

ANALISIS SENTIMEN ULASAN GROWTOPIA DI GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER UNTUK IDENTIFIKASI KEBUTUHAN PENGGUNA

Rafli Putra Setiawan^{1*}, Bambang Irawan², Willy Prihartono³

^{1,2}Teknik Informatika, STMIK IKMI Cirebon; Jl. Perjuangan No.10B, Kota Cirebon 45131; Telp. & Fax. (0231) 490481

³Komputerisasi Akuntansi, STMIK IKMI Cirebon; Jl. Perjuangan No.10B, Kota Cirebon 45131; Telp. & Fax. (0231) 490481

Received: 7 Maret 2025
Accepted: 27 Maret 2025
Published: 14 April 2025

Keywords:

Sentiment Analysis, Naïve Bayes Classifier, Growtopia, User Reviews, Google Play Store

Correspondent Email:

rafliputra24ss@gmail.com

Abstrak. Analisis sentimen memainkan peran penting dalam memahami bagaimana pengguna menilai aplikasi digital, termasuk game. Dalam penelitian ini, penulis menganalisis sentimen pengguna terhadap game Growtopia di Google Play Store dengan menggunakan algoritma Naïve Bayes Classifier. Sebanyak 2.000 ulasan berbahasa Indonesia dikumpulkan melalui proses web scraping. Tahapan preprocessing mencakup beberapa langkah penting, yaitu pembersihan teks, tokenisasi, penghapusan stopwords, dan stemming. Setelah itu, algoritma Naïve Bayes digunakan untuk mengklasifikasikan ulasan tersebut ke dalam dua kategori: sentimen positif dan negatif. Hasil analisis menunjukkan bahwa algoritma ini mampu memberikan akurasi yang baik dengan mendeteksi 27 ulasan sebagai true positive dan 456 ulasan sebagai true negative. Ulasan dengan sentimen positif umumnya menyoroti aspek kreativitas gameplay dan interaksi komunitas, sedangkan ulasan negatif banyak membahas masalah seperti bug, kendala saat login, dan koneksi yang tidak stabil. Temuan ini menegaskan bahwa metode Naïve Bayes efektif dalam menganalisis sentimen sekaligus memberikan masukan berharga bagi pengembang untuk meningkatkan kualitas permainan.

Abstract. Sentiment analysis plays an important role in understanding how users rate digital applications, including games. In this study, the authors analyzed user sentiment towards the Growtopia game on the Google Play Store using the Naïve Bayes Classifier algorithm. A total of 2,000 Indonesian reviews were collected through a web scraping process. The preprocessing stage includes several important steps, namely text cleaning, tokenization, stopwords removal, and stemming. After that, the Naïve Bayes algorithm was used to classify the reviews into two categories: positive and negative sentiment. The analysis showed that the algorithm was able to provide good accuracy by detecting 27 reviews as true positives and 456 reviews as true negatives. Reviews with positive sentiment generally highlighted aspects of gameplay creativity and community interaction, while negative reviews mostly discussed issues such as bugs, login issues, and unstable connections. This finding confirms that the Naïve Bayes method is effective in analyzing sentiment while providing valuable feedback for developers to improve the quality of the game.

1. PENDAHULUAN

Perkembangan teknologi informasi telah memberikan dampak yang signifikan terhadap industri permainan, termasuk dalam memahami pengalaman pengguna melalui analisis sentimen. Game Growtopia, yang dikenal dengan elemen kreatif dan interaktifnya, telah menarik perhatian banyak pemain di seluruh dunia. Namun, seiring dengan pertumbuhannya, berbagai masalah seperti lag, bug, dan kesulitan dalam proses login sering kali dikeluhkan oleh pengguna. Dalam konteks ini, analisis sentimen dapat memberikan wawasan yang berharga bagi pengembang untuk meningkatkan kualitas permainan [1]. Pemahaman terhadap sentimen pengguna sangat penting, karena dapat membantu pengembang dalam menciptakan pengalaman bermain yang lebih baik dan mengurangi keluhan yang sering muncul. Dengan menerapkan algoritma Naïve Bayes, analisis terhadap ulasan pengguna Growtopia dapat dilakukan secara efisien, sehingga memberikan gambaran yang jelas mengenai persepsi pemain terhadap permainan ini.[2] menekankan pentingnya adaptasi teknologi dalam memahami dinamika pasar permainan. Melalui pendekatan ini, diharapkan pengembang dapat merespons kebutuhan pengguna dengan lebih baik, serta meningkatkan kepuasan dan loyalitas pemain terhadap game Growtopia [3]. Dengan demikian, analisis sentimen tidak hanya berfungsi sebagai alat untuk mengidentifikasi masalah yang dihadapi oleh pengguna, tetapi juga sebagai sarana untuk meningkatkan interaksi antara pengembang dan komunitas pemain. Pengembang menciptakan pengalaman bermain yang lebih memuaskan. Hal ini sejalan dengan tujuan utama dalam industri game, yaitu memberikan pengalaman yang menyenangkan dan berkualitas tinggi bagi setiap pemain. Analisis ini didasarkan pada teori mengenai analisis sentimen dalam game online, yang menunjukkan bahwa ulasan pengguna dapat memberikan wawasan mendalam tentang pengalaman bermain dan aspek teknis yang perlu diperbaiki. Menurut penelitian sebelumnya, analisis sentimen dalam game sering digunakan untuk mengidentifikasi pola kepuasan dan keluhan pengguna, sehingga pengembang dapat merancang strategi perbaikan yang lebih tepat sasaran. Tujuan

utama dari penelitian ini adalah untuk menerapkan algoritma Naïve Bayes Classifier dalam menganalisis sentimen ulasan pengguna tentang game Growtopia yang terdapat di Google Play Store. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen positif, negatif, dan netral dari ulasan tersebut, sehingga memberikan wawasan yang lebih mendalam mengenai pengalaman dan kebutuhan pengguna. Signifikansi penelitian ini terletak pada kemampuannya untuk mengisi kesenjangan pengetahuan terkait analisis sentimen dalam konteks permainan, serta memberikan kontribusi yang berarti bagi pengembang game dalam memahami persepsi pengguna terhadap produk mereka [4].

2. TINJAUAN PUSTAKA

Penelitian yang dibahas dalam jurnal [5] berfokus pada analisis sentimen pengguna terkait game Genshin Impact di Google Play. memberikan masukan berharga bagi pengembang untuk melakukan perbaikan berdasarkan opini pengguna. Metode yang digunakan menggabungkan algoritma Naïve Bayes dengan teknik pembobotan TF-IDF. Proses analisis mencakup beberapa tahapan penting, seperti pembersihan teks (text preprocessing), pelabelan data, hingga pemberian bobot menggunakan TF-IDF. Dataset yang dianalisis berisi ulasan pengguna dengan rating mulai dari 1 hingga 5 bintang. Hasil akhir menunjukkan bahwa ulasan cenderung bernada positif, dengan tingkat akurasi mencapai 87%, precision 67%, recall 98%, dan f1-score sebesar 67%. Penelitian ini berkontribusi secara teoritis dalam mengembangkan model klasifikasi sentimen di dunia game sekaligus memberikan wawasan praktis yang bermanfaat bagi pengembang, termasuk saran untuk mempertimbangkan penggunaan simbol emoji dalam analisis sentimen ke depannya.

Sementara itu, jurnal [6] menyoroti pentingnya memahami kecanduan game online pada remaja dan bagaimana algoritma Naïve Bayes Classifier dapat dimanfaatkan untuk mengklasifikasikan tingkat kecanduan tersebut. Meski fokus utamanya berbeda, penelitian ini memiliki kemiripan dengan studi yang tengah saya lakukan, yaitu penggunaan metode Naïve Bayes untuk menganalisis data. Dalam penelitian saya yang berjudul 'Penggunaan

Algoritma Naïve Bayes Classifier dalam Analisis Sentimen Ulasan Growtopia di Google Play Store: Manfaat bagi Pengembang Game dalam Mengidentifikasi Kebutuhan Pengguna', pendekatan yang digunakan berfokus pada analisis sentimen pengguna guna memahami persepsi mereka terhadap game tersebut.

Penelitian yang dibahas dalam jurnal [7] turut menyoroti peran penting analisis sentimen dalam memahami pandangan pemain terhadap sebuah game. Hal ini sangat relevan dengan penelitian saya, karena kedua studi sama-sama menggunakan metode Naïve Bayes untuk mengolah ulasan pengguna. Fokus utama dari pendekatan ini adalah memahami bagaimana sentimen pemain dapat memengaruhi keputusan bisnis dan pengembangan game di masa mendatang.

Jurnal [8] mengupas analisis sentimen pada ulasan pengguna aplikasi streaming Vidio yang tersedia di Google Play Store. Dalam penelitian ini, algoritma Naïve Bayes diterapkan pada 1.500 ulasan yang dikumpulkan melalui web scraping. Setelah dilakukan pembersihan data, jumlah ulasan yang tersisa adalah 1.475, terdiri dari 1.310 ulasan negatif dan 165 ulasan positif. Penelitian ini mencakup langkah-langkah seperti pelabelan manual, preprocessing, dan pembobotan TF-IDF sebelum akhirnya diklasifikasikan. Model yang dikembangkan mencapai akurasi sebesar 80,28%, dengan precision sebesar 24,18% dan recall 35,76%. Visualisasi wordcloud yang dibuat menampilkan kata 'nonton' sebagai istilah yang paling dominan. Temuan ini memberikan wawasan berharga terkait pengalaman pengguna yang dapat menjadi referensi berharga bagi pengembang untuk meningkatkan layanan mereka.

Dalam jurnal [9], penelitian menyoroti betapa pentingnya analisis sentimen dalam memahami persepsi pengguna terhadap aplikasi atau game di platform digital. Penelitian ini memiliki relevansi dengan studi saya, terutama karena sama-sama menggunakan metode Naïve Bayes untuk mengolah sentimen pengguna pada ulasan game di Google Play Store. Pendekatan yang diterapkan dalam studi tersebut dapat menjadi referensi yang berguna dalam menganalisis ulasan pengguna terhadap game Growtopia.

Selanjutnya, jurnal sapu [10] membahas efektivitas algoritma Machine Learning seperti

Multinomial Naïve Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen ulasan aplikasi di Google Play Store. Penelitian ini memiliki keterkaitan dengan studi saya karena turut menggunakan algoritma Machine Learning untuk menganalisis sentimen. Meski metode yang diterapkan sedikit berbeda, baik penelitian ini maupun studi saya memiliki tujuan yang sama, yaitu memahami persepsi pengguna melalui analisis sentimen di platform Google Play Store.

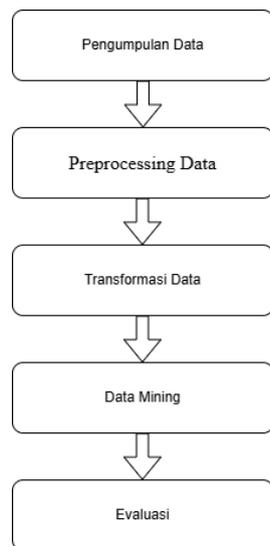
Jurnal [11] membahas penggunaan algoritma Naïve Bayes dengan metode seleksi fitur berbasis Chi-square untuk mengklasifikasikan email spam dan non-spam. Penelitian ini mencatat bahwa meskipun model yang diusulkan memiliki tingkat precision yang cukup baik, nilai recall yang rendah menandakan bahwa model tersebut masih memiliki keterbatasan dalam mendeteksi semua spam dengan akurat. Sebagai solusi, penelitian ini merekomendasikan penggunaan dataset yang lebih besar serta eksplorasi metode preprocessing yang lebih canggih agar performa model dapat lebih optimal.

Terakhir, Jurnal [12] yang membahas tentang metode dan teori yang digunakan untuk mendeteksi serangan pada jaringan komputer dengan pendekatan berbasis data. Metode utama yang digunakan adalah Naïve Bayes untuk klasifikasi data dan Principal Component Analysis (PCA) untuk reduksi dimensi data guna meningkatkan akurasi model. Penelitian ini berfokus pada meningkatkan kinerja deteksi serangan dengan menggabungkan PCA untuk menyaring fitur yang kurang relevan dan Naïve Bayes untuk klasifikasi yang lebih akurat.

3. METODE PENELITIAN

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD) sebagai pendekatan utama dalam proses analisis data. KDD merupakan metode yang dirancang untuk menganalisis data terstruktur dengan tujuan menemukan informasi baru dan mengidentifikasi pola tersembunyi pada kumpulan data yang besar dan kompleks [13]. Salah satu elemen penting dalam KDD adalah data mining, yaitu proses eksplorasi data menggunakan algoritma tertentu untuk membangun model dan mengungkap pola yang tidak terlihat secara kasat mata. Dalam penerapan KDD pada analisis teks, langkah-

langkah yang diterapkan sejalan dengan tahapan data mining, yaitu: (1) Pemilihan Data (Data Selection), (2) Pra-pemrosesan (Preprocessing), (3) Transformasi Data (Transformation), (4) Penggalian Data (Data Mining), dan (5) Evaluasi (Evaluation).[6]



Gambar 1 Tahapan Penelitian

1. Data Selection dilakukan dengan memilih data yang relevan untuk memastikan data sesuai dengan tujuan prediksi.
2. Preprocessing membersihkan data dari nilai yang hilang atau tidak konsisten, serta menormalkan data agar model dapat bekerja secara optimal.
3. Transformation, data yang telah diproses diubah menjadi format yang cocok untuk analisis, misalnya dengan teknik standarisasi atau pengurangan dimensi seperti PCA agar fitur-fitur yang dipertimbangkan lebih efisien.
4. Data Mining, algoritma Naive Bayes diterapkan untuk melakukan klasifikasi dengan menghitung probabilitas kelas berdasarkan fitur-fitur yang ada, dengan asumsi bahwa setiap fitur bersifat independen.
5. Interpretation and Evaluation, performa model dievaluasi dengan metrik seperti akurasi, precision, dan recall, yang membantu memastikan bahwa model telah belajar pola yang signifikan dari data dan memberikan prediksi yang akurat.

4. HASIL DAN PEMBAHASAN

Hasil penelitian menjadi bagian yang krusial dalam sebuah studi ilmiah, karena mencakup temuan, data, dan analisis yang diperoleh melalui pendekatan yang sistematis dan metodologis. Temuan ini berperan penting dalam menjawab pertanyaan penelitian atau menguji hipotesis yang telah dirumuskan, sehingga menghasilkan kesimpulan yang relevan. Dalam studi ini, proses mengumpulkan data dilakukan menggunakan bahasa pemrograman Python melalui platform Google Colaboratory dengan bantuan pustaka google-play-scraper. Metode ini dipilih karena mampu menjamin akurasi sekaligus efisiensi selama proses pengambilan data. Penelitian ini berfokus pada analisis sentimen ulasan pengguna aplikasi Growtopia di Google Play Store, dengan tujuan memberikan pemahaman yang lebih mendalam terkait persepsi pengguna. Hasil analisis ini diharapkan dapat menjadi referensi bagi pengembang dalam memahami kebutuhan pengguna serta menjadi dasar dalam meningkatkan fitur dan layanan aplikasi. Keberhasilan penelitian ini sangat bergantung pada validitas data yang dikumpulkan serta kualitas analisis yang dilakukan. Dengan demikian, temuan dari studi ini diharapkan mampu memberikan kontribusi yang berarti bagi kemajuan ilmu pengetahuan, kebijakan, serta penerapan teknologi di masa depan [15]

A. Pengumpulan Data

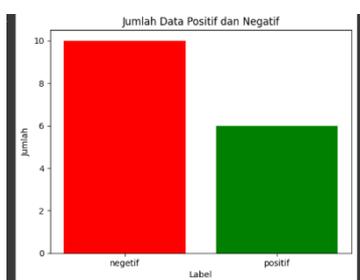
Penelitian ini melibatkan pengumpulan 2000 ulasan pengguna berbahasa Indonesia dari Google Play Store terkait game Growtopia. Data ulasan dipilih menggunakan metode purposive sampling dengan mengutamakan dua kategori utama, yakni ulasan formal yang membahas layanan pelanggan (customer service) dan ulasan non-formal yang berkaitan dengan fitur permainan. Ulasan-ulasan tersebut kemudian dikelompokkan ke dalam dua kategori sentimen utama, yakni positif dan negatif.

B. Preprocessing Data

Hasil preprocessing data mencakup pada hasil pelabelan, case folding, stopword removal, tokenisasi dan stemming.

• Hasil Pelabelan

Pada tahap berikutnya, penulis melakukan proses pelabelan dengan menggunakan skor sebagai acuan untuk menentukan label. Skor yang bernilai kurang dari 3 diberi label ulasan negatif, sedangkan skor 4 dan 5 diberi label ulasan positif. Selanjutnya, label tersebut diberi nilai numerik, di mana label positif diberi nilai 1 dan label negatif diberi nilai 0. Proses ini digambarkan seperti pada Gambar 1.1 ilustrasi berikut



Gambar 1. 1 Hasil Pelabelan

• Case Folding

Case folding adalah langkah penting dalam proses pemrosesan teks yang berfungsi untuk menyamakan semua huruf menjadi bentuk kecil. Dalam tahap ini, huruf besar seperti 'A'-'Z' akan diubah menjadi huruf kecil 'a'-'z'. Proses ini bertujuan untuk memastikan konsistensi data sehingga tidak ada perbedaan makna hanya karena perbedaan penggunaan huruf besar atau kecil. Dengan penerapan case folding, teks yang sebelumnya mengandung huruf kapital akan dikonversi sepenuhnya menjadi huruf kecil. Hasil dari langkah ini dapat dilihat pada Tabel 1 berikut.

Tabel 1 Hasil Case Folding

Sebelum	Sesudah	Rasio
Server benar benar rusak dah,sering error connecting, logout sendiri, 5 menit disconnectnya bisa 3x, developernya ga niat fix servernya atau gimana dari dulu masalahnya kek gini "developer sampah"	server benar benar rusak dahsering error connecting logout sendiri menit disconnectnya bisa x developernya ga niat fix servernya atau gimana dari dulu masalahnya kek gini developer sampah	4,8% (9 dari 188 kata berhasil di Case folding)
tolong untuk developer saya ga main casino/reme/qeme seperti judi lainnya,tetapi saya di suspend padahal saya baru buat akun keesokan nya ke suspend tolong di perhatikan	tolong untuk developer saya ga main casinoremeqeme seperti judi lainnyatetapi saya di suspend padahal saya baru buat akun keesokan nya ke suspend tolong di perhatikan	7,6% (13 dari 171 kata berhasil di Case folding)

• Stopword Removal

Stopwords adalah kata-kata yang umum digunakan dalam suatu bahasa, namun memiliki nilai informasi yang rendah saat dilakukan analisis teks. Karena kemunculannya yang sangat sering di berbagai dokumen, kata-kata ini biasanya dihapus atau diabaikan dalam proses pengolahan teks, khususnya pada analisis berbasis Natural Language Processing (NLP). Menghilangkan stopwords bertujuan untuk menyederhanakan teks dan menyoroti kata-kata yang lebih penting dalam memahami makna atau isi dokumen. Hasil dari proses penghapusan stopwords dapat dilihat pada Tabel 2 berikut.

Tabel 2 Hasil Stopword Removal

Sebelum	Sesudah	Rasio
server benar benar rusak dahsering error connecting logout sendiri menit disconnectnya bisa x developernya ga niat fix servernya atau gimana dari dulu masalahnya kek gini developer sampah	server rusak dahsering error connecting logout menit disconnectnya x developernya ga niat fix servernya gimana kek gini developer sampah	83%(5 dari 6 kalimat berhasil di Stopword)
tolong untuk developer saya ga main casino/reme/qeme seperti judi lainnya,tetapi saya di suspend padahal saya baru buat akun keesokan nya ke suspend tolong di perhatikan	olong developer ga main casinoremeqeme judi lainnyatetapi suspend akun keesokan nya suspend tolong perhatikan	83%(5 dari 6 kalimat berhasil di Stopword)

• Tokenisasi

Tahap tokenisasi merupakan proses yang memecah suatu teks menjadi bagian-bagian kecil yang disebut token. Proses ini dilakukan setelah menyelesaikan tahap Stopword Removal. Hasil dari proses tokenisasi dapat dilihat pada tabel dan gambar di bawah ini.

Tabel 3 Hasil Tokenisasi

Sebelum	Sesudah	Rasio Tokenize
server rusak dahsering error connecting logout menit disconnectnya x developernya ga niat fix servernya gimana kek gini developer sampah	[server, rusak, dahsering, error, connecting, logout, menit, disconnectnya, x, developernya, ga, niat, fix, servernya, gimana, kek, gini, developer, sampah]	19 Tokenize
tolong developer ga main casino/reme/qeme seperti judi lainnya,tetapi suspend akun keesokan nya suspend tolong perhatikan	[tolong, developer, ga, main, casinoremeqeme, judi, lainnyatetapi, suspend, akun, keesokan, nya, suspend, tolong, perhatikan]	14 Tokenize

- Stemming

Stemming ialah tahap dalam pemrosesan teks yang berguna untuk menghapus imbuhan, baik awalan maupun akhiran, dari suatu kata untuk mengembalikannya ke bentuk dasar. Proses ini berfungsi menyederhanakan kata-kata sehingga kata-kata dengan akar yang sama dapat lebih mudah dikenali dalam analisis teks atau pemrosesan bahasa alami. Pada penelitian ini, proses stemming dilakukan dengan memanfaatkan pustaka Sastrawi untuk memetakan dan menguraikan kata-kata menjadi bentuk dasarnya. Hasil dapat dilihat pada tabel 4 dibawah ini.

Tabel 4 Hasil Stemming

Sebelum	Sesudah	Rasio
server rusak dahsering error connecting logout menit disconnectnya x developernya ga niat fix servernya gimana kek gini developer sampah	server_rusak,dahsering,error ,connecting,logout,menit, disconnectnya,x,developernya,ga,niat,fix, servernya,gimana,kek,gini,developer,samp ah	19 Stemmin g
tolong developer ga main casinoremeqem e judi lainnyatetapi suspend akun keesokan nya suspend tolong perhatikan	tolong,developer,ga,,main,casinoremeqeme ,judi,lainnyatetapi,suspend,akun,keesokan, nya ,suspend,tolong,perhatikan	14 Stemmin g

C. Transformasi Data

Transformasi data berperan penting dalam membantu algoritma mengidentifikasi pola yang terdapat dalam data. Beberapa metode yang diterapkan meliputi one-hot encoding, yang mengonversi data kategorikal menjadi format numerik; term frequency, yang menghitung seberapa sering sebuah kata muncul dalam teks; serta TF-IDF (Term Frequency-Inverse Document Frequency), yang memberikan bobot pada kata berdasarkan frekuensinya dalam dokumen tertentu dibandingkan dengan keseluruhan dataset. Proses transformasi ini bertujuan untuk menyiapkan data agar dapat dianalisis secara optimal, sehingga mampu memberikan wawasan yang lebih mendalam bagi pengembang dalam memahami kebutuhan pengguna dan meningkatkan kualitas layanan yang mereka tawarkan.

- TDF-IF

Proses ini melibatkan manipulasi data untuk menyesuaikannya dengan kebutuhan analisis yang akan dilakukan. Term Frequency (TF) digunakan untuk menghitung seberapa sering sebuah kata muncul dalam suatu dokumen, dengan cara membandingkan jumlah kemunculan kata tersebut terhadap total kata dalam dokumen tersebut. Sementara itu, Inverse Document Frequency (IDF) mengukur tingkat keunikan suatu kata dengan melihat seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen. Kata yang jarang ditemukan akan memiliki nilai IDF yang lebih tinggi, menandakan bahwa kata tersebut lebih informatif. Kombinasi kedua konsep ini menghasilkan TF-IDF, yang memberikan bobot pada setiap kata berdasarkan seberapa penting kata tersebut dalam suatu dokumen relatif terhadap keseluruhan dataset. Dalam implementasinya, terdapat dua pendekatan umum untuk menghitung TF-IDF. Dengan TfidfTransformer, proses dimulai dengan menghitung jumlah kata menggunakan CountVectorizer, dilanjutkan dengan perhitungan nilai IDF, dan akhirnya menghasilkan skor TF-IDF. Sementara dengan TF-IDF Vectorizer, ketiga langkah tersebut dilakukan sekaligus dalam satu proses yang lebih praktis dan efisien. Hasil dari perhitungan ini akan ditampilkan pada gambar berikut.

```

[ ] from sklearn.feature_extraction.text import CountVectorizer

vectorizer = CountVectorizer()
vectorizer.fit(X_train)

[ ] X_train = vectorizer.transform(X_train)
X_test = vectorizer.transform(X_test)

[ ] from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(tfidf_train, y_train)
    
```

Gambar 2 Pembobotan TDF-IF

D. Data Mining

Proses data mining merupakan pendekatan yang bertujuan untuk mengekstraksi pola atau informasi berharga dari kumpulan data yang besar dengan memanfaatkan berbagai metode dan algoritma. Fokus utama dari proses ini adalah mengidentifikasi pola tersembunyi yang dapat memberikan wawasan penting guna

mendukung pengambilan keputusan yang lebih baik. Dalam konteks penelitian ini, data mining diterapkan untuk menganalisis sentimen pengguna terhadap aplikasi Growtopia, dengan mengelompokkan ulasan ke dalam kategori positif dan negatif. Metode yang digunakan adalah Naïve Bayes, sebuah algoritma berbasis probabilitas yang efektif dalam klasifikasi teks. Proses analisis ini mencakup beberapa tahapan penting, mulai dari pengumpulan data, preprocessing untuk membersihkan dan menyiapkan data, ekstraksi fitur untuk memperoleh informasi yang relevan, pembagian data ke dalam kelompok pelatihan dan pengujian, pembangunan model klasifikasi, hingga evaluasi kinerja model guna mengukur tingkat akurasi yang dicapai.

• Split Data

Pembagian data yang menjadi dataset pelatihan dan pengujian penting untuk melatih dan menguji model secara efektif. Dataset pelatihan digunakan untuk mengenali pola, sedangkan dataset pengujian menilai kemampuan model dalam memprediksi data baru. Berbagai metode, seperti pembagian acak atau validasi silang, dapat digunakan sesuai kebutuhan analisis. Dalam penelitian ini, data dibagi dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian (test_size 0.3) menggunakan random state 0 untuk memastikan evaluasi yang objektif.

```
[ ] notebook data menjadi data training dan testing dengan test_size = 0.3 dan random state nya 0
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_clean['content'], data_clean['label'],
                                                test_size = 0.3,
                                                random_state = 0)

[ ] from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)

[ ] print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

[ ] (1295,)
(1295,)
(388,)
(388,)
```

Gambar 3 Hasil Split Data

• Klasifikasi Naive Bayes

Proses data mining merupakan langkah penting dalam menggali informasi berharga dari kumpulan data yang besar dan kompleks. Dengan memanfaatkan berbagai metode, algoritma, dan teknik analisis, proses ini bertujuan untuk menemukan pola tersembunyi atau hubungan tertentu yang dapat memberikan wawasan mendalam. Melalui pendekatan ini,

data yang awalnya tampak acak dan tidak beraturan dapat diolah menjadi informasi yang lebih bermakna. Dalam penelitian ini, metode Naïve Bayes diterapkan untuk mengklasifikasikan sentimen pengguna aplikasi Growtopia ke dalam kategori positif dan negatif. Selain itu, metode ini juga digunakan untuk mengukur tingkat akurasi model dalam mengidentifikasi pola sentimen berdasarkan ulasan yang dikumpulkan. Proses ini diharapkan dapat memberikan pemahaman yang lebih baik terkait persepsi pengguna terhadap aplikasi tersebut.

```
[ ] X_train = vectorizer.transform(X_train)
X_test = vectorizer.transform(X_test)

[ ] from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(tfidf_train, y_train)

[ ] X_train.toarray()

array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```

Gambar 4 Hasil Klasifikasi Naïve Bayes

E. Evaluasi

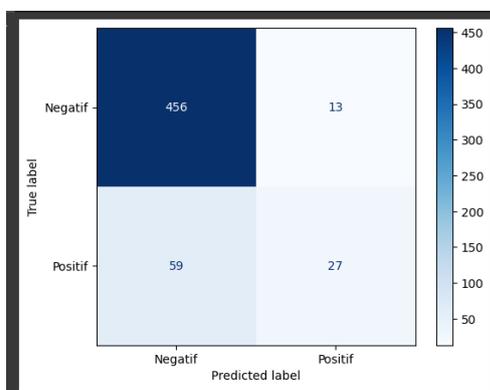
Setelah proses klasifikasi selesai, langkah berikutnya adalah tahap evaluasi. Tahapan ini berperan penting sebagai langkah akhir dalam penelitian, yang bertujuan untuk menilai seberapa baik algoritma Naïve Bayes dalam melakukan klasifikasi data. Evaluasi dilakukan dengan mengukur tingkat akurasi model menggunakan metode MultinomialNB. Berdasarkan hasil analisis, model tersebut menunjukkan tingkat akurasi sebesar 87%, yang mengindikasikan bahwa model mampu mengklasifikasikan sentimen ulasan dengan performa yang cukup baik.

```
MultinomialNB Accuracy: 0.8723021582733813
MultinomialNB Precision: 0.8859315589353612
MultinomialNB Recall: 0.9769392033542977
MultinomialNB f1_score: 0.9292123629112662
confusion_matrix:
[[466 11]
 [ 60 19]]
```

	precision	recall	f1-score	support
Negatif	0.89	0.98	0.93	477
Positif	0.63	0.24	0.35	79
accuracy			0.87	556
macro avg	0.76	0.61	0.64	556
weighted avg	0.85	0.87	0.85	556

Gambar 5 Hasil

Evaluasi menggunakan Confusion Matrix dilakukan dengan membagi data ke dalam dua kelompok, yaitu data pelatihan (training) dan data pengujian (testing). Dari keseluruhan data yang tersedia, sebanyak 556 data dialokasikan sebagai data pengujian. Hasil analisis melalui Confusion Matrix menunjukkan bahwa terdapat 27 data yang berhasil diidentifikasi dengan benar sebagai kelas positif (True Positive/TP). Di sisi lain, terdapat 13 data yang seharusnya masuk ke dalam kelas positif, namun keliru diklasifikasikan sebagai kelas negatif (False Negative/FN). Untuk kelas negatif, ditemukan 59 data yang seharusnya diklasifikasikan sebagai kelas negatif, tetapi salah diidentifikasi sebagai kelas positif (False Positive/FP). Sementara itu, sebanyak 456 data diklasifikasikan dengan benar sebagai kelas negatif (True Negative/TN). Hasil evaluasi ini memberikan gambaran jelas mengenai kinerja model dalam membedakan antara sentimen positif dan negatif pada ulasan pengguna aplikasi Growtopia. Rincian lebih lengkap dapat dilihat pada visualisasi Confusion Matrix berikut.



Gambar 6 Konfusi Matrix

Pembahasan

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna game Growtopia di Google Play Store menggunakan algoritma Naïve Bayes Classifier. Analisis ini didasarkan pada teori mengenai analisis sentimen dalam game online, yang menunjukkan bahwa ulasan pengguna dapat memberikan wawasan mendalam tentang

pengalaman bermain dan aspek teknis yang perlu diperbaiki. Menurut penelitian sebelumnya, analisis sentimen dalam game sering digunakan untuk mengidentifikasi pola kepuasan dan keluhan pengguna, sehingga pengembang dapat merancang strategi perbaikan yang lebih tepat sasaran.

Algoritma Naïve Bayes Classifier terpilih karena kemampuannya yang cukup dalam mengklasifikasikan teks berdasarkan probabilitas, dengan asumsi independensi antar fitur. Teori ini menyatakan bahwa meskipun asumsi independensi sering kali tidak sepenuhnya akurat dalam analisis teks, metode ini tetap efektif dalam menangani dataset yang besar dan bervariasi, seperti ulasan pengguna di platform digital. Studi sebelumnya juga menunjukkan bahwa Naïve Bayes Classifier digunakan dalam analisis sentimen karena kecepatan komputasi dan kemampuannya dalam menangani teks yang tidak terstruktur, meskipun memiliki keterbatasan dalam mengenali konteks yang lebih kompleks.

Dalam penelitian ini, data ulasan yang sebelumnya dikumpulkan melalui web scraping, yang kemudian diproses menggunakan teknik preprocessing, termasuk case folding, tokenizing, stopwords removal, dan stemming. Setelah itu, data diklasifikasikan menjadi positif dan negatif menggunakan Naïve Bayes Classifier, dan hasil klasifikasinya dievaluasi dengan Confusion Matrix.

Dari hasil analisis terbaru, model menunjukkan akurasi sebesar 87%, dengan presisi 88.59% dan recall 97.69%, menandakan bahwa algoritma Naïve Bayes mampu melakukan klasifikasi dengan cukup baik. Namun, setelah validasi ulang menggunakan dataset tambahan, ditemukan bahwa 29 ulasan dikategorikan sebagai True Positive (TP), 20 ulasan sebagai False Negative (FN), 23 ulasan sebagai False Positive (FP), dan 300 ulasan sebagai True Negative (TN).

Lebih lanjut, temuan menunjukkan bahwa ulasan positif umumnya mengapresiasi kreativitas dalam gameplay dan komunitas yang aktif, sedangkan ulasan negatif lebih dominan, dengan keluhan utama mengenai bug,

kesulitan login, dan gangguan koneksi. Implikasi dari hasil ini menunjukkan bahwa pengembang perlu lebih fokus pada perbaikan teknis, terutama dalam meningkatkan stabilitas server dan pengalaman pengguna dalam login.

Dengan menggunakan analisis sentimen ini, pengembang dapat lebih memahami kebutuhan dan ekspektasi pengguna, sehingga dapat menyusun strategi pengembangan yang lebih tepat sasaran. Selain itu, penelitian ini membuktikan bahwa algoritma Naïve Bayes dapat menjadi alat yang efektif dalam klasifikasi ulasan pengguna di platform digital, khususnya dalam domain game online.

5. KESIMPULAN

Penelitian mengenai hal penggunaan Naïve Bayes Classifier dalam menganalisis sentimen ulasan pengguna game Growtopia di Google Play Store, dapat kami disimpulkan bahwa algoritma ini efektif dalam mengklasifikasikan sentimen pengguna terhadap game tersebut. Proses analisis diawali dengan pengumpulan data ulasan pengguna, yang kemudian melalui tahapan preprocessing, seperti case folding, tokenizing, stopwords removal, dan stemming. Data yang telah diproses diklasifikasikan menjadi positif dan negatif. dengan evaluasi menggunakan Confusion Matrix.

Hasil evaluasi menunjukkan bahwa akurasi model mencapai 87%, dengan presisi sebesar 88.59% dan recall sebesar 97.69%, mengindikasikan bahwa Naïve Bayes mampu mengklasifikasikan sentimen dengan cukup baik. Namun, hasil evaluasi ulang dengan dataset tambahan menunjukkan bahwa model mengklasifikasikan 29 data sebagai True Positive (TP), 20 data sebagai False Negative (FN), 23 data sebagai False Positive (FP), dan 300 data sebagai True Negative (TN).

Lebih lanjut, ditemukan bahwa sentimen negatif lebih dominan dibandingkan dengan sentimen positif. Ulasan positif umumnya mengapresiasi kreativitas dalam gameplay serta interaksi komunitas, sementara ulasan negatif didominasi oleh keluhan terkait bug, kesulitan login, dan gangguan koneksi. Hal ini menunjukkan bahwa meskipun game memiliki fitur menarik, aspek teknisnya masih perlu perbaikan yang signifikan.

Hasil penelitian ini menegaskan bahwa analisis sentimen dapat menjadi alat yang efektif bagi pengembang dalam memahami kebutuhan pengguna. Dengan memahami pola sentimen pengguna, pengembang dapat lebih responsif dalam mengatasi permasalahan teknis dan menyusun strategi pengembangan game yang lebih sesuai dengan ekspektasi pemain.

Sebagai rekomendasi, disarankan agar pengembang lebih aktif dalam memonitor dan menanggapi ulasan pengguna, terutama dalam menangani bug dan kendala teknis yang sering dikeluhkan. Selain itu, penelitian di masa depan dapat mengeksplorasi algoritma lain, seperti Support Vector Machine (SVM) atau Deep Learning, untuk meningkatkan akurasi dan efektivitas analisis sentimen.

UCAPAN TERIMA KASIH

Peneliti mengungkapkan banyak terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyelesaian penelitian ini. Ucapan terima kasih khusus ditujukan kepada pembimbing yang telah memberikan arahan dan masukan yang diperlukan selama proses penelitian.

DAFTAR PUSTAKA

- [1] P. Aditya, A. Azzahra, and A. Wijaya, "Analisis Sentimen Pemain Subway Surf Melalui Metode Naive Bayes Menurut Ulasan Play Store," *Jurnal Simasi : Jurnal Ilmiah Sistem Informasi*, vol. 3, no. 2, pp. 267–275, 2023.
- [2] M. A. Nugroho, M. A. Nurwahid, K. M. Fauzan, L. Nurrahman, and F. Nopriani, "Peran E-business dalam Ekonomi Virtual Game Growtopia," vol. 3, no. 1, pp. 45–51, 2024, doi: 10.33379/jusifor.v3i1.4438.
- [3] Hariatih & Sukardi, "Transformasi Digital Dalam Perekonomian Modern," *Jurnal PenKoMi:kajian pendidikan & Ekonomi*, vol. 5, no. 1, pp. 62–74, 2022.
- [4] A. Kumar, R. Khorwal, and S. Chaudhary, "A survey on sentiment analysis using swarm intelligence," *Indian J Sci Technol*, vol. 9, no. 39, 2016, doi: 10.17485/ijst/2016/v9i39/100766.
- [5] Primandani Arsi, Pungkas Subarkah, and Bagus Adhi Kusuma, "Analisis Sentimen Game Genshin Impact pada Play Store Menggunakan Naïve Bayes Clasifier," *Jurnal Ilmiah Teknik Mesin, Elektro dan Komputer*, vol. 3, no. 1, pp.

- 161–170, 2023, doi: 10.51903/juritek.v3i1.1962.
- [6] T. Suci Pania and R. Hidayati, “KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Kecanduan Bermain Game online Pada Remaja Menggunakan Metode Naïve Bayes Classifier Berbasis Website,” *Media Online*, vol. 4, no. 5, pp. 2518–2526, 2024, doi: 10.30865/klik.v4i5.1782.
- [7] K. Khadijah, N. Sabilly, 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.v12i01.328.
- [8] Cep Lukman Rohmat, R. Rosdiana Ningsih, A. Rizki Rinaldi, and F. Fatturohman, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Canva Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2297–2302, 2024, doi: 10.36040/jati.v8i2.9139.
- [9] M. R. A. Surya, Martanto, and U. Hayati, “Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 75–80, 2024.
- [10] R. A. Saputra *et al.*, “Analisis Sentimen Aplikasi Tokocrypto Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Naïve Bayes,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 2028–2036, 2024, doi: 10.30865/klik.v4i4.1707.
- [11] M. Gamma, A. Hakim, and F. Irwiensyah, “Analisis Sentimen Terhadap Ulasan Pengguna Pada Aplikasi BCA Mobile Menggunakan Metode Naïve Bayes,” *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 911–921, 2024, doi: 10.47065/josh.v5i4.5343.
- [12] R. Zulfiqri, B. N. Sari, and T. N. Padilah, “ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MEDIA SOSIAL INSTAGRAM PADA SITUS GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4995.
- [13] X. Lian *et al.*, “SingPro : a kno wledg e base pro viding single-cell prot eomic data,” no. October 2023, pp. 552–561, 2024.
- [14] T. Agustiranti, A. K. Izzati Kurdiana, B. Al Ghiffari, E. Dwi Juniar, and D. Gita Purnama, “Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh),” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 297–305, 2024, doi: 10.55338/jikomsi.v7i1.2946.