

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI OJOL THE GAME MENGGUNAKAN ALGORITMA NAÏVE BAYES

Sumini Aulia¹, Ismail²

^{1,2}Sistem Informasi, Universitas Indonesia Membangun; Jl. Soekarno Hatta No. 448, Batununggal, Kec. Bandung Kidul, Kota Bandung, Jawa Barat 40266

Received: 14 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Analisis sentimen, *Ojol The Game*, *Google Play Store*, Naïve Bayes Classifier (NB), Rappid miner

Correspondent Email:

auliasumini@gmail.com

Abstrak. *Game* merupakan suatu *platform* permainan yang sangat menarik terutama untuk *game online*. Aplikasi *Ojol The Game* merupakan aplikasi *game* simulator yang cukup populer dikalangan anak muda terutama pada gen Z. Meskipun banyaknya peminat dan pengguna yang menyukai aplikasi *game* ini, tidak menutup kemungkinan banyak pengguna yang memiliki kesan kurang baik selama melakukan permainan di *game* tersebut. Banyaknya opini-opini dari para pengguna membuat peneliti tertarik untuk melakukan analisis sentimen pada aplikasi *Ojol The Game* di *google play store*. Tujuan dilakukannya penelitian ini untuk menganalisis sentimen pengguna aplikasi *Ojol The Game* dengan model Naïve Bayes Classifier. Penelitian ini menggunakan data sebanyak 1.198 ulasan, pengambilan data dilakukan teknik *web scrapping* dengan bahasa pemrograman *python*. Data-data tersebut dilakukan preprocessing dengan 4 tahapan yaitu, *filter tokenize*, *filter stopwords*, *filter transform cases*, dan *filter tokens*. Kemudian *dataset* tersebut dibagi menjadi dua yaitu data latih dan data uji dengan menggunakan perbandingan 80:20. Hasil akhir dari pengujian ini adalah menghasilkan nilai akurasi ketepatan dari model naïve bayes classifier dalam melakukan pelabelan dan klasifikasi *dataset*. Hasil pengujian menunjukkan bahwa akurasi mendapat nilai sebanyak 95.00%, *recall* 97,98%, dan presisi 94,19%. Dari hasil penelitian ini mendapatkan kesimpulan sentimen positif lebih besar dengan jumlah *presentase* 62% dibandingkan sentimen negatif dengan *persentase* 38%.

Abstract. *Games* are a very interesting gaming platform, especially for online games. The *Ojol The Game* application is a simulator game application which is quite popular among young people, especially Gen Z. Even though there are many enthusiasts and users who like this game application, it does not rule out the possibility that many users will have a bad impression while playing this game. The large number of opinions from users makes researchers interested in conducting sentiment analysis on the *Ojol The Game* application on the Google Play Store. The aim of this research is to analyze the sentiment of users of the *Ojol The Game* application using the Naïve Bayes Classifier model. This research used data from 1,198 reviews, data collection was carried out using web scrapping techniques using the Python programming language. This data was preprocessed in 4 stages, namely, tokenize filter, stopwords filter, transform cases filter, and tokens filter. Then the dataset is divided into two, namely training and test data using a ratio of 80:20. The final result of this test is to produce an accuracy value for the accuracy of the naïve Bayes classifier model in labeling and classifying the dataset. The test results showed that accuracy was 95.00%, recall was 97.98%, and precision was 94.19%. From the results of this research, it was concluded that positive

sentiment was greater with a percentage of 62% compared to negative sentiment with a percentage of 38%.

1. PENDAHULUAN

Game online sangat diminati oleh semua kalangan masyarakat terutama pada generasi milenial dan generasi Z. *Online game* merupakan media hiburan yang mudah didapat dan tidak bisa dipisahkan dari sebagian golongan masyarakat selama ini. *Game online* ialah *platform* yang banyak menarik perhatian masyarakat dimana para pengguna dapat terhubung dengan pengguna lain secara *online* dengan *platform* yang sudah diberikan pada *game* itu sendiri [1].

Salah satu *game online* yang saat ini sedang hangat dibicarakan pada media sosial X dan TikTok adalah aplikasi *Ojol The Game*. *Ojol The Game* adalah aplikasi simulasi menjadi *driver* ojek *online* yang banyak mengambil contoh aktivitas yang ada didalam kehidupan nyata. Didalam aplikasi *Ojol The Game* pengguna harus mengambil orderan penumpang, barang, dan makanan yang akan diantarkan ke lokasi yang sudah ditentukan. Pengguna dapat menerima rating dan uang yang bisa digunakan untuk meng-upgrade kendaraan dan aksesoris lainnya agar lebih efisien ketika menyelesaikan misi pada *game*. Selain di dunia nyata, pengguna akan mendapatkan sensasi menjadi *driver* ojek *online* dalam dunia *virtual*.

Aplikasi *Ojol The Game* dikembangkan oleh *CodeXplore* yang merupakan *developer* asal Indonesia. Saat ini aplikasi *Ojol The Game* sedang hangat dibicarakan pada media sosial X dan TikTok. Menurut katadata.co.id aplikasi *Ojol The Game* dirilis pada tanggal 18 Januari 2021. *CodeXplore* sendiri merupakan suatu pengembang *game* indie di Indonesia yang berfokus pada pengembangan *game mobile*. *Ojol The Game* menjadi aplikasi *game* paling populer dan menempati peringkat pertama di *google play store* dan *app store*. Hingga saat ini *game* tersebut sudah terinstal lebih dari 10 juta kali dan mendapatkan rating 4,3 dengan 168 ribu ulasan. Ulasan atau opini yang diberikan pengguna aplikasi *Ojol The Game* di *Google Play Store* sangat banyak dan kalimatnya sangat tidak terstruktur. Sehingga diperlukan cara untuk mengetahui bagaimana penilaian pengguna aplikasi *Ojol The Game* dalam memberikan opini terkait aplikasi yang mereka digunakan [2]. Banyaknya komentar ataupun pembahasan dari pengguna mengenai permasalahan dikala bermain permainan, akan lebih baik apabila dicoba analisis sentimen pada informasi pembahasan aplikasi *Sausage Man* di *Google Play Store* proses untuk menciptakan pengkategorian komentar lewat pembahasan yang

diberikan pada sesuatu topik jadi berlabel positif dan negatif. Dengan terdapatnya analisis sentimen ini diharapkan dapat menolong pengembang permainan untuk mengenali kekurangan dari aplikasi yang mereka buat dari pembahasan pengguna aplikasi tersebut. Pembahasan pengguna aplikasi dapat jadi rujukan untuk pengembang buat meningkatkan aplikasi permainan yang mereka buat [3].

Naive Bayes yakni algoritma yang bisa mengklasifikasikan dengan tata metode sederhana serta sangat *universal* digunakan. Model klasifikasi Naive Bayes menghitung *probabilitas posterior* sesuatu kelas bersumber pada distribusi kata dalam dataset (Insan et angkatan laut (AL), 2023). Di sisi lain itu tergantung pada representasi dataset yang sangat simpel pada *Bag of words*. Model ini bekerja dengan mengekstraksi fitur *bag of words* yang mengabaikan posisi kata dalam dataset. Metode ini menggunakan *Teorema Bayes* buat memprediksi kemunculan atau probabilitas bila set fitur yang diberikan mempunyai label tertentu (Primandani Arsi et angkatan laut (AL), 2023). Sebagian *riset* yang telah dicoba lebih dahulu dengan memakai algoritma *naïve bayes classifier* memperoleh hasil yang memuaskan. Salah satu rujukan yang digunakan oleh peneliti ialah mengulas tentang pembahasan aplikasi satu sehat dengan memakai algoritma *naïve bayes*. *Riset* tersebut memakai 1.080 informasi hasil penemuan *riset* tersebut memperoleh nilai akurasi 81. 65%. *Riset* analisis sentimen selanjutnya ialah pembahasan pada aplikasi *brimo riset* ini memakai 1.550 informasi serta memperoleh hasil prediksi pembahasan negatif sebanyak 53. 09% serta pembahasan positif 46. 90% dengan nilai akurasi sebesar 84, 52%. Penelitian ini dilakukan karna memiliki tujuan untuk menganalisis sentimen terhadap pengguna aplikasi *Ojol The Game* dengan algoritma *naive bayes classifier* untuk mengklasifikasi ulasan dan opini pengguna terkait dengan aplikasi *Ojol The Game*. selain itu, tujuan dilakukannya penelitian ini yaitu untuk mengukur seberapa besar tingkat akurasi yang dihasil kan metode *naive bayes* dengan bantuan *tools* *rappid miner*. Bersumber pada penjelasan diatas periset tertarik buat melaksanakan *riset* menimpa analisis sentimen pembahasan aplikasi *Ojol The Game* pada *google play store* serta memakai aplikasi *rappid miner* selaku perlengkapan bantu buat mengklasifikasi informasi *riset*.

2. TINJAUAN PUSTAKA

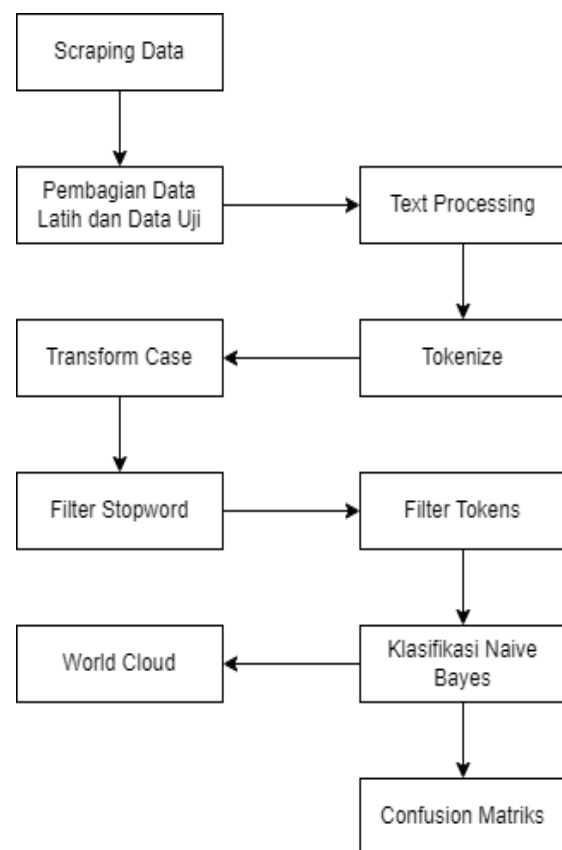
Sentimen analisis adalah suatu teknik dalam mengolah dataset dalam bentuk kalimat yang berfungsi untuk mendapatkan suatu informasi dari sebuah opini atau sentimen yang di dalamnya terdapat nilai sentimen positif dan sentimen negatif [4]. Dalam analisis sentimen yaitu mengelompokkan pola-pola dalam kalimat yang terdapat dalam *dataset*, kata, ataupun pendapat. Analisis sentimen digunakan untuk mengenali pengetahuan perasaan lebih cenderung masuk kedalam kategori sentimen positif maupun sentimen negatif [5].

Penilaian rating di aplikasi *Google Play Store* diiringi dengan pembahasan dari para pengguna terhadap aplikasi *Ojol The Game*. Pembahasan tersebut memiliki opini dari para pengguna yang berkaitan dengan aplikasi tersebut, opini-opini ini dapat berupa rasa kecewa, rasa suka, serta masukan-masukan untuk aplikasi tersebut. Opini yang sudah diberikan oleh pengguna dapat memberikan pengaruh pada calon pengguna sebagai bahan pertimbangan untuk menginstal aplikasi tersebut atau tidaknya [6]. Dengan adanya media sosial aplikasi *Ojol The Game* menjadi *viral* pada tahun 2024 sedangkan aplikasi *game* ini sudah diluncurkan sejak tahun 2021. Aplikasi *Ojol The Game* banyak digemari oleh kalangan anak muda terutama pada gen Z. Aplikasi *Ojol The Game* merupakan aplikasi simulator menjadi seorang *driver ojol*.

3. METODE PENELITIAN

Metode penelitian merupakan metode yang digunakan secara *universal* karena metode ini dapat memperoleh informasi yang akurat dengan tujuan menciptakan data dari hasil *riset* yang telah dicoba. Pada *riset* ini penulis memakai model dari Naïve Bayes Classifier dengan bantuan *software* Rappid Miner. Metode ini bertujuan untuk melakukan klasifikasi sentimen pada ulasan pengguna aplikasi *Ojol The Game*. Model klasifikasi Naïve Bayes didasarkan pada probabilitas yang simpel serta dirancang untuk digunakan dengan anggapan independensi antara variabel penjelas. Algoritma ini menekankan pendidikan probabilitas (Siahaan et angkatan laut (AL)., 2023). Menurut (Meter. Afriansyah et angkatan laut (AL)., 2024) Naïve Bayes Classifier merupakan suatu model klasifikasi yang berdasar pada *teorema Bayes*. Prosedur pengklasifikasian ini memakai tata cara probabilitas serta statistik yang pertama kali ditemukan oleh ilmuwan asal Inggris bernama *Thomas Bayes*, ialah sesuatu cara untuk memprediksi kesempatan di masa depan yang bersumber pada pengalaman di masa lampau, sehingga model penelitian ini dikenal dengan nama *Teorema Bayes*.

Algoritma Naïve Bayes Classifier banyak diterapkan pada penelitian yang berkaitan dengan sentimen karena dapat menghasilkan nilai yang akurat dan memiliki keberhasilan di berbagai *konteks* [7]. Cara kerja dari metode Naïve Bayes Classifier dengan menghitung probabilitas dan memprediksi dari klasifikasi yang sudah dilakukan. Metode Naïve Bayes merupakan kategori machine learning, machine learning sendiri memiliki sifat *predictive* dan *descriptive* yang memiliki tujuan untuk memprediksi berdasarkan data dan informasi yang didapatkan [8]. Naïve bayes ialah model pengklasifikasian yang sangat populer digunakan untuk penelitian analisis sentimen karena naïve bayes mampu menghasilkan tingkat keakuratan yang cukup baik. Untuk memudahkan penelitian ini peneliti membuat kerangka konsep yang dapat dilihat pada gambar 3.1 sebagai berikut:



Gambar 3. 1 Kerangka Konsep Penelitian

Dalam *riset* ini menggunakan metode kualitatif dengan cara melakukan pengumpulan data menggunakan teknik *web scrapping*. *Web scrapping* merupakan bahasa pemrograman dari *python* yang bersifat *open source*. Peneliti menggunakan *tools* dari *google colab* sebagai alat untuk melakukan pemrograman *web scrapping* [9]. Untuk mengambil data dalam jumlah yang cukup besar ada proses yang harus dilakukan dengan cara menginstal *google-*

play-scraper dan menggunakan *library pandas* serta mencantumkan *url* aplikasi *Ojol The Game* pada proses coding. Hal ini dilakukan untuk mengambil data ulasan pada aplikasi *Ojol The Game* di *google play store*. Pada penelitian ini peneliti berhasil mendapatkan 1.198 data ulasan pengguna aplikasi *Ojol The Game*. Data yang diambil merupakan data terbaru dari tanggal 31 Januari 2024 sampai tanggal 15 Maret 2024. Data yang berhasil ditarik akan otomatis tersimpan di *google drive* dengan format *csv*.

Langkah berikutnya penulis menggunakan *software rapid miner* untuk memproses, menganalisa dan memprediksi data ulasan *Ojol The Game* yang didapatkan dalam proses *scrapping*. Pada *software rapid miner* ini dilakukan pembagian data menjadi dua yaitu, data latih dan data uji, pembagian data menggunakan perbandingan rasio 80% : 20% untuk mendapatkan tingkat keberhasilan yang baik. Pada penelitian ini 80% digunakan untuk data latih atau data *training* dan 20% digunakan untuk data uji atau data *testing*. Semakin besar data latih atau data *training* yang digunakan maka semakin baik pula *machine learning* dalam membaca dan memahami pola-pola data yang ada. Model data latih ialah representasi pembelajaran yang akan digunakan dalam memprediksi suatu kelas pada data baru atau data yang belum ada sebelumnya. Setelah dilakukan pembagian data untuk masuk ketahap akurasi dan prediksi dengan model *naïve bayes* perlu dilakukan pembersihan data atau *preprocessing*. Data yang didapatkan dari hasil *scrapping* merupakan data yang tidak terstruktur dan memiliki banyak *noise*. Dalam tahap *preprocessing* ini terdapat beberapa tahapan yaitu *tokenize*, *transform cases*, *filter stopwords*, dan *filter tokens*. Pada tahapan *tokenize* data ulasan akan dipecah dari berbentuk kalimat menjadi kata per kata, pada tahap *tokenize* juga dapat menghilangkan simbol-simbol atau *regular expression* seperti `([-!\"#$%&'()*+,-./:;<=>?@[\\\]_`{|}~])`.

Berikutnya pada tahap *transform cases* memiliki fungsi untuk mengubah data kalimat yang masih memiliki huruf besar diubah menjadi huruf kecil [10]. Tahap ini perlu dilakukan untuk menghindari dan memudahkan *machine learning* dalam membaca sebuah data. Setelah semua data berhasil diubah menjadi huruf kecil perlu dilakukannya *filter stopwords* untuk mengurangi jumlah kata dalam sebuah data yang bertujuan untuk mengurangi kata-kata yang tidak penting atau tidak memiliki arti. Pada tahap ini peneliti menggunakan kamus *stopwords* dari *kaggle.com* dan tahap terakhir yaitu *filter tokens* yang berfungsi untuk menghilangkan kata-kata yang kurang dari 4 huruf/karakter dan tidak lebih dari 25 huruf/karakter. Hal ini perlu dilakukan untuk mempercepat

performa *machine learning* dalam memproses suatu data.

Setelah tahap *preprocessing* berhasil dilakukan data ulasan *Ojol The Game* akan diklasifikasi menjadi dua label yaitu, ulasan bersifat positif dan ulasan bersifat negatif. *Naïve Bayes* akan memprediksi dan memberikan label pada data uji dengan cara mempelajari pola-pola data dan probabilitas dari data latih. Model *Naïve Bayes Classifier* ialah model yang banyak digunakan untuk melakukan klasifikasi sesuatu sentimen dengan baik. *Naïve Bayes* bisa mengelompokkan pendapat menjadi sentimen positif maupun sentimen negatif terhadap suatu aplikasi ataupun isu yang lagi tumbuh dimasyarakat [5]. Berikutnya adalah *confusion matriks* merupakan tahap untuk melihat seberapa baik performance dari algoritma *naïve bayes classifier*. *Confusion matriks* memiliki tolak ukur yang digunakan seperti nilai *accuracy*, *precision*, dan *recall*. *Accuracy* berfungsi untuk mengetahui seberapa besar model *naïve bayes* dalam memprediksi suatu data dengan nilai yang benar atau aktual [11]. *Precision* berfungsi untuk mengenali seberapa pas ataupun akurat model dari yang diprediksi dalam kategori kelas positif. *Precision* merupakan ukuran yang baik buat membenarkan nilai *False Positif* pada suatu model yang besar. *Recall* berperan buat menghitung seberapa banyak dari nilai Actual/benar Positif yang berhasil diidentifikasi oleh model memakai pelabelan *True Positif*. *Recall* merupakan metrik model yang digunakan buat memilah model terbaik jika ada nilai yang besar dengan *False Negatif*. *Confusion matrix* ialah sesuatu tabel yang terdiri atas banyaknya baris informasi uji yang diprediksi benar serta tidak benar oleh model klasifikasi. Tabel *confusion matrix* diperlukan buat membenarkan kinerja suatu model klasifikasi [12].

1. *True positif (TP)* : ialah jumlah *dataset* positif yang bernilai benar dari hasil pengklasifikasian.
2. *True negatif (TN)* : ialah jumlah *dataset* negatif yang bernilai benar dari hasil pengklasifikasian.
3. *False positif (FP)* : ialah jumlah data positif yang bernilai salah dari hasil pengklasifikasian.
4. *False negatif (FN)* : ialah jumlah data negatif yang bernilai salah dari hasil pengklasifikasian

Pada tahapan terakhir untuk melihat topik dan pembahasan yang sering diberikan oleh pengguna aplikasi *Ojol The Game* penulis menggunakan *word cloud* yang berfungsi untuk melihat kata-kata yang sering muncul pada ulasan aplikasi *Ojol The Game*. *Word Cloud* ialah sesuatu visualisasi data yang umumnya berbentuk teks. Kata-kata yang terdapat ditafsirkan dalam sesuatu visualisasi dimana besar

kecilnya kata dalam gambar bergantung pada frekuensi kemunculan kata tersebut dalam sebuah *dataset*. Sempelnya, kata terbanyak ialah modus dalam data teks tersebut.

4. HASIL DAN PEMBAHASAN

4.1 Hasil *Scrapping* Data

Hasil pengumpulan data dengan teknik *scrapping* pada *google colab* dengan menggunakan *library* *pandas* pada aplikasi *Ojol The Game* menarik data sebesar 1.198 ulasan dari tanggal 1 Januari 2024 sampai dengan 15 Maret 2024. Data yang didapat menarik beberapa atribut seperti *review id*, *user name*, *user image*, *content*, *score*, *thumbs up count*, *review created version*, *at*, *reply content*, *reply at*, *app version*, *sort order*, dan *app id*. Tetapi dari semua atribut yang didapat, hanya atribut *content* dan *score* saja yang akan diproses dalam kedalam *rappid miner*. Berikut proses *scrapping* dapat dilihat pada gambar 4.1

```
[ ] from google.colab import drive
drive.mount('/content/gdrive', force_remount=True)
path = "gdrive/MyDrive/"

Mounted at /content/gdrive

[ ] app_packages = [
    'codexplore.ojol'
]

[ ] app_reviews = []

for ap in tqdm(app_packages):
    for score in list(range(1,6)):
        for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]:
            rvs, _ = reviews(
                ap,
                lang='id',
                country='id',
                sort=sort_order,
                count=1000 if score == 3 else 100,
                filter_score_with = score
            )
```

Gambar 4.1 Proses *Scrapping* Data

4.2 *Preprocessing*

Preprocessing merupakan tahapan kedua setelah *scrapping data*. pada tahapan ini dilakukan pembersihan *dataset* dengan menggunakan beberapa tahapan *filter* seperti *filter tokenize*, *filter transform cases*, *filter stopwords*, *filter tokens*. Pada *filter* pertama yaitu *filter tokenize* yang memiliki fungsi untuk memecah kalimat menjadi kata-kata dan menghapus karakter-karakter yang terdapat pada teks. Hasil dari *filter tokenize* dapat dilihat pada tabel 4.1

Tabel 4.1 Hasil *Filter Tokenize*

Sebelum <i>Tokenize</i>	<i>Filter</i>	Sesudah <i>Tokenize</i>	<i>Filter</i>
----------------------------	---------------	----------------------------	---------------

<p>Bagus! Untuk game nya juga sangat asik,tapi menurut saya tolong perbaiki bug nya Saya sering terkena bug di game ojol the game ini...dan mungkin banyak juga yang terkena bug tersebut,entah rumah dan orang orang yang tiba tiba menghilang,dan sebagainya...sisanya sudah bagus dan memuaskan. Itu saja pendapat saya tentang game ini! Terimakasih dan semoga cepat diperbaiki untuk bug nya...makasii 4</p>	<p>Bagus Untuk game nya juga sangat asik tapi menurut saya tolong perbaiki bug nya Saya sering terkena bug di game ojol the game ini dan mungkin banyak juga yang terkena bug tersebut entah rumah dan orang orang yang tiba tiba menghilang dan sebagainya sisanya sudah bagus dan memuaskan Itu saja pendapat saya tentang game ini Terimakasih dan semoga cepat diperbaiki untuk bug nya makasii</p>
--	---

Filter kedua yaitu *filter transform case*, *filter* ini berfungsi untuk mengubah kata-kata yang memiliki huruf besar menjadi huruf kecil atau disebut dengan *lowercases*. Hasil dari proses *filter transform cases* dapat dilihat pada tabel 4.2

Tabel 4.2 Hasil *Filter Transform Cases*

Sebelum <i>Filter Transform Case</i>	<i>Filter</i>	Sesudah <i>Filter Transform Case</i>	<i>Filter</i>
<p>Bagus Untuk game nya juga sangat asik tapi menurut saya tolong perbaiki bug nya Saya sering terkena bug di game ojol the game ini dan mungkin banyak juga yang terkena bug tersebut entah rumah dan orang orang yang tiba tiba menghilang dan sebagainya sisanya sudah bagus dan memuaskan Itu saja pendapat saya tentang game ini Terimakasih dan semoga cepat diperbaiki untuk bug nya makasii</p>		<p>bagus untuk game nya juga sangat asik tapi menurut saya tolong perbaiki bug nya saya sering terkena bug di game ojol the game ini dan mungkin banyak juga yang terkena bug tersebut entah rumah dan orang orang yang tiba tiba menghilang dan sebagainya sisanya sudah bagus dan memuaskan itu saja pendapat saya tentang game ini terimakasih dan semoga cepat diperbaiki untuk bug nya makasii</p>	

Filter ketiga yaitu *filter stopwords* yang berfungsi untuk menghapus kata-kata yang tidak memiliki arti atau makna contohnya seperti kata “di”, “dan”, “entah” dan lain lain. Hasil dari *filter stopwords* dapat dilihat pada tabel 4.3

Tabel 4.3 Hasil *Filter Stopwords*

Sebelum <i>Filter Stopwords</i>	Sesudah <i>Filter Stopwords</i>
Bagus untuk game nya juga sangat asik tapi menurut saya tolong perbaiki bug nya saya sering terkena bug di game ojol the game ini dan mungkin banyak juga yang terkena bug tersebut entah rumah dan orang orang yang tiba tiba menghilang dan sebagainya sisanya sudah bagus dan memuaskan itu saja pendapat saya tentang game ini terimakasih dan semoga cepat diperbaiki untuk bug nya makasii	seluruhan nya bagus grafiknya lumayan seru dah bug pas jalan tambahin pejalan kaki kayak dlu yg dlu bagus yg segitu aja sihh kalo pejalan kaki bug nya benerin auto bintang dah

Filter ke empat yaitu *filter tokens* yang berfungsi untuk menghapus kata-kata yang kurang dari 3 huruf dan tidak lebih dari 25 huruf. Hal ini perlu dilakukan untuk mempermudah *machine learning* dalam melakukan klasifikasi data. Hasil dari *filter tokens* dapat dilihat pada tabel 4.4

Tabel 4.4 Hasil *Filter Tokens*

Sebelum <i>Filter Tokens</i>	Sesudah <i>Filter Tokens</i>
seluruhan nya bagus grafiknya lumayan seru dah bug pas jalan tambahin pejalan kaki kayak dlu yg dlu bagus yg segitu aja sihh kalo pejalan kaki bug nya benerin auto bintang dah	seluruhan bagus grafiknya lumayan seru jalan tambahin pejalan kaki kayak bagus segitu sihh kalo pejalan kaki benerin auto bintang

Setelah dilakukan pembersihan data, tahap berikutnya yaitu *splitting* data. Hal ini dilakukan untuk membagi data latih yang akan digunakan sebagai metode pembelajaran bagi *machine learning* pada model naïve bayes, dan data uji yang akan digunakan sebagai *dataset* yang akan diuji dari hasil pembelajaran naïve bayes sebelumnya. Untuk mendapatkan tingkat keberhasilan yang tinggi maka diperlukan data latih yang lebih banyak dengan *rasio* perbandingan 80 : 20. Data latih akan diberikan pelabelan secara manual oleh penulis berdasarkan *score* yang diberikan oleh pengguna aplikasi *Ojol*

The Game. Apabila pengguna memberikan *score* 3, 4, dan 5 maka penulis akan menganggap sebagai sentimen positif. Jika pengguna memberikan *score* 1 dan 2 maka penulis akan memberikan label sebagai sentimen negatif. Pada data uji akan diberikan label secara otomatis oleh naïve bayes dirappid miner berdasarkan dari hasil pembelajaran pada data latih. Proses pembelajaran yang dilakukan oleh naïve bayes dengan cara melihat atau menghitung probabilitas dari setiap klasifikasi yang dilakukan. Pada tabel 4.5 menunjukkan hasil pembagian jumlah data latih dan data uji yang sudah diklasifikasi atau diberikan label.

Tabel 4.5 Pembagian Jumlah Data Latih dan Data Uji

Klasifikasi	Jumlah	Data Latih (80%)	Data Uji (20%)
Positif	743	600	143
Negatif	455	358	97
Total	1.198	958	240

4.3 Hasil Klasifikasi Naïve Bayes

Terdapat rumusan untuk menentukan ulasan pengguna *Ojol The Game* masuk kedalam label positif atau negatif dengan cara membuat kategori kelas positif (1) dan kelas negatif (0). Berikut merupakan salah satu ulasan dari data uji “seluruhan bagus grafiknya lumayan seru jalan tambahin pejalan kaki kayak bagus segitu sihh kalo pejalan kaki benerin auto bintang”

Tabel 4.6 Contoh Sentimen Dari Data Uji

Kata	Positif	Negatif
Seluruhan	1	0
Bagus	347	103
Grafiknya	14	3
Lumayan	46	17
Seru	212	57
Jalan	107	25
Tambahin	56	5
Pejalan	9	2
Kaki	11	3
Kayak	34	12
Bagus	374	103
Segitu	14	1
Sihh	5	4
Kalo	210	42
Pejalan	9	2
Kaki	11	3
Benerin	8	8
Auto	1	0
Bintang	125	61
Total	1.594	451

$$\begin{aligned} \text{Class Positif (1)} &= \frac{1}{1} \times \frac{347}{450} \times \frac{14}{17} \times \frac{46}{63} \times \\ &\frac{212}{269} \times \frac{107}{132} \times \frac{56}{61} \times \frac{9}{11} \times \frac{11}{14} \times \frac{34}{46} \times \\ &\frac{269}{374} \times \frac{14}{15} \times \frac{5}{9} \times \frac{210}{252} \times \frac{9}{11} \times \frac{11}{14} \times \frac{8}{16} \times \\ &\frac{1}{1} \times \frac{125}{186} \times \frac{743}{1198} \\ &= \\ &\frac{386.411.519.338.890.000.000.000.000.000,00}{65.887.000.597.957.900.000.000.000.000.000,00} \\ &= 0,005864761 \end{aligned}$$

$$\begin{aligned} \text{Class Negatif (0)} &= \frac{0}{1} \times \frac{103}{450} \times \frac{3}{17} \times \frac{17}{63} \times \\ &\frac{57}{269} \times \frac{25}{132} \times \frac{5}{61} \times \frac{2}{11} \times \frac{3}{14} \times \frac{12}{46} \times \frac{103}{477} \times \\ &\frac{1}{15} \times \frac{4}{9} \times \frac{42}{252} \times \frac{2}{11} \times \frac{3}{14} \times \frac{8}{16} \times \frac{0}{1} \times \\ &\frac{61}{186} \times \frac{455}{1198} \\ &= \\ &0 \\ &= \frac{\text{jumlah (class=1)}}{(\text{jumlah (class=0)} + \text{jumlah (class=1)})} \\ &= \frac{0,005864761}{0+0,005864761} \\ &= 1 \end{aligned}$$

Dari hasil perumusan dapat disimpulkan bahwa kalimat tersebut masuk kedalam kelas "Positif"

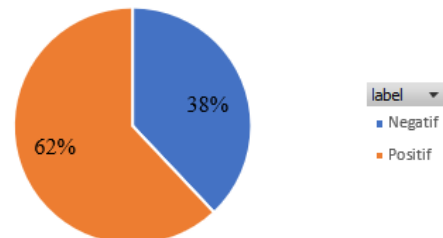
Berdasarkan perhitungan diatas probabilitas ulasan pada data uji yang telah diklasifikasi oleh naïve bayes menghasilkan sentimen positif dengan nilai sebesar 0,005864761, sedangkan hasil dari sentimen negatif yaitu 0. Karena hasil nilai dari sentimen positif lebih besar dari sentimen negatif maka kalimat tersebut masuk kedalam kelas sentimen positif. Berikut merupakan contoh dari pelabelan sentimen positif dan sentimen negatif yang diberikan secara otomatis oleh naïve bayes dapat dilihat pada tabel 4.7

Tabel 4.7 Hasil Pelabelan Sentimen Positif dan Negatif

Teks	Sentimen
gamenya bagus banget kasih bintang gamenya tolong dikembangkan tambahkan misi misi perluas tambahkan beli rumah terbeli bosan gada tujuan saran makasih	Positif

kecewa karna level lumayan riset ngulang capek capek ngumpulin uang naikin levelnya	Negatif
---	---------

Hasil Persentase Ulasan Ojol The Game



Gambar 4.2 Diagram Sentimen Ojol The Game

Berikut merupakan perhitungan dari hasil klasifikasi sentimen positif dan sentimen negatif. Total data sentimen positif dibagi dengan total keseluruhan data, dan total data sentimen negatif dibagi dengan total keseluruhan data.

Presentase dari sentimen positif

$$= \frac{743}{1198} \times 100\% = 62,02\%$$

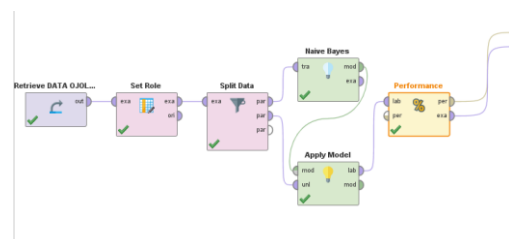
Presentase dari sentimen negatif

$$= \frac{455}{1198} \times 100\% = 37,97\%$$

Berdasarkan hasil dari model naïve bayes classifier didapatkan klasifikasi dengan prediksi sentimen positif berjumlah 743 ulasan dengan *presentase* 62,02%. Dan untuk klasifikasi sentimen negatif diprediksi berjumlah 455 ulasan dengan *presentase* 37,97% dengan total 1.198 ulasan. Hal ini menunjukkan bahwa aplikasi *Ojol The Game* memiliki sentimen positif lebih banyak dibandingkan sentimen negatif.

4.4 Confusion Matrix

Untuk mengetahui tingkat *accuracy* dari model naïve bayes maka pada rapid miner diperlukan atribut performance untuk melihat seberapa besar nilai akurasi yang dihasilkan oleh model naïve bayes dalam mengklasifikasi suatu data.



Gambar 4.3 Proses *Confusion Matrix*

Tabel 4.8 Hasil *Confusion Matrix*

	True Positif	True Negatif
Prediksi Positif	146	9
Prediksi Negatif	3	82

Hasil prediksi dapat dilihat pada tabel 4.8. Banyaknya informasi yang berkategori negatif akan diprediksi negatif (diprediksi dengan pas) oleh *machine learning* yang disebut dengan *true* negatif dan menghasilkan sebanyak 82 data. Banyaknya informasi yang berkategori positif akan diprediksi positif (diprediksi dengan pas) oleh *machine learning* yang disebut dengan *true* positif, dengan jumlah sebanyak 146 data. Banyaknya informasi yang berkategori positif namun ada kesalahan prediksi yang disebut dengan *false* positif mendapatkan 3 data. Banyaknya informasi yang berkategori negatif namun ada kesalahan prediksi yang disebut dengan *false* negatif mendapatkan sebanyak 9 data. Untuk memastikan ketepatan peneliti menggunakan perhitungan manual pada confusion matrix yang dimana akan menghitung nilai *accuracy*, *recall* dan *precision*. Berikut rumus dari confusion matrix.

Arti singkatan :

TP = True Positif

TN = True Negatif

FP = False Positif

FN = False Negatif

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{146+82}{146+82+9+3}$$

$$= \frac{228}{240} \times 100\% = 95,00\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{146}{146+3}$$

$$\frac{146}{149} \times 100\% = 97,98\%$$

$$Presicion = \frac{TP}{TP+FP} = \frac{146}{146+9}$$

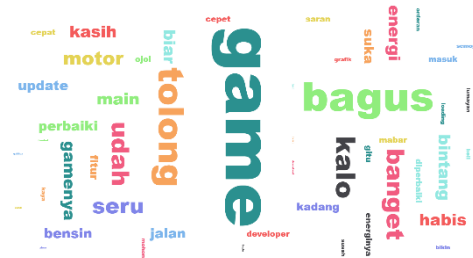
$$= \frac{146}{155} \times 100\% = 94,19\%$$

Berdasarkan hasil *confusion matrix* didapat nilai *accuracy* sebesar 95.00%, *recall* 97.98%, dan *precision* 94.19%. Hasil perhitungan manual tersebut sama dengan hasil perhitungan di *software* rapid miner.

4.5 Wordcloud

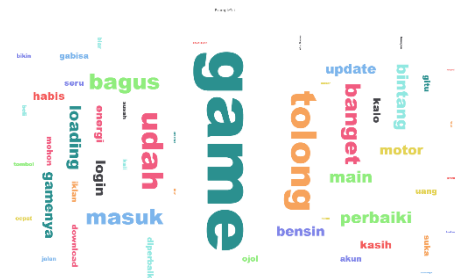
Tahapan terakhir yaitu *wordcloud* yang berfungsi untuk memahami gagasan atau topik yang sering dibicarakan oleh pengguna aplikasi *Ojol The Game*. *Wordcloud* sendiri merupakan teknik yang

digunakan untuk melihat frekuensi kemunculan suatu kata, semakin banyak frekuensi kata tersebut maka *wordcloud* akan memvisualisasikan kata tersebut dengan ukuran yang besar, dan sebaliknya semakin sedikit frekuensi kemunculan suatu kata maka *wordcloud* akan memvisualisasikan kata tersebut dengan ukuran yang kecil. Berikut merupakan visualisasi dari *wordcloud* dapat dilihat pada gambar 4.4 dan 4.5



Gambar 4.4 Wordcloud Sentimen Positif

Pada gambar 4.4 menunjukkan visualisasi dari kata-kata yang sering muncul di sentimen positif. Dapat dilihat bahwa kata *game* merupakan visualisasi yang paling besar, hal ini dikarenakan kata *game* memiliki jumlah frekuensi sebanyak 458 kata, dan frekuensi kedua yang sering muncul yaitu kata bagus dengan jumlah 288 kata, dan kata tolong berjumlah 255 dan seterusnya.



Gambar 4.5 Wordcloud Sentimen Negatif

Pada gambar 4.5 menunjukkan visualisasi dari kata-kata yang sering muncul di sentimen negatif. Dapat dilihat bahwa kata *game* merupakan visualisasi yang paling besar, kata *game* pada sentimen negatif memiliki jumlah frekuensi sebanyak 243 kata, dan frekuensi kedua yang sering muncul yaitu kata tolong dengan jumlah 127 kata, dan kata ketiga yaitu udah berjumlah 109 dan seterusnya. Jika dilihat kata-kata yang muncul merujuk pada kualitas dari *aplikasi Ojol The Game*. Yang artinya pengguna memberikan keluhan dan masukkan terkait aplikasi yang mereka gunakan.

5. KESIMPULAN

- 1) Kesimpulan dari penelitian ini adalah analisis sentimen ulasan pengguna aplikasi

Ojol The Game menggunakan algoritma naïve bayes mendapatkan hasil yang memuaskan. Penelitian ini menggunakan data sebanyak 1.198 ulasan dari aplikasi *Ojol The Game* yang ada di *google play store*, data yang didapat telah di melewati beberapa tahapan dimulai dari pembersihan data dengan *filter tokenize*, *filter transform cases*, *filter stopwords*, *filter tokens*. Hasil akhir dari penelitian ini yaitu dilakukannya *confusion matrix* dan *wordcloud*. Penelitian menggunakan perbandingan data 80:20 untuk data uji dan data latih.

- 2) Hasil klasifikasi sentimen menggunakan metode Naïve Bayes Classifier dengan *tools* RapidMiner memperoleh tingkat akurasi sebesar 95,00%, *recall* 97,98%, dan presisi 94,19%, yang artinya metode Naïve Bayes mampu memberikan prediksi sentimen dengan tepat dan menghasilkan data yang akurat.
- 3) Hasil dari penelitian ini memperoleh data sentimen positif lebih banyak dibandingkan sentimen negatif. Klasifikasi sentimen positif mendapatkan 62% dari 743 data dan untuk sentimen negatif mendapat 38% dari 455 data. Hasil ini menunjukkan bahwa ulasan dari pengguna aplikasi *Ojol The Game* tidak seimbang, karena banyak pengguna yang menyukai aplikasi tersebut.

UCAPAN TERIMA KASIH

Pada kesempatan ini saya ingin mengucapkan banyak terima kasih kepada :

1. Dua orang yang paling berjasa dalam hidup saya, Mama dan Bapak. Terima kasih atas kepercayaan serta pengorbanan, cinta dan motivasi, semangat dan juga nasihat.
2. Adik terkasih Ulfa dan Mega yang memberikan semangat dan celotehannya, tapi saya yakin itu adalah bentuk dukungan dan motivasi.
3. Keluarga besar yang selalu memberikan dukungan baik secara moril maupun material.
4. Teman – teman yang selalu support
5. Terima kasih untuk diri sendiri, Sumini Aulia sudah menepikan ego dan memilih untuk kembali bangkit dan menyelesaikan semua ini. Terima kasih telah mengendalikan diri dari berbagai tekanan di luar keadaan dan tak pernah mau memutuskan untuk menyerah, kamu hebat...

DAFTAR PUSTAKA

- [1] Primandani Arsi, Pungkas Subarkah, and Bagus Adhi Kusuma, "Analisis Sentimen *Game Genshin Impact* pada *Play Store* Menggunakan Naïve Bayes Clasifier," *J. Ilm. Tek. Mesin, Elektro dan Komput.*, vol. 3, no. 1, pp. 161–170, 2023, doi: 10.51903/juritek.v3i1.1962.
- [2] F. Fandhilah and A. Ardiansyah, "Analisis Sentimen Aplikasi Quizizz pada *Google PlayStore* Menggunakan NaïveBayes," *SATIN –Sains dan Teknol. Inf.*, vol. 9, no. 2, pp. 176–183, 2023, doi: 10.33372/stn.v9i2.1056.
- [3] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Inf. Syst. Public Heal.*, vol. 4, no. 1, pp. 28–36, 2019.
- [4] A. N. Hasanah and B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada *Google Play* Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, pp. 90–96, 2024, doi: 10.23960/jitet.v12i1.3628.
- [5] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi *Google Play* Menggunakan Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.
- [6] M. Haikal, M. Martanto, and U. Hayati, "Analisis Sentimen Terhadap Penggunaan Aplikasi *Game Online Pubg Mobile* Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3275–3281, 2024, doi: 10.36040/jati.v7i6.8174.
- [7] G. Darmawan, S. Alam, and M. I. Sulisty, "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.
- [8] V. A. W and A. W. -, "Analisis Sentimen Ulasan Aplikasi *Capcut* Menggunakan Metode Naive Bayes," *J. Sains Sist. Inf.*, vol. 2, no. 1, p. 9, 2024, doi: 10.31602/jssi.v2i1.13372.
- [9] M. F. Y. Herjanto and C. Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi *Sirekap* Pada *Play Store* Menggunakan Algoritma Random Forest Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [10] A. Nurian, T. N. Padilah, and G. Garro, "Analisis Sentimen Terhadap Pelayanan *Disdukcapil Karawang* Menggunakan Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4178.
- [11] I. Syahrohim, S. D. Saputra, R. W. Saputra, V. H. Pranatawijaya, and R. Priskila, "Perbandingan Analisis Sentimen Setelah

- Pilpres 2024 Di Twitter Menggunakan Algoritma Machine Learning,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4249.
- [12] M. Afriansyah, Joni Saputra, V. Y. P. Ardhana, and Yuan Sa’adati, “Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna,” *J. Inf. Syst. Manag. Digit. Bus.*, vol. 1, no. 2, pp. 236–248, 2024, doi: 10.59407/jismdb.v1i2.438.