

ANALISIS SENTIMEN ULASAN “OJOL THE GAME” DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN MODEL EKSTRAKSI FITUR TF-IDF UNTUK MENINGKATKAN KUALITAS GAME

Rafi Rahmadani^{1*}, Abdul Rahim², Rudiman³

¹Universitas Muhammadiyah Kalimantan Timur, Indonesia

²Universitas Muhammadiyah Kalimantan Timur, Indonesia

³Universitas Muhammadiyah Kalimantan Timur, Indonesia

Received: 14 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Sentiment analysis, Naïve bayes, TF-IDF, Ojol The Game, Google Play Store

Correspondent Email:

ar622@umkt.ac.id

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap game "OJOL THE GAME" di Google Play Store memanfaatkan algoritma Naive Bayes dan model ekstraksi fitur TF-IDF. Data ulasan dikumpulkan melalui teknik web scraping menggunakan Python, kemudian diproses dengan tahapan preprocessing meliputi pembersihan data, case folding, stop word removal, tokenizing, dan stemming. Data yang telah diproses kemudian dianalisis menggunakan algoritma Metode Naive Bayes digunakan untuk mengklasifikasikan sentimen positif dan negatif. Hasil penelitian mengindikasikan bahwa kombinasi antara algoritma Naive Bayes dan TF-IDF memberikan akurasi sebesar 94,12%, menunjukkan efektivitas tinggi dalam mengidentifikasi sentimen pengguna. Temuan ini memberikan wawasan berharga dalam memahami opini pengguna, meningkatkan kualitas game.

Abstract. This study aims to analyze user sentiment towards the game "OJOL THE GAME" on Google Play Store using the Naive Bayes algorithm and the TF-IDF feature extraction model. User review data was collected through web scraping techniques using Python, then processed through preprocessing stages including data cleaning, case folding, stop word removal, tokenizing, and stemming. The processed data was then analyzed using the Naive Bayes algorithm to classify positive and negative sentiments. The results of the study show that the combination of the Naive Bayes algorithm and TF-IDF yielded an accuracy of 94.12%, demonstrating high effectiveness in identifying user sentiment. These findings provide valuable insights into understanding user opinions and improving the quality of the game.

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat ternyata dapat membuat perubahan dalam segala lapisan kehidupan sehari-hari sehingga membangun kreativitas manusia untuk mendorong temuan baru dalam dunia teknologi [1]. Pengembangan industri game

tercermin dari pendapatan global yang terus meningkat dan terus naik pesat hingga tahun 2025 [2]. Saat ini tengah muncul salah satu game yaitu ojol the game yang membuat gempar di dunia virtual, sensasi menjadi driver ojek tidak hanya di rasakan dalam dunia nyata melainkan dapat dirasakan secara online. Ojol

the game dikembangkan oleh *Codexplore* developer asal Indonesia yang didirikan pada bulan januari tahun 2021 [3].

Popularitas game “Ojol The Game” semakin berkembang dan banyak orang yang memainkan game tersebut [4]. Saat ini game tersebut mendapatkan peringkat pertama game paling populer di playstore dengan unduhan sebanyak 10 juta sehingga banyak ulasan yang dikirimkan untuk aplikasi ini, baik ulasan positif maupun negatif. Analisis sentiment pada game “Ojol The Game” bertujuan untuk mendapatkan informasi dari ulasan atau komentar pengguna, yang terbagi menjadi dua yaitu positif dan negatif [5]. Analisis sentiment pada review pengguna adalah proses penyelidikan ulasan untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan [6].

Analisis sentiment dapat dilakukan dengan menggunakan analisis *Naïve Bayes*, algoritma yang memiliki akurasi yang cukup tinggi dan waktu yang efisien dibandingkan algoritma klasifikasi lainnya [7]. Dengan memahami opini dan pengalaman pengguna, pengembang game dapat melakukan perbaikan, peningkatan kualitas, penyesuaian strategi pengembangan untuk meningkatkan kepuasan dan pengalaman bermain pengguna game Ojol The Game.

Menurut penelitian yang dilakukan oleh Gilbert Darmawan dkk [7]. berjudul "Analisis Sentimen Ulasan Pengguna Aplikasi MyPertamina di Google Playstore dengan Metode *Naïve Bayes*" dengan jumlah ulasan 3948 data, dengan hasil ulasan pengguna MyPertamina tergolong *negative* dengan presentase nilai akurasi 91%, nilai *precision* 92%, dan *recall*nya 100%. Nilai tersebut menunjukkan bahwa klasifikasi algoritma *naïve bayes* dinilai cukup baik dalam pemrosesan data ulasan *play store*, karena hasil presentase akurasi mencapai 91%.

Penelitian sebelumnya juga dilakukan oleh Abdul Aziz dan Fauziah [8]. yang berjudul Analisis Sentimen untuk Mengidentifikasi Opini tentang Produk, Layanan, dan Kebijakan Perusahaan dengan Menggunakan Algoritma TF-IDF dan *SentiStrength* dimana hasil klasifikasi Tweet memiliki sentiment positif, negatif, atau netral. Pada data uji dengan kata kunci “*vaksin anak*” sebanyak 1000 tweet

diperoleh hasil sentiment positif 54%, sentiment negatif 20% dan sentiment netral 26%. Kemudian perbandingan dengan analisis data yang sama menggunakan algoritma *naïve bayes* mendapatkan hasil sentiment positif 55%, negatif 16% dan netral 29%. Sedangkan *Decision Tree* mendapatkan hasil positif 61%, negatif 14% dan netral 25%.

Menurut penelitian sebelumnya oleh Dianati Suci Putri dkk [9] yang berjudul Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier* dengan hasil penggunaan algoritma *naïve bayes* didapatkan *accuracy score* sebesar 0.8 atau 80% hal ini berarti system mampu memprediksi secara akurat dari total data testing sebesar 20%. Klasifikasi *tweet* mengenai DPR sebanyak 95 positif, 693 netral dan 758 negatif dari hasil *crawling* sebanyak 1546.

Berdasarkan uraian diatas peneliti tertarik untuk melakukan penelitian terkait bagaimana penerapan *naïve bayes* dan model ekstraksi fitur TF-IDF dalam melakukan analisis sentiment positif dan negatif pada “Ojol The Game” dan sejauh mana akurasi algoritma *naïve bayes* dan model ekstraksi fitur TF-IDF dalam menganalisis sentimen ulasan pengguna terhadap game “Ojol The Game” yang tengah viral sekarang ini sehingga hasil dari analisis yang dilakukan dapat memberikan informasi berharga bagi pengembang game.

2. TINJAUAN PUSTAKA

2.1 *Ojol The Game*

Ojol The Game merupakan game virtual yang dimana game ini terinspirasi dari dunia nyata yaitu Ojol (Ojek Online) ojol the game ini memberikan sensasi memainkan game Ojol online di dunia virtual kini sensasi menjadi driver ojek online tidak hanya di rasakan di dunia nyata, tetapi juga di dalam sebuah permainan yang sedang viral saat ini. (XI Max, 2024).

2.2 *Naïve Bayes*

klasifikasi *naïve bayes* merupakan metode untuk klasifikasi yang ringan dalam pemrosesan dan memiliki akurasi yang tinggi jika diaplikasikan ke klasifikasi teks. Metode klasifikasi merupakan metode dengan

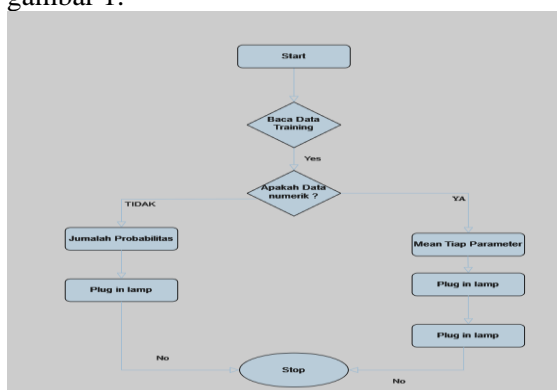
penggunaan algoritma yang dinilai sederhana namun memiliki akurasi kecepatan yang tinggi. [10].

2.3 TF – IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah proses yang mengubah data tekstual menjadi data numerik. Proses ini bertujuan untuk Menetapkan bobot untuk setiap kata atau fitur yang ada. Tahapan disini juga berguna untuk menunjukkan seberapa penting sebuah kata dalam setiap dokumen [11]

3. METODE PENELITIAN

Penelitian yang dilakukan melibatkan tahapan untuk menganalisis sentiment dari data Google Play Store dengan alur yang dimulai dari pengumpulan data ulasan menggunakan web Scraping dengan Python, dalam tahap ini data dari 1000 ulasan dikumpulkan. Selanjutnya dilakukan *Filtering* untuk mendapatkan informasi penting mengenai *username*, skor, tanggal dan isi ulasan. Setelahnya dilakukan pelabelan untuk mengkategorikan skor atau bintang pada game “Ojol The Game”. Tahap berikutnya adalah *Cleansing* atau pembersihan data, termasuk menangani *missing values* dengan mengabaikan tuple yang tidak lengkap. Selanjutnya dilakukan *Preprocessing teks* yang meliputi *Case Folding*, penghapusan kata umum (*Stopword removal*), *Tokenizing*, dan *Stemming*. Setelah preprocessing, data dibagi menjadi dua set: training set dan testing set. Penggunaan Naive Bayes dalam tahap ini sangat penting untuk menentukan sentimen positif dan negatif dalam ulasan game "Ojol The Game". Selanjutnya, metode pembobotan TF-IDF digunakan untuk mengevaluasi pentingnya sebuah kata dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Scraping

Scraping menggunakan web scraping untuk mengambil data sangat berguna dalam mengumpulkan data web terstruktur secara otomatis dengan menggunakan kode pemrograman Python. Proses ini memungkinkan ekstraksi data dalam jumlah besar secara cepat dan otomatis [12].

Data yang digunakan dalam penelitian ini adalah ulasan pengguna dari Google Play Store untuk permainan "OJOL THE GAME". ulasan aplikasi game "Ojol The Game" simulator, yang diperoleh dari scraping situs web Google Play Store dengan bantuan Google Colab. Sebanyak 1000 data ulasan berhasil diambil dan disimpan dalam format CSV untuk analisis sentimen lebih lanjut.

3.2. Preprocessing

Preprocessing data adalah langkah penting dalam analisis data dan *machine learning* yang bertujuan untuk mengubah data mentah menjadi data yang bersih dan siap digunakan dalam model analisis dan pembelajaran mesin. analisis atau pembuatan model. Berikut tahapan dari *Preprocessing* :

3.2.1. Filtering

Pada tahap ini, dilakukan proses *filtering* dan *cleansing* dengan menggunakan *select attribute* dan *remove duplicate*. Langkah ini bertujuan untuk mengambil kata-kata penting, membersihkan dokumen dari kata-kata yang tidak diperlukan, dan menghapus kata-kata yang duplikat. *Filtering* berfungsi untuk memfilter data mana saja yang akan di proses kemudian setelah dilakukan nya pemfilteran selanjutnya ke tahap pelabelan untuk melabelkan teks.

3.2.2. Pelabelan

Pada tahap analisis ini, pelabelan data ulasan dari game "OJOL THE GAME" dilakukan menggunakan skor penilaian untuk menentukan sentimen pengguna. Fungsi pelabelan mengkategorikan ulasan dengan skor 1 dan 2 sebagai negatif, dan skor 4 dan 5 sebagai positif. Awalnya, skor 3 tidak diberi label, sehingga beberapa nilai menjadi None. Setelah menambahkan kondisi untuk skor 3 sebagai 'Netral', seluruh data ulasan berhasil diberi label

dengan tepat, memungkinkan analisis sentimen yang akurat terhadap *feedback* pengguna.

3.2.3. Case Folding

Pada tahapan case folding, transform cases digunakan untuk mengubah teks menjadi huruf kecil. Tujuan dari case folding ini adalah untuk menyamaratakan penggunaan huruf kapital dalam teks.

3.2.4. Stop Word Removal

Stopword removal adalah proses menghapus kata-kata umum dari teks. Stopword adalah kata-kata yang sering muncul dalam teks namun tidak memberikan makna signifikan. suatu bahasa, seperti "dan", "di", "yang", yang biasanya tidak memberikan informasi signifikan dalam analisis teks.

3.2.5. Tokenizing

Tahap ini melakukan penghapusan terhadap semua karakter non alphabet, misalnya symbol, spasi, angka, tanda baca yang tidak penting.

3.2.6. Stemming

Proses dilanjutkan dengan menghilangkan imbuhan pada masing – masing kata sehingga menjadi kata – kata dasar misalnya kata – kata “berlari”, “berlarian”, dan “berlarilah” akan diubah menjadi bentuk dasar seperti “lari”. Ini membantu mengurangi variasi kata yang memiliki kata yang memiliki makna serupa dalam teks bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat contoh data yang telah di stemming [13].

3.2.7. Handling Missing Value-Ignore Tuple

Handling Missing Values – Ignore Tuple adalah metode untuk menangani data yang hilang dengan cara menghapus baris yang memiliki nilai hilang. Metode ini cocok digunakan ketika jumlah data yang hilang sangat sedikit dibandingkan dengan ukuran keseluruhan dataset. Tujuan tahapan *Handling Missing Value* ini untuk menghilangkan atau membersihkan nilai 3 atau null yang tidak di perlukan dalam data [9].

3.3. Split Data

Pemisahan data (*split data*) merujuk pada pembagian dataset menjadi dua subset: data latihan (*training data*) dan data uji (*testing data*). Proses ini umumnya digunakan dalam tugas data *mining* dan *machine learning* untuk mengevaluasi kinerja dan kemampuan generalisasi dari sebuah model. Data latihan

digunakan untuk melatih model, sedangkan data uji digunakan untuk mengukur seberapa baik model tersebut dapat membuat prediksi pada data yang belum pernah dilihat sebelumnya. Pembagian data ini penting untuk memastikan bahwa model tidak *overfitting* dan dapat melakukan generalisasi dengan baik pada data baru [14].

3.4. TF-IDF

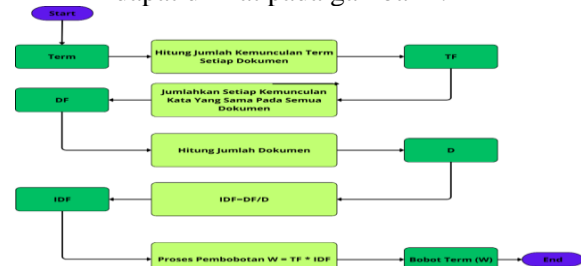
Penggunaan TF-IDF pada penelitian ini berfungsi untuk pembobotan suatu kata dan menghitung frekuensi kemunculan kata pada dokumen yang di gunakan berdasarkan 2 faktor :

3.4.1. Tern Frequency (TF)

Tern Frequency (TF) untuk mengukur seberapa sering suatu kata muncul dalam dokumen, semakin sering kata tersebut muncul maka semakin tinggi nilai TF-nya. $TF = \text{Jumlah pengulangan kata dalam dokumen} / \text{total jumlah kata dalam dokumen}$.

3.4.2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) digunakan untuk mengukur seberapa umum atau jarang suatu kata dalam seluruh korpus (kumpulan dokumen). Kata yang jarang muncul di korpus akan mendapatkan bobot yang lebih tinggi, IDF logaritma dari rasio jumlah seluruh dokumen dengan jumlah dokumen yang mengandung kata tersebut [8]. Bagan proses TF-IDF dapat dilihat pada gambar 2.



Gambar 2. Pembobotan TF-IDF

Sumber: Rsdm

Berdasarkan bagan di atas rumus TF-IDF adalah sebagai berikut:

$tf_{t,d}$

$$= \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d}$$

Inverse Document Frequency (IDF) mengukur seberapa penting sebuah kata dengan memperhitungkan seberapa sering kata tersebut muncul dalam semua dokumen. N melambangkan jumlah total dokumen dalam

kumpulan teks. Mengenai rumus TF-IDF score yakni :

$$idf_d = \log \frac{N}{n_t}$$

Keterangan :

t = Kata kunci, term

d = Dokumen

t.d = nilai TF-IDF untuk kata t dalam dokumen d

Tf = Banyaknya t (kata) yang dicari dalam dokumen

Idf = Banyak t kebalikan dari kata yang dicari

3.5. Naïve bayes

Teknik prediksi yang menggunakan probabilitas sederhana berdasarkan penerapan teorema Bayes dengan asumsi bahwa data bersifat independen, yaitu tidak saling berhubungan.

$$P(A|B) = \frac{P(A|B) \cdot P(A)}{P(B)}$$

Keterangan:

A =hipotesis data B adalah kelas khusus

B = data dengan kelas tidak dikenal

P(A|B) = probabilitas hipotesis A didasarkan pada kondisi B

P(B|A) = probabilitas hipotesis B didasarkan pada kondisi A

P(A) = probabilitas hipotesis A

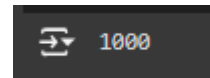
P(B) = probabilitas B

Jika ingin menggunakan algoritma *naïve bayes* alangkah baiknya jika menggunakan data latih yang cukup banyak, dikarenakan semakin banyak data yang digunakan maka semakin banyak hasil prediksi yang dihasilkan.

4. HASIL DAN PEMBAHASAN

4.1. Scraping/Pengambilan Data

Dalam penelitian ini, data yang dikumpulkan mencakup 1000 ulasan yang di-scraping. Namun, data yang telah di-scraping ini belum dilakukan pembersihan dan pelabelan. Pembersihan data diperlukan untuk



memastikan kualitas data yang akan dianalisis, sementara pelabelan diperlukan untuk mengidentifikasi kategori atau sentimen dari ulasan tersebut. Berikut data yang di dapatkan hasil scraping.

Gambar 3 Hasil Scraping

Data di bawah ini adalah hasil scraping yang belum mengalami pembersihan dan pelabelan. Data yang ditampilkan masih dalam bentuk aslinya, dan tabel 1 ini hanya menampilkan dua data sebagai contoh hasil scraping data yang di dapatkan dari hasil scraping 1000 data.

Tabel 1. Hasil Scraping

NO	revi wid	username	userimage	content	score	thumb pcount	versi	at	replycon tent	repliedat	versi
1	0f162 eeb- 8613- 458a- 8fd3- 1847e 9ebfc 08	Pris P	https://pla y- lh.googleu sercontent. com/a- /ALV-U..	bagusss dan ciri khas unik Indonesia , sangat h...	5	216	2.6.1	2024- 05-30 10:30: 31	None	None	2.6.1
2	b159a fcd- c1cb- 4c43- 937f- 27a3a 019e2 9d	Radit Kaunang	https://pla y- lh.googleu sercontent. com/a- /ALV-U..	Tolong dikasih efek ataupun apalah biar ga nge.	3	129	2.6.1	2024- 05-14 09:55: 08	None	None	2.6.1

4.2. Preprocessing

Pada tahap preprocessing, data yang telah di-scraping akan diproses lebih lanjut untuk dibersihkan dan disiapkan. mengalami serangkaian langkah pembersihan dan

pemrosesan untuk mempersiapkannya sebelum analisis lebih lanjut dilakukan. Langkah ini sangat penting untuk memastikan kualitas dan keakuratan hasil analisis.

4.2.1. Filtering

Pada tahap filtering, data ulasan yang telah di-scraping kemudian difilter untuk menentukan data apa saja yang perlu diambil. Pada tahap ini, ulasan diseleksi untuk memastikan hanya data yang relevan seperti username, skor, tanggal, dan isi ulasan yang diambil, seperti ditunjukkan pada gambar 4 Kodingan Filtering1.

```
[ ] my_df = sorted_df[['userName', 'score','at', 'content']]
```

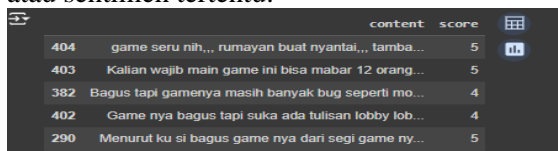
Gambar 4 Code Filtering

Setelah data yang dipilih pada tahap sebelumnya dirasa kurang, data difilter lagi untuk tahap pemrosesan. Pada tahap ini, hanya data skor (bintang) dan isi ulasan (content) yang diambil, seperti ditunjukkan pada gambar 5 di bawah ini yang menampilkan data yang sudah difilter.

```
[ ] my_df=my_df[['content', 'score']]
```

Gambar 5 Code Filtering

Setelah melakukan filtering, data yang didapatkan akan seperti pada gambar 6 Hasil Filtering Content dan Skor. Setelah data dirasa sudah sesuai, kita lanjutkan ke tahap pelabelan, di mana data diberi label sesuai dengan kategori atau sentimen tertentu.

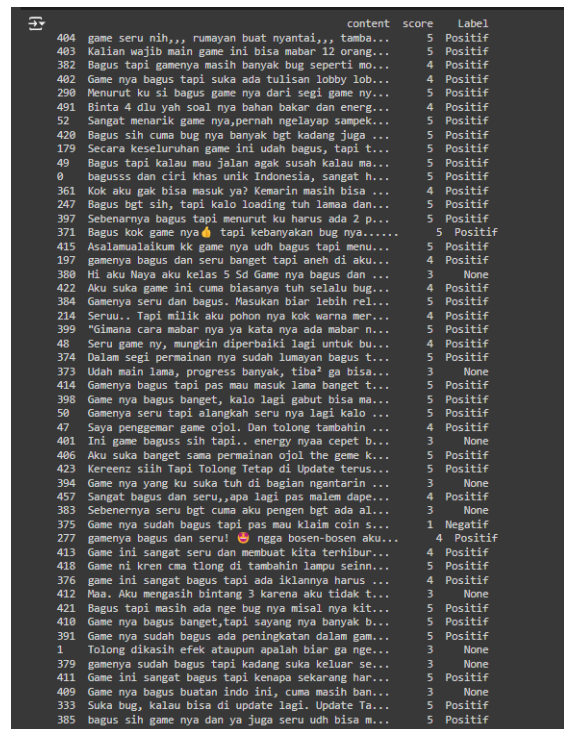


	content	score
404	game seru nih,,, rumayan buat nyantai,,, tamba...	5
403	Kalian wajib main game ini bisa mabar 12 orang...	5
382	Bagus tapi gamenya masih banyak bug seperti mo...	4
402	Game nya bagus tapi suka ada tulisan lobby lob...	4
290	Menurut ku si bagus game nya dari segi game ny...	5

Gambar 6 Hasil Filtering

4.2.2. Pelabelan

Pada tahap pelabelan data, ulasan yang telah melalui preprocessing dan filtering diberi label sesuai dengan kategori bintang yang diperoleh. Ulasan dengan penilaian bintang 1 hingga 2 dianggap negatif, sementara ulasan dengan penilaian bintang 4 dan 5 dianggap positif. Ulasan dengan penilaian bintang 3 tidak dimasukkan dalam analisis karena dianggap netral atau kurang konsisten. Gambar 7 menunjukkan data ulasan yang telah dilabeli, menampilkan 50 baris ulasan dengan bintang dan kategori yang sesuai. Hanya ulasan bintang 1, 2, 4, dan 5 yang digunakan dalam analisis sentiment.

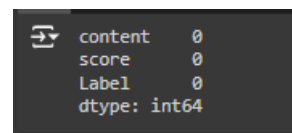


	content	score	Label
404	game seru nih,,, rumayan buat nyantai,,, tamba...	5	Positif
403	Kalian wajib main game ini bisa mabar 12 orang...	5	Positif
382	Bagus tapi gamenya masih banyak bug seperti mo...	4	Positif
402	Game nya bagus tapi suka ada tulisan lobby lob...	4	Positif
290	Menurut ku si bagus game nya dari segi game ny...	5	Positif
491	Binta 4 dlu yah soal nya bahan bakar dan energ...	4	Positif
52	Sangat menarik game nya,pernah ngelayap sampek...	5	Positif
420	Bagus sih cuma bug nya banyak bgt kadang juga ...	5	Positif
179	Secara keseluruhan game ini udah bagus, tapi t...	5	Positif
49	Bagus tapi kalau mau jalan agak susah kalau ma...	5	Positif
0	baguss dan ciri khas unik Indonesia, sangat h...	5	Positif
361	Kok aku gak bisa masuk ya? Kemarin masih bisa ...	4	Positif
247	Bagus bgt sih, tapi kalo loading tuh lama dan...	5	Positif
397	Sebenarnya bagus tapi menurut ku harus ada 2 p...	5	Positif
371	Bagus kok game nya 🍌 tapi kebanyakan bug nya....	5	Positif
415	Assalamualaikum kk game nya udh bagus tapi men...	5	Positif
197	gamenya bagus dan seru banget tapi aneh di aku...	4	Positif
380	Hi aku Naya aku kelas 5 SD Game nya bagus dan ...	3	None
422	Aku suka game ini cuma biasanya tuh selalu bug...	4	Positif
384	Gamenya seru dan bagus. Masukan biar lebih rel...	5	Positif
214	Seruu.. Tapi milik aku pohon nya kok warna mer...	4	Positif
399	"Gimana cara mabar nya ya kata nya ada mabar n...	5	Positif
48	Seru game ny, mungkin diperbaiki lagi untuk bu...	4	Positif
374	Dalam segi permainan nya sudah lumayan bagus t...	5	Positif
373	Udah main lama, progress banyak, tiba-tiba ga bisa...	3	None
414	Gamenya bagus tapi pas mau masuk lama banget t...	4	Positif
398	Game nya bagus banget, kalo lagi gabut bisa ma...	5	Positif
50	Gamenya seru tapi alangkah seru nya lagi kalo ...	5	Positif
47	Saya penggemar game ojol. Dan tolong tambahn ...	4	Positif
401	Ini game baguss sih tapi.. energy nya cepet bu...	3	None
406	Aku suka banget sama permainan ojol the game k...	5	Positif
423	Kereenz siih Tapi Tolong Tetap di Update terus...	5	Positif
384	Game nya yang ku suka tuh di bagian ngantarin ...	3	None
457	Sangat bagus dan seru,apa lagi pas maten dape...	4	Positif
383	Sebenarnya seru bgt cuma aku pengen bgt ada al...	3	None
375	Game nya sudah bagus tapi pas mau kllam coin s...	1	Negatif
277	Gamenya bagus dan seru! 🍌 ngga bosan-bosan aku...	4	Positif
413	Game ini sangat seru dan membuat kita terhibur...	4	Positif
418	Game ni kren cma tlong di tambahn lampu sein...	5	Positif
376	game ini sangat bagus tapi ada iklannya harus ...	4	Positif
412	Man. Aku mengasih bintang 3 karena aku tidak t...	3	None
421	Bagus tapi masih ada nge bug nya misal nya kit...	5	Positif
410	Game nya bagus banget,tapi sayang nya banyak b...	5	Positif
391	Game nya sudah bagus ada peningkatan dalam gam...	5	Positif
1	Tolong dikasih efek ataupun apalah biar ga nge...	3	None
379	gamenya sudah bagus tapi kadang suka keluar se...	3	None
411	Game ini sangat bagus tapi kenapa sekarang har...	5	Positif
409	Game nya bagus buatan indo ini, cuma masih ban...	3	None
333	Suka bug, kalau bisa di update lagi! Update Ta...	5	Positif
385	bagus sih game nya dan ya juga seru udh bisa m...	5	Positif

Gambar 7 Hasil Pelabelan

4.2.3. Cleansing atau Pembersihan Data

Pada tahap ini, digunakan metode handling missing values dengan menghapus baris yang memiliki nilai yang hilang (ignore tuple). Langkah ini termasuk dalam proses pembersihan data untuk memastikan data yang digunakan dalam analisis lebih lengkap dan akurat. Selanjutnya, jumlah nilai yang hilang ('NaN') di setiap kolom dihitung dan ditampilkan untuk memverifikasi bahwa nilai hilang dalam kolom ('label') sudah dihapus gambar 8.



```
content    0
score      0
Label      0
dtype: int64
```

Gambar 8. Hasil Pembersihan Data

4.2.4. Case Folding

Pada tahap case folding ini proses pengolahan teks di mana semua karakter dalam teks diubah menjadi huruf kecil. Sepeti 'Game Ini Seru Fun' setelah menggunakan case folding menjadi 'game ini seru fun' setelah data yang kita proses menggunakan case folding ini hasil seperti di bawah ini

	content	score	Label	text_clean
0	game seru nih... rumayan buat nyantai... tamba...	5	Positif	game seru nih rumayan buat nyantai tambahan aj...
1	Kalian wajib main game ini bisa mabar 12 orang...	5	Positif	kalian wajib main game ini bisa mabar orang c...
2	Bagus tapi gamenya masih banyak bug seperti mo...	4	Positif	bagus tapi gamenya masih banyak bug seperti mo...
3	Game nya bagus tapi suka ada tulisan lobby lob...	4	Positif	game nya bagus tapi suka ada tulisan lobby lob...
4	Menurut ku si bagus game nya dari segi game ny...	5	Positif	menurut ku si bagus game nya dari segi game ny...
5	Binta 4 du yah soal nya bahan bakar dan energi...	4	Positif	binta 4 du yah soal nya bahan bakar dan energi...
6	Sangat menarik game nya.pemah ngelayap sampek...	5	Positif	sangat menarik game nyapemah ngelayap sampek...
7	Bagus sih cuma bug nya banyak bgt kadang juga...	5	Positif	bagus sih cuma bug nya banyak bgt kadang juga...
8	Secara keseluruhan game ini udah bagus, tapi l...	5	Positif	secara keseluruhan game ini udah bagus tapi lo...
9	Bagus tapi kalau mau jalan agak susah kalau ma...	5	Positif	bagus tapi kalau mau jalan agak susah kalau ma...

Gambar 9. Hasil Case Folding

4.2.5. Stop Word Removal

Pada tahap pemrosesan stop word removal ini kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam Bahasa Indonesia “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan stopwords yaitu dengan menghapus kata – kata yang memiliki informasi rendah sebuah teks. Pada gambar 10 di bawah ini adalah hasil pemrosesan stop word removal.

	content	score	Label	text_clean	text_stopword
0	game seru nih... rumayan buat nyantai... tamba...	5	Positif	game seru nih rumayan buat nyantai tambahan aj...	game seru nih rumayan buat nyantai tambahan aj...
1	Kalian wajib main game ini bisa mabar 12 orang...	5	Positif	kalian wajib main game ini bisa mabar orang c...	wajib main game mabar orang mabar orang c...
2	Bagus tapi gamenya masih banyak bug seperti mo...	4	Positif	bagus tapi gamenya masih banyak bug seperti mo...	bagus gamenya bug motor hilang nabrak dinding...
3	Game nya bagus tapi suka ada tulisan lobby lob...	4	Positif	game nya bagus tapi suka ada tulisan lobby lob...	game nya bagus suka tulisan lobby lobby lob...
4	Menurut ku si bagus game nya dari segi game ny...	5	Positif	menurut ku si bagus game nya dari segi game ny...	ku si bagus game nya dari segi game nya...
5	Binta 4 du yah soal nya bahan bakar dan energi...	4	Positif	binta 4 du yah soal nya bahan bakar dan energi...	binta 4 du yah nya bahan bakar energi nya capai...
6	Sangat menarik game nya.pemah ngelayap sampek...	5	Positif	sangat menarik game nyapemah ngelayap sampek...	menarik game nyapemah ngelayap sampek rumaya...
7	Bagus sih cuma bug nya banyak bgt kadang juga...	5	Positif	bagus sih cuma bug nya banyak bgt kadang juga...	bagus sih bug nya bgt kadang kadang kadang s...
8	Secara keseluruhan game ini udah bagus, tapi l...	5	Positif	secara keseluruhan game ini udah bagus tapi lo...	game udah bagus tolong ditambahin fitur keun i...
9	Bagus tapi kalau mau jalan agak susah kalau ma...	5	Positif	bagus tapi kalau mau jalan agak susah kalau ma...	bagus jalan susah mabar gampang jalan mabar...
10	bagusss dan dari khas unik indonesia, sangat h...	5	Positif	bagusss dan dari khas unik indonesia sangat hi...	bagusss dari khas unik indonesia hebat game ba...
11	Kok aku gak bisa masuk ya? Kemarin masih bisa...	4	Positif	kok aku gak bisa masuk ya kemarin masih bisa t...	gak masuk ya kemarin main gak masuk dan pakek...
12	Bagus bgt sih, tapi kalo boading tuh lamaa dan...	5	Positif	bagus bgt sih tapi kalo boading tuh lamaa dan...	bagus bgt sih kalo boading tuh lamaa boading ka...
13	Sebenarnya bagus tapi menurut ku harus ada 2 p...	5	Positif	sebenarnya bagus tapi menurut ku harus ada pe...	bagus ku perubahan grafik nya bakin ya tolong...
14	Bagus kok game nya... tapi kebanyakan bug nya...	5	Positif	bagus kok game nya tapi kebanyakan bug nya...	bagus game nya kebanyakan bug nya ditambahin l...
15	Asalamualaikum hi game nya udi bagus tapi men...	5	Positif	asalamualaikum hi game nya udi bagus tapi men...	asalamualaikum hi game nya udi bagus menaruh...
16	gamenya bagus dan seru banget tapi aneh di aku...	4	Positif	gamenya bagus dan seru banget tapi aneh di aku...	gamenya bagus seru banget aneh aku ga dapet...
17	Aku suka game ini cuma bawanya tuh sedikit bag...	4	Positif	aku suka game ini cuma bawanya tuh sedikit bag...	suka game tuh bug trus keek kalo ditambah...
18	Gamenya seru dan bagus. Mainan bar lebih reli...	5	Positif	gamenya seru dan bagus. Mainan bar lebih reli...	gamenya seru bagus mainan bar lebih reli...

Gambar 10. Hasil Stop Word Removal

4.2.6. Tokonezing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisis. Kata, angka, simbol, tanda baca, dan entitas penting lainnya dapat dianggap sebagai token. Pada gambar 11 di bawah ini data teks ulasan yang telah dilakukan nya pemrosesan tokenizing.

	content	score	Label	text_clean	text_stopword	text_tokens
0	game seru nih... rumayan buat nyantai... tamba...	5	Positif	game seru nih rumayan buat nyantai tambahan aj...	game seru nih rumayan buat nyantai tambahan aj...	(game, seru, nih, rumayan, nyantai, tamba...
1	Kalian wajib main game ini bisa mabar 12 orang...	5	Positif	kalian wajib main game ini bisa mabar orang c...	wajib main game mabar orang mabar orang c...	(wajib, main, game, mabar, orang, mabar...
2	Bagus tapi gamenya masih banyak bug seperti mo...	4	Positif	bagus tapi gamenya masih banyak bug seperti mo...	bagus gamenya bug motor hilang nabrak dinding...	(bagus, gamenya, bug, motor, hilang, nabrak, d...
3	Game nya bagus tapi suka ada tulisan lobby lob...	4	Positif	game nya bagus tapi suka ada tulisan lobby lob...	game nya bagus suka tulisan lobby lobby lob...	(game, nya, bagus, suka, tulisan, lobby, lobby...
4	Menurut ku si bagus game nya dari segi game ny...	5	Positif	menurut ku si bagus game nya dari segi game ny...	ku si bagus game nya dari segi game nya...	(ku, si, bagus, game, nya, dari, segi, game, nya, ba...

Gambar 11. Hasil Tokonezing

4.2.7. Stemming

Proses stemming dalam analisis ulasan ini adalah langkah pemrosesan yang mengubah kata-kata yang terinfeksi atau memiliki akhiran menjadi bentuk dasarnya. Tujuan dari stemming adalah untuk mengurangi kata – kata yang berbeda ke bentuk dasarnya sehingga variasi kata dengan mana yang sama dapat diperlakukan sebagai satu kata. Misalnya, kata – kata seperti “running”, “runs”, dan “ran”, akan reduksi menjadi “run”. Pada gambar 12 kata yang di proses setiap kata sebanyak 3087 ini adalah data ulasan yang di proses steming.

3087	-----
1	: game : game
2	: seru : seru
3	: nih : nih
4	: rumayan : rumayan
5	: nyantai : nyantai
6	: tambahan : tambah
7	: aja : aja
8	: tolong : tolong
9	: motornya : motor
10	: ditambah : tambah
11	: harga : harga
12	: tarifnya : tarif
13	: sesuaikan : sesuai
14	: yg : yg
15	: diindonesia : indonesia
16	: maju : maju
17	: wajib : wajib
18	: main : main
19	: mabar : mabar
20	: orang : orang
21	: mabarnya : mabarnya
22	: liat : liat
23	: kiri : kiri
24	: gambar : gambar
25	: pala : pala
26	: tinggal : tinggal
27	: picit : picit
28	: server : server
29	: terseher : seher
30	: ngasih : ngasih
31	: tau : tau
32	: kalo : kalo
33	: bug : bug
34	: gak : gak
35	: kawan : kawan
36	: terima : terima
37	: kasih : kasih
38	: support : support
39	: bagus : bagus
40	: gamenya : gamenya
41	: motor : motor
42	: hilang : hilang
43	: nabrak : nabrak
44	: dinding : dinding
45	: tim : tim
46	: codexlore : codexlore
47	: mohon : mohon
48	: perbaiki : baik
49	: bugnya : bugnya
50	: ya : ya
51	: nya : nya
52	: suka : suka
53	: tulisan : tulis
54	: lobby : lobby

Gambar 12. Hasil Stemming

4.2.8. Spilting Data

Pada tahap ini, digunakan metode handling missing values dengan menghapus baris yang memiliki nilai yang hilang (ignore tuple). Langkah ini termasuk dalam proses pembersihan data untuk memastikan data yang digunakan dalam analisis lebih lengkap dan akurat. Selanjutnya, jumlah nilai yang hilang ('NAN') di setiap kolom dihitung dan

ditampilkan untuk memverifikasi bahwa nilai yang hilang sudah dihapus. Pada tahap splitting data, Data dibagi menjadi dua bagian: data pelatihan dan data pengujian, dengan parameter `test_size = 0.2` dan `random_state = 0`. Fungsi `train_test_split` digunakan untuk memisahkan dataset menjadi dua subset: satu untuk melatih model dan yang lainnya untuk mengujinya. Parameter `test_size = 0.2` menunjukkan bahwa 20% dari data akan digunakan untuk pengujian, sementara `random_state = 0` memastikan hasil yang konsisten. memastikan bahwa pembagian data acak dapat direproduksi untuk konsistensi hasil. Ini adalah langkah penting dalam pemodelan machine learning untuk memastikan bahwa model dapat diuji dengan data yang belum pernah diproses sebelumnya guna mengukur kinerjanya secara akurat. Gambar 13 menunjukkan kode sumber pemrosesan dalam splitting data sebelum masuk ke proses pembobotan TF-IDF.

```
[ ] #membagi data menjadi data training dan testing dengan test_size = 0.2 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 = 0.2,
                                                  random_state = 0)
```

Gambar 13. Code Splitting Data

4.2.9. Pembobotan TF-IDF

Pada penggunaan pembobotan TF – IDF ini digunakan untuk menilai pentingnya kata dalam suatu dokumen relative terhadap Kumpulan dokumen lainnya. Pada gambar 14 di bawah ini menunjukan source code dari TF – IDF pada pemrosesan sebelum lanjut pada pemrosesan naïve bayes sebelum langsung ke naïve bayes.

```
pembobotan tf-idf

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)
```

Gambar 14. Code TF-IDF

4.2.10. Naïve Bayes

Pada pemrosesan Naïve Bayes pada data ulasan game "Ojol the Game," hasil yang diperoleh cukup memuaskan dengan akurasi sebesar 94.12%. Dari 1000 ulasan yang digunakan, setelah memfilter ulasan berdasarkan bintang 1, 2, 4, dan 5, rata-rata ulasan menunjukkan kecenderungan positif. Untuk ulasan negatif, precision, recall, dan f1-score semuanya bernilai 0.00 dengan support

sebanyak 6. Sedangkan untuk ulasan positif, precision bernilai 0.94, recall 1.00, dan f1-score 0.97 dengan support sebanyak 96.

```
MultinomialNB Accuracy:94.12%
MultinomialNB Precision: 0.0
MultinomialNB Recall: 0.0
MultinomialNB f1_score: 0.0
confusion_matrix:
[[ 0  6]
 [ 0 96]]
```

	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	6
Positif	0.94	1.00	0.97	96
accuracy			0.94	102
macro avg	0.47	0.50	0.48	102
weighted avg	0.89	0.94	0.91	102

Gambar 15. Hasil Naïve Bayes

Hasil ini ditampilkan pada gambar 15 yang menunjukkan akurasi analisis sentimen ulasan "Ojol the Game". Algoritma Naïve Bayes yang digunakan bersama dengan TF-IDF menghasilkan akurasi yang sangat baik, dengan 20% data digunakan sebagai data uji, sementara 80% sisanya digunakan sebagai data latih. Pada penelitian ini, akurasi mencapai nilai 94.12%, sementara F1- score bernilai 0.94. Macro average precision mendapatkan nilai 0.47, recall 0.50, dan f1-score 0.48 dengan support sebanyak 102. Sedangkan untuk weighted average, precision mendapatkan nilai 0.89, recall 0.94, dan f1-score 0.91 dengan support sebanyak 102. Dengan hasil ini, ulasan "Ojol the Game" menunjukkan sentimen positif yang dapat digunakan untuk mengembangkan dan meningkatkan kualitas permainan lebih lanjut.

5. KESIMPULAN

- Hasil penelitian menunjukkan bahwa analisis sentimen menggunakan algoritma Naive Bayes dan pembobotan TF-IDF pada ulasan game "OJOL THE GAME" menghasilkan akurasi sebesar 94,12%.
- Data ulasan dikumpulkan melalui web scraping menggunakan Python dan diolah dengan teknik preprocessing yang mencakup pembersihan data dan pelabelan. Tingginya akurasi ini mengindikasikan bahwa kombinasi Naive Bayes dan TF-IDF efektif dalam mengidentifikasi sentimen positif dan negatif dari ulasan pengguna, memberikan wawasan berharga bagi pengembang untuk

meningkatkan kualitas game berdasarkan umpan balik pengguna.

UCAPAN TERIMA KASIH

Terimakasih kepada pihak – pihak yang telah membantu saya dalam penyelesaian jurnal ini.

DAFTAR PUSTAKA

- [1] Sapto Irawan and Dina Siska W., “Faktor-faktor Yang Mempengaruhi Kecanduan Game Online Peserta Didik,” *J. Konseling Gusjigang*, vol. 7, no. 1, pp. 9–19, 2021.
- [2] M. Mustofa, J. L. Putra, and C. Kesuma, “Penerapan Game Development Life Cycle Untuk Video Game Dengan Model Role Playing Game,” *Comput. Sci.*, vol. 1, no. 1, pp. 27–34, 2021, doi: 10.31294/coscience.v1i1.158.
- [3] XI Max, “Ojol The Game: Sensasi Menjadi Driver Ojek Online Di Dunia Virtual – Berita dan Informasi,” *umsu.ac.id*. Accessed: Mar. 04, 2024. [Online]. Available: <https://umsu.ac.id/berita/ojol-the-game-sensasi-menjadi-driver-ojek-online-di-dunia-virtual/>
- [4] R. W. Hayuningtyas, “Benarkah Uang Ojol The Game Bisa Dicairkan? Ini Faktanya!,” *MOMS MONEY.id*. Accessed: Mar. 09, 2024. [Online]. Available: <https://momsmoney.kontan.co.id/news/benarka-h-uang-ojol-the-game-bisa-dicairkan-ini-faktanya#>:
- [5] F. N. H. Fauzan Setya Ananto1*, “Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 450–462, 2023, doi: 10.31539/intecom.v6i1.6117.
- [6] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [7] Gilbert, Syariful Alam, and M. Imam Sulistyo, “Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123/storage.v2i3.2333.
- [8] A. Aziz, “Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. s1, p. 115, 2022.
- [9] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [10] L. B. Ilmawan and M. A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [11] A. N. Hasanah, B. N. Sari, U. S. Karawang, T. Timur, and J. Barat, “JASA OJEK ONLINE MAXIM PADA GOOGLE PLAY,” vol. 12, no. 1, pp. 90–96, 2024.
- [12] M. Kopravi and W. S. Putra, “Implementasi Web Scraping pada Google Cendekia sebagai Sarana Profiling Penelitian Dosen,” *Sci. Tech J. Ilmu Pengetah. dan Teknol.*, vol. 9, no. 1, pp. 59–72, 2023, doi: 10.30738/st.vol9.no1.a14188.
- [13] N. Widya Utami and M. Artana, “Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma K-Nearest Neighbor,” *J. Inform. Teknol. dan Sains*, vol. 4, no. 2, pp. 140–148, 2022, doi: 10.51401/jinteks.v4i2.2034.
- [14] E. Muningsih, “Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan SpliT DATA,” vol. 16, no. 1, pp. 113–118, 2022.