menunjukkan bahwa sebagian besar data netral dan negatif tidak berhasil diklasifikasikan dengan benar oleh model.

Perbedaan performa ini menunjukkan bahwa model cenderung lebih fokus dalam mengenali sentimen positif dibandingkan kategori lainnya. Hal ini dipengaruhi oleh ketidakseimbangan distribusi data (*class imbalance*), dimana jumlah data sentimen positif jauh lebih dominan dibandingkan kategori lainnya.

Selain itu, karakteristik ulasan pada *Google Play* yang cenderung lebih singkat menyebabkan informasi yang tersedia menjadi terbatas, sehingga model kesulitan dalam membedakan sentimen negatif dan netral secara lebih akurat.

3.6. Hasil Perbandingan Performa Model

Pada bagian ini dilakukan perbandingan performa model klasifikasi sentimen antara *dataset Steam* dan *Google Play* berdasarkan hasil evaluasi menggunakan metode *5-Fold Cross Validation*. Perbandingan performa model pada kedua *dataset* disajikan pada Tabel 7.

Tabel 7. Perbandingan Performa Model Klasifikasi Sentimen

| Datas et | Accu racy | Preci sion (Pos) | Recall (Pos) | F1 (Pos) | F1 (Netra l) | F1 (Negat if) |
|----------------|-----------|------------------|--------------|----------|--------------|---------------|
| Steam 1 | 0.815 | 0.81 | 0.99 | 0.89 | 0.57 | 0.18 |
| Googl e Play 2 | 0.838 | 0.84 | 1.00 | 0.91 | 0.24 | 0.06 |

Nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang digunakan dalam perbandingan ini diperoleh dari hasil perhitungan berdasarkan *confusion matrix* pada masing-masing *dataset*. Nilai *precision* dan *recall* difokuskan pada kategori sentimen positif karena kategori tersebut merupakan kelas dominan dalam *dataset* dan memiliki kontribusi terbesar terhadap performa model secara keseluruhan.

Berdasarkan Tabel 7, model menunjukkan performa yang lebih baik pada *dataset Google Play* dengan nilai akurasi sebesar 0.8382 dibandingkan *dataset Steam* sebesar 0.8151.

Pada kategori sentimen positif, kedua *dataset* memiliki performa yang sangat baik dengan nilai *F1-score* sebesar 0.89 pada *Steam* dan 0.91 pada *Google Play*. Hal ini menunjukkan bahwa model mampu mengenali sentimen positif secara efektif pada kedua *platform*.

Namun, pada kategori sentimen netral dan negatif terdapat perbedaan performa yang cukup signifikan. *Dataset Steam* memiliki nilai *F1-score* sebesar 0.57 untuk kategori netral dan 0.18 untuk kategori negatif, sedangkan pada *dataset Google Play* masing-masing hanya sebesar 0.24 dan 0.06.

Perbedaan ini menunjukkan bahwa model lebih mampu mengenali sentimen non-positif pada *dataset Steam* dibandingkan *Google Play*. Hal ini diduga dipengaruhi oleh karakteristik ulasan pada *Steam* yang cenderung lebih panjang dan informatif, sehingga memberikan konteks yang lebih jelas bagi model dalam melakukan klasifikasi.

Selain itu, ketidakseimbangan distribusi data (*class imbalance*) juga mempengaruhi performa model, dimana dominasi data sentimen positif menyebabkan model lebih optimal dalam mengenali sentimen positif dibandingkan kategori lainnya.

Secara keseluruhan, kombinasi metode TF-IDF dan Multinomial Naïve Bayes mampu memberikan performa klasifikasi yang cukup baik, meskipun masih memiliki keterbatasan dalam mengidentifikasi sentimen negatif dan netral.

4. Kesimpulan

Penelitian ini bertujuan menganalisis perasaan konsumen kepada penilaian *game Stardew Valley* di *platform Steam* dan *Google Play* dengan metode text mining dan machine learning. Proses analisis dilakukan dalam beberapa tahap, diantaranya adalah mengumpulkan data, memproses teks sebelumnya, memberi label sentimen dengan metode VADER, mengekstrak fitur menggunakan TF-IDF, serta mengklasifikasikan data dengan algoritma Multinomial Naïve Bayes.

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh *dataset* sebanyak 25.099 ulasan dari *platform Steam* dan 25.594 ulasan dari *platform Google Play*. Hasil pelabelan sentimen menunjukkan bahwa mayoritas ulasan pada kedua *platform* menunjukkan sentimen positif, yang menunjukkan bahwa *game Stardew Valley* mendapatkan respons yang baik dari para pemain.

Hasil pengujian dengan metode *5-Fold Cross Validation* menunjukkan indikasi bahwa algoritma *Multinomial Naïve Bayes* mampu memberikan kinerja klasifikasi yang memadai. Model mencapai rata-rata akurasi sebesar 0,8151 pada *dataset Steam* dan 0,8382 pada *dataset Google Play Store*. Perbedaan tersebut menunjukkan bahwa performa model cenderung lebih baik pada *dataset Google Play Store* dibandingkan dengan *dataset Steam*.

Perbedaan performa tersebut kemungkinan dipengaruhi oleh karakteristik teks ulasan pada masing-masing *platform*. Ulasan pada *Google Play* umumnya lebih singkat dan memiliki struktur kalimat yang lebih sederhana sehingga pola kata lebih mudah dipelajari oleh model klasifikasi. Sebaliknya, ulasan pada *platform Steam* cenderung lebih panjang dan memiliki variasi bahasa yang lebih kompleks.

Secara umum, penelitian ini menunjukkan bahwa menggunakan metode VADER untuk memberi label pada sentimen, TF-IDF untuk mengambil fitur, serta *Multinomial Naïve Bayes* untuk mengklasifikasikan dapat dilakukan dengan baik dalam menganalisis sentimen dari ulasan pengguna di *platform* distribusi *game* digital.

Untuk penelitian berikutnya, sebaiknya dicoba metode klasifikasi lain seperti *Support Vector Machine*, *Random Forest*, atau model *deep learning*. Selain itu, bisa juga digunakan teknik untuk mengimbangkan data agar model lebih akurat dalam mengklasifikasikan sentimen negatif dan netral.

Daftar Rujukan

- [1] D. Sulistyio, F. Ahda, and V. Aida Fitria, "Epistemologi dalam Natural Language Processing," *J. Inov. Teknol. dan Edukasi Tek.*, vol. 1, pp. 652–664, 2021, doi: 10.17977/um068v1i92021p652-664.
- [2] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," pp. 216–225.
- [3] D. A. Sulistyio, "LSTM-Based Machine Translation for Madurese-Indonesian," vol. 4, no. 3, pp. 190–199, 2023.
- [4] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed. 2026. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/>
- [5] J. Iqbal et al., "Sentiment Analysis Classification of E-commerce User Reviews Using Natural Language Processing (NLP) and Support Vector Machine (SVM) Methods," *Bull. Inf. Technol.*, vol. 6, no. 2, pp. 196–205, 2025, doi: 10.47065/bit.v5i2.2018.
- [6] M. Y. Febrianta, S. Widiyanesti, and S. R. Ramadhan, "Analisis Ulasan Indie Video Game Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *J. Animat. Games Stud.*, vol. 7, no. 2, pp. 117–144, 2021, doi: 10.24821/jags.v7i2.5162.
- [7] A. R. D. Khusnul Khotimah, Martanto and A. Rifa'i, "GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA RoBERTa," *J. Comasie*, vol. 13, no. 03, 2025.
- [8] R. Aprinastya, M. Jazman, S. Syaifullah, M. Rahmawita, S. Siregar, and E. Saputra, "Comparative Analysis of Naïve Bayes Classifier and Support Vector Machine for Multilingual Sentiment Analysis: Insights from Genshin Impact User Reviews," *JUSIFO (Jurnal Sist. Informasi)*, vol. 10, no. 2, pp. 117–126, 2024, doi: 10.19109/jusifo.v10i2.24876.
- [9] U. Atma and J. Yogyakarta, "Survey Paper tentang Analisis Sentimen," vol. 2, no. 1, pp. 201–206, 2022.
- [10] B. Liu, "Sentiment Analysis and Opinion Mining," no. May, 2012.
- [11] L. Zhang, "Features extraction based on Naive Bayes algorithm and TF-IDF for news classification," *PLoS One*, vol. 20, no. 7, p. e0327347, 2025, doi: 10.1371/journal.pone.0327347.
- [12] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008. [Online]. Available: <https://nlp.stanford.edu/IR-book/>
- [13] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, Mar. 2002, doi: 10.1145/505282.505283.
- [14] R. Li, M. Liu, D. Xu, J. Gao, F. Wu, and L. Zhu, "A Review of Machine Learning Algorithms for Text Classification," in *Cyber Security*, W. Lu, Y. Zhang, W. Wen, H. Yan, and C. Li, Eds., Singapore: Springer Nature Singapore, 2022, pp. 226–234.
- [15] F. Ahda, A. Wibawa, D. Prasetya, and D. Sulistyio, "Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation," *JOIV Int. J. Informatics Vis.*, vol. 8, p. 231, 2024, doi: 10.62527/joiv.8.1.1818.
- [16] D. Arbian, A. Prasetya, D. Dwi, and F. Almu, "Multilingual Parallel Corpus for Indonesian Low-Resource Languages," vol. 9, no. September, 2025.
- [17] D. A. Sulistyio, D. D. Prasetya, F. A. Ahda, and A. P. Wibawa, "Pivoted Low Resource Multilingual Translation with NER Optimization," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 24, no. 5, May 2025, doi: 10.1145/3727876.
- [18] D. Science, A. S. Wicaksono, U. Sultan, and A. Tirtayasa, "Twitter sentiment analysis of electric vehicle subsidy policy using naïve bayes algorithm," vol. 01, no. June, pp. 1–9, 2023.
- [19] C. C. Aggarwal and C. Zhai, "A Survey of Text Classification Algorithms," in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds., Boston, MA: Springer US, 2012, pp. 163–222. doi: 10.1007/978-1-4614-3223-4_6.
- [20] J. Wilson and C. Hernández-Hall, "Octava Conferencia Internacional AAAI sobre Weblogs y Redes Sociales," *Eighth Int. AAAI Conf. Weblogs Soc. Media*, p. 18, 2014, [Online]. Available: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>
- [21] J. Wilson and C. Hernández-Hall, "Octava Conferencia Internacional AAAI sobre Weblogs y Redes Sociales," *Eighth Int. AAAI Conf. Weblogs Soc. Media*, p. 18, 2014.
- [22] H. Akshay Koushik, R. B. Bharadwaj, R. P. E. Naik, G. Ramesh, M. J. Yogesh, and S. Habeeb, "Detection and Classification of Diseased Mangoes," in *2020 International Conference on Computer Science and Its Application in Agriculture (ICOSICA)*, 2020, pp. 1–8. doi: 10.1109/ICOSICA49951.2020.9243277.
- [23] K. Harmandini and K. L., "Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment," *Sinkron*, vol. 8, pp. 929–937, 2024, doi: 10.33395/sinkron.v8i2.13376.