



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 5 Tahun 2024 Page 3177-3193

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Penerapan Fitur Ekstraksi TF-IDF untuk Analisis Sentimen Ulasan Game Bus Simulator Indonesia dengan Algoritma Naive Bayes

Thoriq Ikhwan Alfawas¹, Abdul Rahim^{2✉}, Rudiman³

Universitas Muhammadiyah Kalimantan Timur

Email: ar622@umkt.ac.id^{2✉}

Abstrak

Di era digital saat ini, permainan video khususnya game simulasi telah menjadi hiburan yang sangat populer di berbagai kalangan. Bus Simulator Indonesia juga menawarkan pengalaman bermain yang autentik dengan mempersembahkan detail-detail khas Indonesia, mulai dari lingkungan sekitar hingga perilaku lalu lintas yang serupa dengan kondisi di negara ini. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap BUSSID dengan menggunakan ekstraksi fitur TF-IDF dan algoritma Naive Bayes. Kesulitan dalam melakukan pembobotan dengan TF-IDF Vectorizer termasuk memahami konsep TF dan IDF serta menangani teks yang tidak biasa seperti Bahasa non-standar atau singkatan. Analisis dilakukan dengan menguji lima variasi pembagian data yang berbeda untuk menentukan akurasi tertinggi dan terendah dari model Multinomial Naive Bayes setelah menggunakan fitur TF-IDF. Berdasarkan hasil pengujian manual dengan 5 variasi data latih dan uji menunjukkan bahwa model Multinomial Naive Bayes memberikan akurasi tertinggi sebesar 85% pada variasi 90%:10%, dengan nilai precision, recall, dan F1-score yang konsisten. Variasi lainnya menunjukkan akurasi sedikit lebih rendah, antara 82% hingga 84%, namun tetap menunjukkan kinerja model yang baik. Hasil ini memberikan gambaran yang jelas tentang efektivitas model Naive Bayes dalam menganalisis sentimen ulasan pengguna setelah penerapan pembobotan TF-IDF.

Kata Kunci: *Bus Simulator Indonesia, Analisis Sentimen, TF-IDF, Naive Bayes.*

Abstract

In the current digital era, video games, especially simulation games, have become very popular entertainment in various circles. Bus Simulator Indonesia also offers an authentic playing experience by presenting typical Indonesian details, the surrounding environment to traffic behavior that is similar to conditions in this country. This research aims to analyze the sentiment user reviews of BUSSID using TF-IDF feature extraction and the Naive Bayes algorithm. Difficulties in performing weighting with TF-IDF Vectorizer include understanding TF and IDF concepts as well as dealing with unusual text such as non-standard language or abbreviations. The analysis was carried out by testing five different variations data division to determine the highest and lowest accuracy of the Multinomial Naive Bayes model after using TF-IDF feature. Based on the results of manual testing with 5 variations training and test data, it shows that the Multinomial Naive Bayes model provides the highest accuracy of 85% at a variation of 90%:10%, consistent precision, recall and F1-score values. Other variations show slightly lower accuracy, between 82% and 84%, but still show good model performance. These results provide a clear picture of the effectiveness of the Naive Bayes model in analyzing user review sentiment after applying TF-IDF weighting.

Keyword: *Bus Simulator Indonesia, Sentiment Analysis, TF-IDF, Naive Bayes.*

PENDAHULUAN

Dalam era digital saat ini permainan video atau game telah menjadi bentuk hiburan yang paling populer diberbagai kalangan. Salah satu genre game yang menarik perhatian banyak pengguna adalah game simulasi. Di Indonesia sendiri perkembangan industri game di Indonesia mengalami pertumbuhan yang pesat seiring dengan meningkatnya minat masyarakat terhadap permainan digital. Salah satu genre game yang populer adalah simulasi dimana pemain dapat merasakan pengalaman dunia nyata melalui permainan digital. Industri game mobile di Indonesia telah mengalami pertumbuhan yang signifikan dalam beberapa tahun terakhir, dengan berbagai jenis permainan yang diminati oleh pengguna termasuk game simulasi. Pada jurnal menjelaskan game simulasi merupakan penggambaran dinamis dari proses nyata yang disederhanakan menjadi elemen yang signifikan. Komponen permainan seperti tujuan, aturan, dan skor menciptakan pengalaman bermain yang kuat dan mendalam.

Bus Simulator Indonesia yang sering disingkat sebagai BUSSID adalah permainan simulasi yang memberikan pengalaman menjadi sopir bus dalam lingkungan dunia nyata. Game ini pertama kali dirilis pada tahun 2017 dikembangkan oleh Maleo pengembang lokal asal Surabaya. Bus Simulator Indonesia merupakan jenis game simulator dimana pemain dapat menjalankan peran sebagai sopir bus dan merasakan pengalaman

mengemudi bus yang mendalam dan realistis. Dalam game Bus Simulator Indonesia juga memiliki berbagai jenis bus yang ada di Indonesia, seperti Harapan Jaya, Sumber Rahayu, Eka, Ramayana, dan Pahala Kencana. Bus Simulator Indonesia juga menawarkan pengalaman bermain yang autentik dengan mempersembahkan detail-detail khas Indonesia, mulai dari lingkungan sekitar hingga perilaku lalu lintas yang serupa dengan kondisi di negara ini. Beberapa keluhan masalah yang sering muncul dari pemain adalah adanya bug dan glitch, pemain sering melaporkan masalah seperti bus yang tiba-tiba melayang, kendaraan yang menembus objek, atau kerusakan grafis yang mengganggu. Masalah-masalah ini secara signifikan mengurangi kesenangan dalam bermain dan mengganggu alur permainan yang menyebabkan pemain merasa frustrasi dan tidak puas. Masalah lain yang juga sering dikeluhkan adalah kontrol yang kurang responsive, beberapa pemain mengeluhkan bahwa kontrol kemudi, rem, dan akselerator dalam game tidak responsif atau sulit digunakan. Hal ini menurunkan kenyamanan dan akurasi dalam mengemudi, sehingga membuat permainan terasa kurang realistis dan tidak menyenangkan. Masalah-masalah ini menunjukkan adanya tantangan signifikan yang dihadapi oleh game Bus Simulator Indonesia dalam memenuhi harapan pengguna.

Dalam zaman digital ini peran ulasan pengguna semakin signifikan dalam membentuk pandangan konsumen terhadap suatu produk atau layanan. Ulasan yang terdapat di Google Play Store merupakan informasi yang dapat diproses dan dijadikan sebagai sumber data yang berguna untuk menginformasikan perbaikan-perbaikan pada aplikasi di masa depan. Aspek yang penting dalam memahami kualitas dan respon pengguna terhadap suatu permainan adalah melalui ulasan dan penilaian yang diberikan oleh pengguna. Platform distribusi utama seperti Google Play Store menjadi tempat utama bagi pengguna untuk berbagi pengalaman mereka dalam menggunakan aplikasi dan game.

Analisis sentimen adalah suatu proses dimana kita mencari menilai dan mengkategorikan ekspresi emosi dalam teks menjadi negatif (-1), netral (0), atau positif (+1). Analisis sentimen mengacu pada sikap, pemikiran, atau penilaian yang dipengaruhi oleh emosi. Juga dikenal sebagai penambangan opini, analisis sentimen mempelajari pandangan orang terhadap entitas tertentu. Analisis sentimen digunakan untuk menilai apakah kinerja produk tersebut dinilai positif atau negatif berdasarkan ulasan pelanggan. Analisis sentimen juga dapat memperbaiki pengambilan keputusan bisnis, data tentang sentimen pengguna dapat dimanfaatkan untuk merancang strategi bisnis yang lebih efektif, mengalokasikan sumber daya dengan lebih bijak, dan merencanakan tahapan

pengembangan produk berdasarkan preferensi serta harapan pengguna.

TF-IDF Vectorizer adalah algoritma yang digunakan dalam pengolahan teks untuk mengubah dokumen teks menjadi representasi vektor numerik. Algoritma ini menggunakan metode TF-IDF untuk mengevaluasi signifikansi kata-kata dalam dokumen berdasarkan dua faktor kunci. [11] TF-IDF bertujuan untuk memberikan nilai atau bobot yang lebih tinggi kepada kata-kata yang sering muncul dalam dokumen tertentu namun jarang muncul dalam dokumen lainnya. Hal ini membantu dalam mengenali kata-kata yang paling signifikan dalam menentukan isi dari suatu dokumen dalam kerangka korpus yang lebih besar.

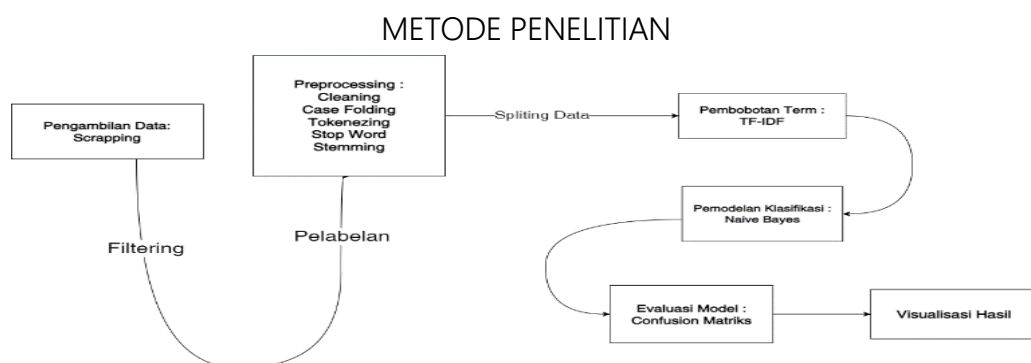
Naïve Bayes merupakan metode yang efisien untuk meningkatkan tingkat akurasi dalam klasifikasi hasil dengan memanfaatkan pendekatan statistik dan teknik probabilitas. Pemilihan Algoritma Naïve Bayes karena Algoritma ini terkenal dengan efisiensinya dalam klasifikasi teks terutama pada dataset yang besar. Algoritma Naïve Bayes terbukti dan telah digunakan secara luas dalam berbagai aplikasi analisis sentimen mulai dari analisis ulasan suatu produk hingga pemantauan media sosial seperti instagram, twitter dan facebook. Selain itu Algoritma Naive Bayes dapat berfungsi sebagai alat analisis sentimen yang efektif terutama karena kemampuannya untuk memproses data teks dengan cepat dan akurat. Algoritma Naïve Bayes digunakan untuk mengidentifikasi sentimen berbeda dari ulasan dengan langkah-langkah berikut: (i) Dengan membagi data menjadi pelatihan dan pengujian lalu melatih model dengan menghitung frekuensi kata serta probabilitas kemunculan kata dalam setiap kategori sentiment. (ii) Selanjutnya melakukan preprocessing teks agar ulasan diproses untuk menghilangkan tanda baca, mengubah semua teks menjadi huruf kecil, dan menghapus kata-kata umum yang tidak memiliki arti khusus dalam analisis sentiment. (iii) Setelah itu ekstraksi fitur menggunakan TF-IDF dan mengklasifikasikan ulasan baru berdasarkan probabilitas sentimen tertinggi serta mengevaluasi kinerja model dengan data pengujian.

Penelitian yang dilakukan [15] berjudul "Studi Komparasi Metode Analisis Sentimen Naïve Bayes, SVM, dan Logistic Regression Pada Piala Dunia 2022," kami menggunakan tiga algoritma berbeda untuk menganalisis sentimen ulasan pengguna: Bernoulli Naive Bayes, Support Vector Classifier, dan Logistic Regression. Hasil analisis menunjukkan bahwa Bernoulli Naive Bayes memperoleh nilai presisi 71%, recall 99%, dan akurasi 76%. Di sisi lain, Support Vector Classifier mencapai presisi 94%, recall 93%, dan akurasi 92%. Logistic Regression juga menunjukkan performa yang baik dengan nilai presisi 93%, recall 93%, dan akurasi 92%. Dari hasil tersebut, dapat disimpulkan bahwa Support Vector

Classifier dan Logistic Regression lebih unggul dalam hal presisi dan akurasi dibandingkan Bernoulli Naive Bayes, meskipun Bernoulli Naive Bayes memiliki nilai recall yang sangat tinggi. Penelitian ini memberikan wawasan tentang efektivitas berbagai algoritma dalam analisis sentimen, khususnya dalam konteks Piala Dunia 2022, serta memberikan kontribusi penting bagi pengembangan model yang lebih akurat di masa mendatang.

Pada penelitian yang mengklasifikasi sentimen review film dengan klasifikasi Naive Bayes. Penelitian ini menggunakan dataset ulasan film sekitar 50.000 data. Akurasi Naive Bayes pada pengolahan data IMDb menggunakan Tf-Idf sebagai seleksi fitur mencapai 86.48%. Sehingga Naive Bayes dengan fitur Tf-Idf pada masalah analisis klasifikasi sentimen review film terbukti memberikan akurasi yang akurat.

Penelitian ini akan menerapkan ekstraksi fitur TF-IDF untuk menganalisis sentimen ulasan pengguna terhadap permainan "Bus Simulator Indonesia" menggunakan algoritma Naive Bayes. Penelitian ini melakukan pengujian dengan 5 variasi pembagian data berbeda untuk mendapatkan hasil akurasi paling tinggi dan paling rendah dari model Multinomial Naive Bayes setelah menggunakan fitur TF-IDF. Tujuan dari penelitian ini adalah untuk memberikan gambaran pola sentimen dari ulasan tersebut, pola sentimen mana yang lebih tinggi dari positif, negatif atau netral, serta mengukur akurasi model dalam mengklasifikasikan sentimen. Memantau perubahan sentimen dari waktu ke waktu dengan menggunakan akurasi memungkinkan untuk mengidentifikasi dan menangkap perubahan dalam pandangan pengguna terhadap game tersebut. Hal ini membantu dalam mendeteksi tren baru atau perubahan signifikan dalam persepsi, memungkinkan respons yang cepat dan tepat terhadap umpan balik pengguna. Dengan demikian, hasil penelitian ini diharapkan memberikan wawasan dasar kepada pembaca, penulis, atau pengembang dalam memahami persepsi dan pengalaman pengguna dari ulasan game tersebut.



Gambar 1. Alur Penelitian

A. Pengambilan Data

Pengambilan data ulasan pengguna game bus simulator indonesia dari platform Play Store diambil menggunakan google colab sebanyak 1500 data ulasan di google play store menggunakan teknik web scraping, kemudian data akan disimpan dalam format dokumen csv. Web scraping adalah teknik yang digunakan untuk mengumpulkan informasi berupa data ulasan dari Play Store dengan tujuan untuk mengekstrak sebuah informasi.

B. *Preprocessing*

Preprocessing merupakan tahap dimana data dibersihkan dan diperbaiki, karena data yang dikumpulkan seringkali didapati data tersebut tidak terstruktur dan mengandung banyak karakter. Tujuan dari preprocessing adalah untuk menghilangkan gangguan dari data tersebut.

Dalam proses preprocessing terdapat beberapa langkah yang harus dijalankan, yaitu:

1. *Cleaning*

Cleaning adalah langkah penting untuk menghapus kata-kata yang tidak diperlukan dari teks, dengan maksud untuk meningkatkan kualitas data dan mengurangi gangguan pada proses klasifikasi. Proses ini juga melibatkan penghapusan noise seperti emoticon dan karakter yang tidak memiliki arti dalam kalimat ulasan.

2. *Case Folding*

Case folding adalah proses mengubah huruf kapital menjadi huruf kecil. Proses ini mempermudah pencarian karena tidak semua dokumen teks menggunakan huruf kapital secara konsisten.

3. *Tokenization*

Tokenisasi adalah langkah dalam pemrosesan teks yang memecah kalimat menjadi unit-unit kata yang lebih kecil. Tujuannya adalah untuk memudahkan pemrosesan teks, termasuk identifikasi dan penghapusan kata-kata penghubung (stopwords).

4. *Stopwords*

Stopwords adalah langkah dalam pra-pemrosesan teks di mana kata-kata yang umumnya dianggap tidak memberikan nilai informatif signifikan dihapus dari teks.

5. *Stemming*

Dengan menghilangkan imbuhan sari awalan dan akhiran, tahapan ini adalah proses yang berfungsi untuk mengubah katakata menjadi kata dasar.

C. Pembobotan Term

Penerapan metode TF-IDF (Term Frequency-Inverse Document Frequency) ini berguna untuk menentukan bobot kata dalam setiap dokumen yang memungkinkan

penentuan pentingnya kata-kata dalam konteks analisis sentimen. Bobot kata yang tinggi menandakan kata-kata yang lebih signifikan dalam analisis sentimen, dengan demikian proses transformasi mempermudah identifikasi kata-kata yang paling relevan dalam analisis sentimen. Term Frequency (TF) mencerminkan seberapa sering kata muncul dalam dokumen, sedangkan Inverse Document Frequency (IDF) mengukur kelangkaan kata dalam kumpulan dokumen. Gabungan nilai TF dan IDF digunakan untuk mengevaluasi pentingnya kata dalam dokumen, dimana nilai TF-IDF yang tinggi menunjukkan kepentingan kata tersebut dalam analisis sentimen.

D. Naïve Bayes

Naive Bayes adalah algoritma pembelajaran mesin yang menggunakan teorema Bayes dengan asumsi bahwa setiap fitur dalam data tidak memiliki ketergantungan satu sama lain.[11] Dalam penelitian ini, digunakan klasifikasi Naive Bayes tipe *Multinomial Naive Bayes*, model ini berdasarkan konsep *term frequency* yaitu menghitung seberapa sering kata tertentu muncul dalam dokumen untuk melakukan klasifikasi. Multinomial Naive Bayes mampu mengklasifikasikan data yang tidak dapat direpresentasikan secara numerik dengan menggunakan pendekatan distribusi multinomial untuk fitur-fiturnya. [23]

E. Evaluasi

Evaluasi dilakukan dengan tujuan menilai seberapa akurat hasil klasifikasi dengan model *Multinomial Naive Bayes*. Tahap evaluasi menggunakan perhitungan tabel matriks klasifikasi (confusion matrix) sebagai dasar evaluasi, confusion matrix digunakan untuk menghitung metrik akurasi, presisi, recall, dan F1-score.

F. Visualisasi Hasil

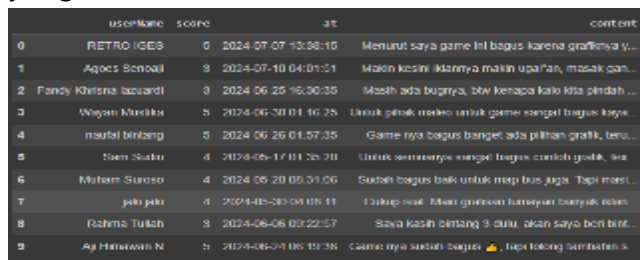
Visualisasi hasil penelitian ini meliputi grafik distribusi sentimen, confusion matrix, grafik performa model, dan word cloud yang bersama-sama memberikan gambaran lengkap tentang efektivitas model Naive Bayes dalam mengklasifikasikan sentimen ulasan pengguna terhadap permainan Bus Simulator Indonesia dengan fitur ekstraksi TF-IDF. Grafik distribusi sentimen baik dalam bentuk batang maupun pie chart, menunjukkan sebaran ulasan yang dikategorikan sebagai positif, negatif, atau netral, sehingga memudahkan identifikasi sentimen dominan. Confusion matrix menampilkan jumlah prediksi benar dan salah dalam setiap kategori sentimen, serta memungkinkan perhitungan akurasi, presisi, recall, dan F1-score. Word cloud menggambarkan kata-kata yang paling sering muncul dalam ulasan pengguna, dengan ukuran kata menunjukkan frekuensinya, membantu mengidentifikasi kata kunci dalam ulasan positif dan negatif. Kombinasi visualisasi ini memberikan wawasan lengkap tentang pola sentimen dan

mendukung pemberian rekomendasi perbaikan yang tepat.

HASIL DAN PEMBAHASAN

A. Pengambilan Data

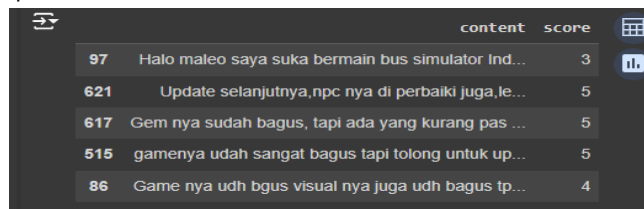
Pengambilan data ini menerapkan teknik web scraping untuk mengumpulkan data ulasan game Bus Simulator Indonesia. Teknik web scraping yang diterapkan di Google Colab terbukti efektif untuk mengumpulkan dan menganalisis data dalam jumlah besar. Proses scraping dilakukan dengan memanfaatkan Google Colab dengan menjalankan kode Python di lingkungan cloud. Proses ini berhasil mengumpulkan sebanyak 1500 data terkait ulasan game ini, yang meliputi berbagai informasi seperti ulasan pengguna, rating, username, at/tanggal dan lain-lain. Berikut ini merupakan capture data dari ulasan game Bus Simulator Indonesia yang berhasil diambil.



	username	score	at	content
0	RETRO ICE8	5	2024-07-07 13:38:15	Menurut saya game ini bagus karena grafiknya y...
1	Aqes Sencal	3	2024-07-10 04:01:51	Makin kesini idannya makin upatan, masak dan...
2	Fandy Khriena Iacuardi	3	2024-06-25 15:30:35	Masih ada bugnya, btw kenapa kalo kita pindah ...
3	Wagan Kuslika	5	2024-06-30 01:16:25	Untuk pindak modke untuk game sangat bagus bagy...
4	nasul bintang	5	2024-06-26 01:57:35	Game nya bagus banget ada pilihan grafik, teru...
5	Sam Sialin	4	2024-06-17 01:35:30	Untuk semanyng sangat bagus control grafik, lea...
6	Muham Suroso	4	2024-06-20 05:31:06	Sudah bagus baik untuk masp bus juga. Tapi masi...
7	jeki jeki	4	2024-05-20 04:03:11	Control real. Main gampang banget banyak objek...
8	Rahma Tulah	3	2024-06-06 02:22:57	Saya kasih bintang 3 dulu. akan saya beri bint...
9	Aj Hinarwan N	5	2024-06-09 06:12:36	Game nya sudah bagus 🍌 tapi tolong tambahkan s...

Gambar 2. Hasil Scrraping Google Colab

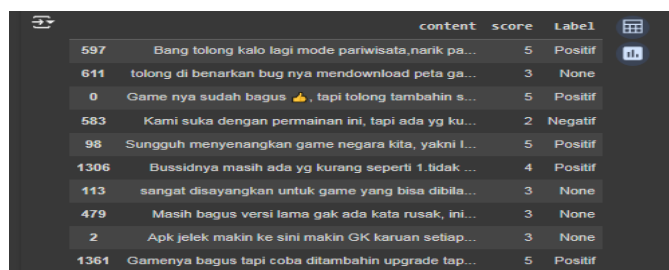
Gambar 2 menunjukkan hasil dari scraping mencakup berbagai atribut seperti reviewId, UserName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, replyContent, replyAt, dan appVersion. Selanjutnya dilanjutkan dilakukan filtering atribut, filtering ini akan menghapus semua atribut kecuali content dan score.



	content	score
97	Halo maleo saya suka bermain bus simulator Ind...	3
621	Update selanjutnya,npc nya di perbaiki juga,le...	5
617	Gem nya sudah bagus, tapi ada yang kurang pas ...	5
515	gamenya udah sangat bagus tapi tolong untuk up...	5
86	Game nya udh bgus visual nya juga udh bagus tp...	4

Gambar 3. *Filtering atribut*

Dari gambar 3 bisa dilihat bahwa atribut yang digunakan hanyalah content dan score.



	content	score	Label
597	Bang tolong kalo lagi mode pariwisata,narik pa...	5	Positif
611	tolong di benarkan bug nya mendownload peta ga...	3	None
0	Game nya sudah bagus 🍌, tapi tolong tambahin s...	5	Positif
583	Kami suka dengan permainan ini, tapi ada yg ku...	2	Negatif
98	Sungguh menyenangkan game negara kita, yakni l...	5	Positif
1306	Bussidnya masih ada yg kurang seperti 1.tidak ...	4	Positif
113	sangat disayangkan untuk game yang bisa dibila...	3	None
479	Masih bagus versi lama gak ada kata rusak, ini...	3	None
2	Apk jelek makin ke sini makin GK karuan setiap...	3	None
1361	Gamenya bagus tapi coba ditambahin upgrade tap...	5	Positif

Gambar 4. Pelabelan

Gambar 4 ini adalah proses pelabelan dimana Python digunakan untuk menetapkan label positif atau negatif pada atribut konten tertentu berdasarkan skor yang diberikan pada komentar. Komentar dengan skor dibawah tiga diberi label "Negatif", sementara komentar dengan skor empat atau lima diberi label "Positif". Proses ini disebut sebagai Pelabelan Data.

```
Jumlah masing-masing label:
Label
Positif    1013
Negatif    273
Name: count, dtype: int64
```

Gambar 5. Jumlah Label

Gambar 5 menunjukkan hasil dari pelabelan sentimen positif dan negatif dengan jumlah positif lebih banyak dengan 1013 ulasan sedangkan negatif hanya 237 ulasan.

```
Positif
content score
0 kalau bisa tempat makan nya di aktifin donk. s... 5
1 saya suka walaupun banyak iklan saya suma dari... 5
4 game nya sih bagus tapi kenapa pas saya telepo... 5
5 Untuk play udah bagus tapi plis ya maleo bussi... 5
7 Gamenya bagus, tp smoga untuk kedepannya bisa ... 5
11 Gamenya udh bagus sih, cuma jujur aja jadi ga ... 4
12 Untuk next update obb bussid nya bisa di kempa... 5
13 Mantap sih game nya,tapi jangan kebanyakan ikl... 4
14 Bagus banget, filturnya bagus, sekarang pun bi... 5
15 Bagus sih bang, btw tambahin mode main bareng ... 5
```

Gambar 6. Hasil Sentimen Positif

```
Negatif
content score
2 Saya tidak suka dengan update kali ini karena ... 1
10 Pembaruan berikutnya tambahin tol cirebon samp... 2
17 Bus ini banyak bug nya dan sering iklan saya b... 2
28 Semua bagus sayang bnget untuk mod interior ik... 2
29 Gamenya sih udah bagus cuman kekurangan nya tu... 1
32 trafik nya tolol, orang belok di tabrak, bener... 2
37 game nya sih udah bagus. cuma kurang beberapa ... 2
41 Game apaan cok, masa gw pake wifi gw bagus pad... 1
43 Bug nya euy.... Habis iklan isi bensin atau l... 2
46 Suara in gamenya kurang keras, masa harus sett... 1
```

Gambar 7. Hasil Sentimen Negatif

B. Preprocessing

Setelah pelabelan selesai, dilakukan serangkaian tahapan preprocessing untuk mempersiapkan data yang akan digunakan dalam analisis lebih lanjut.

Tabel 1. Hasil *Preprocessing*

processing	Content
Sebelum	Kenapa tiap mau mabar mesti DC atau loading lama banget ya Dev? Tolong solusinya Karna kalau masalah signal, signal saya termasuk bagus N lancar untuk game yg lain.

Cleaning	Kenapa tiap mau mabar mesti DC atau loading lama banget ya Dev Tolong solusinya Karna kalau masalah signal signal saya termasuk bagus N lancar untuk game yg lain
Case Folding	kenapa tiap mau mabar mesti dc atau loading lama banget ya dev tolong solusinya karna kalau masalah signal signal saya termasuk bagus n lancar untuk game yg lain
Stop Word	mabar mesti dc loading banget ya dev tolong solusinya karna signal signal bagus n lancar game yg
Tokenize	mabar,mesti,dc,loading,banget,ya,dev,tolong,solusinya,karna,signal signal, bagus,n,lancar,game,yg
Stemming	mabar mesti dc loading banget ya dev tolong solusi karna signal signal bagus n lancar game yg

Pada Tabel 1 menjelaskan hasil proses preprocessing teks melibatkan beberapa langkah. (i) cleaning digunakan untuk menghapus karakter atau data yang tidak relevan dari teks. (ii) case folding mengubah semua huruf dalam teks menjadi huruf kecil agar perbedaan kapitalisasi tidak mempengaruhi analisis. (iii) stop word removal menghapus kata-kata penghubung yang umumnya tidak memberikan nilai tambah dalam analisis teks. Keempat, tokenization membagi teks menjadi potongan-potongan lebih kecil seperti kata atau frasa. Terakhir, stemming mereduksi kata-kata ke bentuk dasarnya dengan menghapus imbuhan. Setelah semua langkah ini dilakukan, data siap diproses lebih lanjut untuk mendapatkan informasi atau pola yang relevan dari teks yang telah diproses.

C. Pembobotan Term

Tabel 2. Variasi pembagian data

Data latih: Data uji
90%:10%
80%:20%
70%:30%
60%:40%
50%:50%

Pada tabel 2 ini dimulai dengan membagi data menjadi dua yaitu data uji dan data latih, pengujian ini bertujuan untuk mendapatkan hasil akurasi Naïve Bayes yang menggunakan pembobotan TF-IDF dengan 5 variasi data latih dan data uji. Pengujian dilakukan sebanyak 1 kali pada setiap variasi, kemudian diambil nilai rata-ratanya. Variasi pembagian data bisa dilihat pada tabel 2 diatas.

Pembobotan TF-IDF

```
[558] 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)

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

(1816,)
(1816,)
(255,)
(255,)

[560] from sklearn.feature_extraction.text import CountVectorizer
      vectorizer = CountVectorizer()
      vectorizer.fit(X_train)

= CountVectorizer
  CountVectorizer()

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

Gambar 8. Pembobotan TF-IDF

Pada proses ini, fitur diekstraksi dari data teks menggunakan metode TF-IDF dan *Count Vectorizer*. Pertama, *TfidfVectorizer* dari *library sklearn.feature_extraction.text* digunakan untuk mengubah teks menjadi numerik berdasarkan frekuensi kemunculan dan bobot tiap kata dalam dokumen. Fungsi *tfidf_vectorizer.fit_transform(X_train)* digunakan untuk melatih model TF-IDF dengan data pelatihan *X_train* dan mengubahnya menjadi matriks, sementara *tfidf_vectorizer.transform(X_test)* digunakan untuk mentransformasi data uji *X_test* menggunakan model TF-IDF yang telah dilatih. Selanjutnya, untuk menghitung frekuensi kata secara sederhana, digunakan *CountVectorizer*. Model ini dilatih dengan *X_train* menggunakan *vectorizer.fit(X_train)*, kemudian data pelatihan dan uji diubah menjadi matriks menggunakan *vectorizer.transform(X_train)* dan *vectorizer.transform(X_test)*. Setelah transformasi ini, data teks *X_train* dan *X_test* siap digunakan untuk pelatihan dan pengujian model machine learning selanjutnya.

Tabel 3. Hasil Pengujian Data

Pembagian	Accuracy	Precision	Recall	F1-Score
90:10	85%	84%	85%	84%
80:20	84%	82%	84%	82%
70:30	83%	81%	83%	81%
60:40	83%	81%	83%	81%
50:50	82%	80%	82%	81%

Pada tabel 3 menjelaskan setelah dilakukan pembobotan TF-IDF lalu di lakukan pengujian manual menggunakan pembagian data latih dan data uji dengan lima variasi yang berbeda menghasilkan beberapa kesimpulan. Variasi 90%:10% menghasilkan akurasi 85%, variasi 80%:20% menghasilkan 84% akurasi, 70%:30% dan Variasi 60%:40% sama-sama mencapai akurasi tertinggi sebesar 83, sedangkan variasi 50%:50% menghasilkan

akurasi 82%.

D. Evaluasi

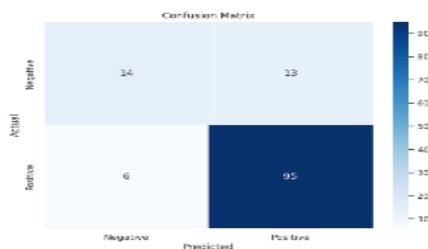
```
MultinomialNB Accuracy: 0.8515625
MultinomialNB Precision: 0.7
MultinomialNB Recall: 0.5185185185185185
MultinomialNB f1_score: 0.5957446888510639
confusion_matrix:
[[14 13]
 [ 6 95]]
-----
              precision    recall  f1-score   support
Negatif         0.70         0.52         0.60         27
Positif         0.88         0.94         0.91        101
accuracy                0.79         0.73         0.85        128
macro avg                0.79         0.73         0.75        128
weighted avg             0.84         0.85         0.84        128
True Positive : 95
True Negative : 14
False Positive : 13
False Negative : 6
```

Gambar 9. Hasil Model

Hasil pengujian manual sebanyak lima kali pada algoritma Naïve Bayes menunjukkan accuracy tertinggi sebesar 85%. Nilai ini sesuai dengan data pada confusion matrix yang mencatat 95 True Positives, 14 True Negatives, 13 False Positives, dan 6 False Negatives.

E. Hasil Visualisasi

Visualisasi Confusion Matrix



Gambar 15. Confusion Matrix

Dari gambar 15 confusion matrix diatas dapat disimpulkan: Confusion matrix di atas menunjukkan hasil dari lima kali pengujian manual pada algoritma Naïve Bayes, dengan akurasi tertinggi sebesar 85%. True Positives (TP) adalah 95 sampel positif yang berhasil diidentifikasi dengan benar, True Negatives (TN) adalah 14 sampel negatif yang berhasil diidentifikasi dengan benar, False Positives (FP) adalah 13 sampel negatif yang salah diidentifikasi sebagai positif, dan False Negatives (FN) adalah 6 sampel positif yang salah diidentifikasi sebagai negatif.

Visualisasi *Word cloud*

Visualisasi menggunakan Word Cloud pada python dari kata yang sering dipakai pada ulasan dapat dilihat pada gambar 10.



Gambar 10. Wordcloud Kata yang paling sering dipakai

Pada gambar 10 menunjukkan kata yang sering dipakai secara keseluruhan.
Visualisasi *Wordcloud* Kelas Positif



Gambar 11. Wordcloud Kelas Positif

Pada gambar 11 kata positif yang sering muncul adalah dan, ada, bisa, di, bagu, game dan masih banyak lagi.

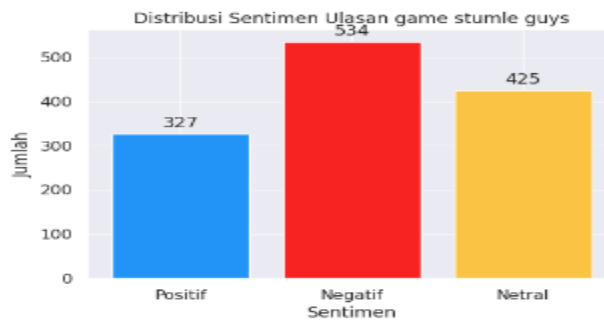
Visualisasi Wordcloud Kelas Negatif



Gambar 18. Wordcloud Kelas Negatif

Pada gambar 18 kata yang menilai negatif adalah game, saya, iklan, tolong, ada dan lain sebagainya.

Visualisasi Pola Sentimen



Gambar 12. Pola Distribusi Ulasan Sentimen

Dari gambar 12 data ulasan pengguna menunjukkan pola sentimen 327 ulasan positif, 534 ulasan negatif, dan 425 ulasan netral. Mayoritas pengguna sebanyak 534 ulasan negatif, merasa tidak puas dengan permainan bus simulator indonesia, menandakan adanya masalah signifikan yang perlu segera diatasi oleh pengembang seperti bug, glitch, atau kontrol yang tidak responsif. Sedangkan 327 ulasan positif menunjukkan bahwa sejumlah pengguna merasa puas dan menikmati beberapa aspek dari permainan seperti grafis, konsep, atau fitur tertentu. Sementara itu 425 ulasan netral

mencerminkan pengalaman pengguna yang rata-rata, di mana mereka tidak menemukan masalah besar tetapi juga tidak menemukan hal yang sangat memuaskan, menunjukkan area yang bisa ditingkatkan lebih lanjut. Secara keseluruhan pola ini menunjukkan bahwa meskipun ada pengguna yang puas, banyak masalah yang mempengaruhi kepuasan sebagian besar pengguna, sehingga mengidentifikasi dan mengatasi masalah-masalah ini dapat membantu meningkatkan pengalaman bermain dan meningkatkan ulasan positif di masa depan.

SIMPULAN

Kesimpulan dari penelitian di atas adalah bahwa analisis sentimen ulasan pengguna game "Bus Simulator Indonesia" dari Google Play Store berhasil mengidentifikasi pola sentimen secara efektif menggunakan ekstraksi fitur TF-IDF dan algoritma Naïve Bayes.

1. Dari 1500 ulasan yang dikumpulkan, 1286 ulasan dianalisis setelah preprocessing, yang menunjukkan distribusi sentimen dengan mayoritas ulasan negatif (534), diikuti oleh ulasan netral (425) dan positif (327).
2. Pengujian dengan 5 variasi data latih dan uji menunjukkan bahwa model Multinomial Naïve Bayes memberikan akurasi tertinggi sebesar 85% pada variasi 90%:10%, dengan nilai precision, recall, dan F1-score yang konsisten. Variasi lainnya menunjukkan akurasi sedikit lebih rendah, antara 82% hingga 84%, namun tetap menunjukkan kinerja model yang baik.

Hasil ini memberikan pandangan yang jelas tentang sentimen pengguna terhadap game dan menunjukkan bahwa penggunaan TF-IDF dan Naïve Bayes efektif dalam menganalisis ulasan. Pengembang dapat menggunakan wawasan ini untuk meningkatkan kualitas game, merespons umpan balik pengguna dengan lebih baik, dan mengembangkan strategi pemasaran yang lebih tepat sasaran.

DAFTAR PUSTAKA

- M. Roca-Puigròs, A. Gerber, M. Ulrich, M. Y. Reich, D. Beat Müller, and P. Wäger, "Linking socio-economic metabolism models and simulation games: Reflections on benefits and challenges," *J. Ind. Ecol.*, vol. 28, no. 2, pp. 182–193, 2024, doi: 10.1111/jiec.13462.
- Tutus Subronto, "Bus Simulator Indonesia, Game Buat Penghobi Dunia Otomotif," *carmudi.co.id*, 2023. <https://www.carmudi.co.id/journal/bus-simulator-indonesia-game-otomotif/> (accessed May 10, 2024).

- M. F. Ravizaldi and N. Z. Saputra, "Analisis Kearifan Lokal Pada Game Bus Simulator Indonesia," pp. 750–755, 2022.
- Tekno, "Bus Simulator Indonesia, Game Lokal yang Wajib Diunduh," <https://kumparan.com/>, 2021. <https://kumparan.com/how-to-teknobus-simulator-indonesia-game-lokal-yang-wajib-diunduh-1x93LobTERc/full> (accessed May 21, 2024).
- R. Maulana, M. Raihan, and I. Santoso, "Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Tokopedia," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 7, no. 2, pp. 177–189, 2023, doi: 10.47111/jti.v7i2.10071.
- M. Izunnahdi, G. Aburrahman, and A. Eko Wardoyo, "Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes Sentiment Analysis on KAI Access Application Review Data on Google PlayStore Using Multinomial Naive Bayes Method," *J. Smart Teknol.*, vol. 4, no. 2, pp. 2774–1702, 2023, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- M. M. Hamed Taherdoost 1, "Artificial Intelligence and Sentiment Analysis : A Review in," *Comput.* 2023, 12, 37. <https://doi.org/10.3390/computers12020037>, no. 0, 2023.
- M. R. Elfansyah, M. R. Perdana, I. N. T. I. Nabawi, and R. Rudiman, "Analisis Sentiment Cyberbullying pada media Youtube menggunakan Algoritma Naive Bayes," *KOMPUTEK J. Tek. Univ. Muhammadiyah Ponorogo*, vol. 8, no. 1, pp. 29–35, 2024, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek>
- A. Patel, P. Oza, and S. Agrawal, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," *Procedia Comput. Sci.*, vol. 218, pp. 2459–2467, 2022, doi: 10.1016/j.procs.2023.01.221.
- A. Wijaya, P. Korespondensi, M. Enjeli, S. Informasi, and U. Katolik Musi Charitas, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, [Online]. Available: <https://doi.org/10.58602/jics.v2i2.24>
- [A. Susanto, I. Agung Dzulkarnain, P. Studi Statistika, and F. Matematika Dan Ilmu Pengetahuan Alam, "Analisis Sentimen Data Twitter Topik Ekonomi Dan Industri Dengan Metode Naive Bayes Dan Random Forest," *J. Ilm. Wahana Pendidikan*, Oktober, vol. 2023, no. 20, pp. 59–65, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.8398895>

- A. Saleh, N. Dharshinni, D. Perangin-Angin, F. Azmi, and M. I. Sarif, "Implementation of Recommendation Systems in Determining Learning Strategies Using the Naïve Bayes Classifier Algorithm," *Sinkron*, vol. 8, no. 1, pp. 256–267, 2023, doi: 10.33395/sinkron.v8i1.11954.
- R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- M. Ramdan, A. Surya, and U. Hayati, "ANALISIS SENTIMEN ULASAN PENGGUNA OVO MENGGUNAKAN," vol. 8, no. 3, pp. 2780–2786, 2024.
- M. Zaki Anbari, M. Zaki Anbari, and B. Sugiantoro, "Studi Komparasi Metode Analisis Sentimen Naïve Bayes, SVM, dan Logistic Regression Pada Piala Dunia 2022," *J. Media Inform. Budidarma*, vol. 7, no. 2, pp. 688–695, 2023, doi: 10.30865/mib.v7i2.5383.
- M. T. Razaq, D. Nurjanah, and H. Nurrahmi, "Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier dengan Fitur TF-IDF," *e-Proceeding Eng.*, vol. 10, no. 2, pp. 1698–1712, 2023, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/19997>
- F. Romadoni, Y. Umaidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, pp. 247–253, 2020, doi: 10.32736/sisfokom.v9i2.903.
- M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di," vol. 7, no. 1, pp. 478–483, 2023.
- S. Informasi, F. Sains, U. Katolik, and M. Charitas, "DALAM GOOGLE PLAY STORE DENGAN MODEL NAÏVE BAYES," vol. 4, no. 2, pp. 374–385, 2023, doi: 10.46576/djtechno.
- Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- T. A. Sari, E. Sinduningrum, and F. Noor Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes," *Media Online*, vol. 3, no. 6, pp. 773–779, 2023, doi: 10.30865/klik.v3i6.884.

- I. Hasmadi, R. Rudiman, K. Huda, D. Putra, and M. F. Jundullah, "Analisis Sentimen Terhadap Kualitas Layanan Driver Gojek Di Aplikasi Play Store Menggunakan Algoritma Naïve Bayes Dan Aplikasi Orange," vol. 2, no. 1, pp. 138–151, 2024, [Online]. Available: <https://doi.org/10.59841/saber.v2i1.673>
- Ernianti Hasibuan and Elmo Alistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," J. Tek. dan Sci., vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.