

# Sentiment Analysis of Chatting Application Reviews on Google Play Store Using Naïve Bayes Classifier Algoritm

Muhammad Luthfi Alfathan\*, Nufhika Fishuri, Dodi Vionanda

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [lutfil466@gmail.com](mailto:lutfil466@gmail.com)

Submitted : 10 Februari 2025

Revised : 20 Februari 2025

Accepted : 28 Februari 2025

## ABSTRACT

*Chatting application is a medium used to connect two or more people through social media platforms. Based on the results of the survey report, there are 5 chat applications that are often used as a medium of communication, including WhatsApp, Facebook, Telegram, Instagram and Line applications. This research aims to see the sentiment of chat application user and see how naive bayes performs in analyzing the sentiment of chat application users. The purpose of sentiment analysis in this research is to assess whether a comment related to an issue is negative or positive, provide a guide in improving the quality or service of a product. From the analysis results obtained, the Naïve Bayes model showed mixed performance depending on the type of application and sentiment. The model generally showed better performance in identifying positive reviews, especially on Facebook, Telegram, and Instagram apps, where recall reached 100%. However, the model performed very poorly in identifying neutral reviews across all apps. To increase accuracy and more balanced sentiment detection capabilities, improvements in data preprocessing, handling data imbalance, or the use of more complex classification methods are needed.*

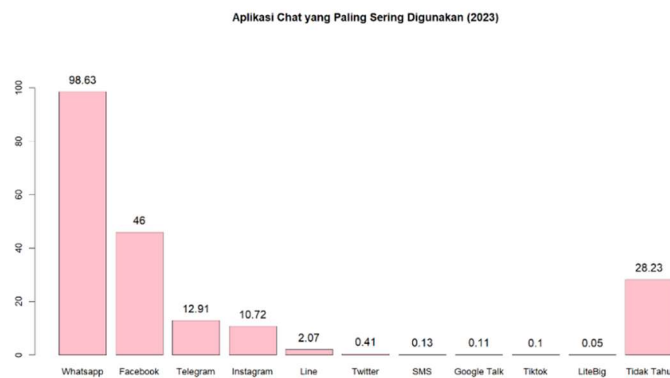
**Keywords:** *Chatting Application, Scrapping, Sentiment Analysis, Naïve Bayes Classifier*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Aplikasi chatting merupakan media yang digunakan untuk menghubungkan dua orang atau lebih melalui platform media sosial. Berdasarkan hasil laporan survey Asosiasi Penyedia Jaringan Internet Indonesia (APJII) dengan 8.510 jiwa responden yang tersebar di berbagai provinsi yang ada di Indonesia diperoleh 5 aplikasi chatting yang sering digunakan sebagai media komunikasi, diantaranya yaitu aplikasi *WhatsApp*, *Facebook*, *Telegram*, *Instagram* dan *Line*.(Databoks.katadata.co.id 2023)



**Gambar 1.** Grafik Persentase Aplikasi Chat yang Paling Sering Digunakan (2023)

Aplikasi tersebut banyak digunakan karena memberi kemudahan bagi para pengguna untuk melakukan percakapan pribadi atau grup, tetapi hal itu tidak bisa menjadi penilaian kepuasan untuk pengguna aplikasi tersebut karena masih terdapat komentar negative yang tertera pada bagian komentar di *Google Play Store* (Saepulrohman, Saepudin, and Gustian 2021).

*Google Play Store* merupakan toko resmi untuk mengunduh aplikasi berbasis android. Di dalam *Google Play Store* terdapat fitur *rating* dan ulasan yang digunakan oleh pengguna untuk memberikan komentar atau menilai seberapa baik aplikasi itu bekerja. Saat ini, banyak pengguna yang mengandalkan *rating* dan ulasan tersebut sebagai bahan pertimbangan untuk menggunakan suatu produk (Masturoh and Pohan 2021).

Penelitian ini bertujuan untuk melihat sentimen pengguna aplikasi chatting, dan melihat bagaimana performa *naive bayes* dalam menganalisis sentimen pengguna aplikasi chatting. Analisis sentimen (*opinion mining*) merupakan jenis pemrosesan bahasa alami yang menggunakan komponen komputasi linguistik, analisis teks, dan verifikasi biometric untuk menganalisis, mengekstrak, mengukur dan mengevaluasi informasi yang unik dan terstruktur (Bhatia, Sharma, and Bhatia 2018). Tujuan dari analisis sentimen pada penelitian ini yaitu untuk menilai apakah suatu komentar yang terkait suatu isu bersifat negatif atau positif, juga sebagai pedoman dalam meningkatkan kualitas atau layanan suatu produk.

Klasifikasi merupakan proses penemuan model yang membedakan kelas data yang bertujuan untuk memprediksi kelas dari objek yang tidak berlabel dan salah satu metode *supervised learning* (Kamber and Han 2006). Salah satu metode klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier* (NBC) karena, ulasan pada play store merupakan data yang tidak terstruktur sehingga penggunaan metode yang sederhana serta memiliki akurasi yang tinggi seperti metode NBC dapat dilakukan untuk mengklasifikasikan ulasan tersebut (Septianingrum and Irawan 2021)

## II. METODE PENELITIAN

### A. Jenis Data dan Sumber Data

Jenis data yang digunakan dalam penelitian adalah jenis data sekunder, yaitu sekumpulan ulasan pengguna aplikasi *chatting* (*WhatsApp*, *Facebook*, *Telegram*, *Instagram* dan *Line*) di Indonesia yang terdapat pada lama *Google Play Store*.

### B. Variabel penelitian

Variabel penelitian yang digunakan pada penelitian ini yaitu opini atau sentimen dari pengguna aplikasi chatting (*Whatsapp*, *Facebook*, *Telegram*, *Instagram*, dan *Line*) sebagai variabel dependen, dan frekuensi kata ke-i yang muncul pada objek (ulasan) sebagai variabel independen.

Table 1. Variabel untuk klasifikasi sentimen

Variabel	Keterangan
Y	0 = Sentimen Negatif 1 = Sentiimen positif
X	Frekuensi kata ke-I yang muncul pada objek (ulasan)

### C. Langkah -Langkah Analisis Data

#### 1. Pengumpulan Data

Pengambilan data dilakukan dengan cara scraping menggunakan web *Google Colaboratory*. Proses scraping ini menggunakan bahasa pemrograman python dengan memasukkan link aplikasi *WhatsApp*, *Facebook*, *Telegram*,

Instagram dan Line yang terdapat pada laman *Google Play Store* ke dalam perintah yang telah dibuat pada *Google Colaboratory*.

## 2. Proses pelabelan data

Pelabelan data dilakukan untuk mengidentifikasi kata-kata positif dan kata-kata negatif dalam suatu ulasan. Sebelum melakukan pelabelan, data diterjemahkan kedalam Bahasa Inggris karena bahasa yang tersedia pada kamus *lexicon* adalah Bahasa Inggris. Setelah proses penerjemahan, dilakukan proses perhitungan *compound score* menggunakan persamaan.

$$\text{compound score (cs)} = \frac{x}{\sqrt{x^2 + \alpha}} \quad (1)$$

Keterangan:

$x$  = total *score polarity* dari setiap *term* yang terdapat dalam dokumen

$\alpha$  = konstanta

Jika nilai *compound score* lebih besar dari 0 maka ulasan tersebut masuk kedalam kelas sentiment negatif.

## 3. Preprocessing Data

### a. Proses *Cleaning Data*

*Cleaning* data dilakukan untuk menghapus elemen-elemen yang tidak perlu ada atau mengganggu ulasan, seperti emoji, tanda baca, dan lain-lain. Proses ini bertujuan untuk mengurangi *noise*, dan mencegah adanya data duplikat, data *outlier*, dan data tidak valid.

### b. *Case Folding*

*Case folding* bertujuan untuk menyamakan kapitalisasi dalam teks ulasan menjadi *lowercase* untuk menghindari *case sensitivitas* pada saat melakukan analisis.

### c. Tokenisasi

Proses ini bertujuan untuk memisahkan kalimat dalam teks ulasan menggunakan spasi atau tanda baca menjadi token-token kata.

### d. *Stopword Removal*

*Stopword Removal* digunakan untuk mengeleminasi kata-kata umum yang sering muncul tetapi tidak berpengaruh dalam teks.

### e. *Stemming*

*Stemming* merupakan proses untuk mengurangi kata-kata berimbuhan supaya kembali kedalam bentuk kata dasar dari kata-kata yang telah melalui proses tokenisasi.

## 4. Pemodelan klasifikasi sentimen

Pemodelan klasifikasi sentimen menggunakan *Naïve Bayes Classifiers* sebelumnya, data akan dipisah menjadi data *training* dan data *testing*. Data *training* digunakan untuk bahan pembelajaran oleh model, sedangkan data *testing* digunakan untuk mengukur kinerja model yang telah dibentuk. Pembagian dari data *training* dan data *testing* yaitu 80:20. (Hamzah 2012)

*Naïve Bayes Classifiers* (NBC) merupakan metode pengklasifikasian statistik berdasarkan teorema Bayes, untuk memprediksi apa yang terjadi di masa depan dengan melihat kejadian sebelumnya dengan menghitung probabilitas. (Tanggraeni & Sitokdana 2022). Untuk menghitung nilai probabilitas NBC disajikan sebagai berikut.

$$P(v_j|a_i) = \frac{P(a_i|v_j)P(v_j)}{P(a_i)} \quad (2)$$

Setelah penghitungan nilai probabilitas NBC, selanjutnya mencari nilai probabilitas tertinggi dari semua kategori kelas yang telah diuji dengan mencari nilai *Maximum A Posteriori* (MAP) menggunakan persamaan sebagai berikut.

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j) \times \prod_i P(a_i | v_j) \tag{3}$$

Nilai  $P(v_j)$  yang dihitung saat *training* dan nilai  $P(v_j | a_i)$  untuk probabilitas kata  $a_i$  setiap kategori diperoleh dari persamaan.

$$P(v_j) = \frac{docs_j}{training} \tag{4}$$

$$P(a_i | v_j) = \frac{n_i + 1}{n + kosakata} \tag{5}$$

Keterangan:

$P(v_j | a_i)$  = Peluang kategori kelas  $j$  ketika kata ke- $i$  muncul

$P(a_i | v_j)$  = Peluang kata ke- $i$  terdapat dalam kategori kelas  $j$

$P(v_j)$  = Peluang jumlah kejadian kategori kelas  $j$

$P(a_i)$  = Peluang munculnya kata ke- $i$

$V_{MAP}$  = Nilai *Maximum A Posteriori*

$docs_j$  = Jumlah dokumen dalam kategori  $j$

5. Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix*. Dalam pengukuran *confusion matrix* terdapat beberapa istilah True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). (Kamber and Han 2006). Pada *confusion matrix* ada beberapa nilai yang akan dihitung yaitu *accuary*, *recall*, dan *precision*. Berikut tabel perhitungan *confusion matrix* dan persamaannya.

**Table 2.** *Confusion Matrix*

Actual Class	Predicted Class	
	Yes	No
Yes	TP ( <i>True Positive</i> )	FP ( <i>False Positive</i> )
No	FN ( <i>False Negative</i> )	TN ( <i>True Negative</i> )

$$Accuary = \frac{TN + TP}{TN + TP + FN + FP} \times 100\% \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Dataset yang digunakan adalah ulasan aplikasi *chatting* pada *goole play store* yang diambil menggunakan teknik *scraping* dari *library google-play-scraper*. Pengambilan data dilakukan pada tanggal 26 Mei 2024. Proses pengambilan data melalui *website google collaborator* dengan memasukkan link aplikasi yang terdapat di laman *Google Play Store* ke dalam perintah yang telah disusun dalam bahasa pemrograman python.. Setelah data ulasan diperoleh, langkah selanjutnya adalah mengubahnya ke dalam bentuk data frame, di mana hanya fitur-fitur yang relevan seperti isi ulasan dan rating yang dipilih untuk dipertahankan. Kemudian, dataframe yang telah disusun tersebut disimpan dalam format csv dan xlsx agar dapat diolah pada tahap analisis selanjutnya. Data yang diperoleh sebagai berikut.

##### a. Whatsapp

Data yang dikumpulkan pada aplikasi *Whatsapp* menggunakan teknik *scraping* dari *library google-play-scraper* didapatkan data 995 data.

##### b. Facebook

Data yang dikumpulkan pada aplikasi *Facebook* menggunakan teknik *scraping* dari *library google-play-scraper* didapatkan data 398 data

##### c. Telegram

Data yang dikumpulkan pada aplikasi *Telegram* menggunakan teknik *scraping* dari *library google-play-scraper* didapatkan data 199 data

##### d. Instagram

Data yang dikumpulkan pada aplikasi *Instagram* menggunakan teknik *scraping* dari *library google-play-scraper* didapatkan data 796 data

##### e. Line

Data yang dikumpulkan pada aplikasi *Line* menggunakan teknik *scraping* dari *library google-play-scraper* didapatkan data 597 data

#### B. Preprocessing Data

Pada tahapan ini, review (*content*) mentah diproses dulu pada tahap preprocessing (*cleaning, case folding, stopword, stemming, tokenisasi*), Tahapan preprocessing ditampilkan sebagai berikut.

##### a. Whatsapp

**Tabel 3.** Hasil *Preprocessing Data Whatsapp*

Proses	Hasil
<i>Dataset</i>	Akun ku di blok entah kenapa padahal akun nya di pinjem org tiba tiba di block
<i>Cleaning</i>	Akun ku di blok entah kenapa padahal akun nya di pinjem org tiba tiba di block
<i>Case Folding</i>	akun ku di blok entah kenapa padahal akun nya di pinjem org tiba tiba di block
<i>Stopword</i>	akun ku blok akun nya pinjem org block
<i>Stemming</i>	akun ku blok akun nya pinjem org block
<i>Tokenisasi</i>	[akun, ku, blok, akun, nya, pinjem, org, block]

##### b. Facebook

**Tabel 4.** Hasil *Preprocessing Data Facebook*

Proses	Hasil
<i>Dataset</i>	Akun tiba tiba di banned, aneh banget

<i>Cleaning</i>	Akun tiba tiba di banned aneh banget
<i>Case Folding</i>	akun tiba tiba di banned aneh banget
<i>Stopword</i>	akun banned aneh banget
<i>Stemming</i>	akun banned aneh banget
<i>Tokenisasi</i>	[akun, banned, aneh, banget]

c. Telegram

**Tabel 5.** Hasil *Preprocessing Data Telegram*

<b>Proses</b>	<b>Hasil</b>
<i>Dataset</i>	Kenapa nomor saya di banned Ini aplikasi lama lama g bermanfaat
<i>Cleaning</i>	Kenapa nomor saya di banned Ini aplikasi lama lama g bermanfaat
<i>Case Folding</i>	kenapa nomor saya di banned ini aplikasi lama lama g bermanfaat
<i>Stopword</i>	kenapa nomor banned aplikasi lama lama g bermanfaat
<i>Stemming</i>	kenapa nomor ban aplikasi lama lama g manfaat
<i>Tokenisasi</i>	['kenapa', 'nomor', 'ban', 'aplikasi', 'lama', 'lama', 'g', 'manfaat']

d. Instagram

**Tabel 6.** Hasil *Preprocessing Data Instagram*

<b>Proses</b>	<b>Hasil</b>
<i>Dataset</i>	Akun saga selalu di nonaktifkan Secara tiba tiba
<i>Cleaning</i>	Akun saga selalu di nonaktifkan Secara tiba tiba
<i>Case Folding</i>	akun saga selalu di nonaktifkan secara tiba tiba
<i>Stopword</i>	akun saga nonaktifkan tiba tiba
<i>Stemming</i>	akun saga aktif secara tiba tiba
<i>Tokenisasi</i>	['akun', 'saga', 'aktif', 'secara', 'tiba', 'tiba']

e. Line

**Tabel 7.** Hasil *Preprocessing Data Line*

<b>Proses</b>	<b>Hasil</b>
<i>Dataset</i>	Kenapa line skrng ribet banget sih?!!!!!! Mau login aja di persulit padahal aku dah punya akun lama mlh di suruh bikin akun baru!!!!!! Kesel anj akun lama ku ga bisa di logonin Mana tuh akun ku dri tahun 2015 asu lah
<i>Cleaning</i>	Kenapa line skrng ribet banget sih Mau login aja di persulit padahal aku dah punya akun lama mlh di suruh bikin akun baru Kesel anj akun lama ku ga bisa di logonin Mana tuh akun ku dri tahun 2015 asu lah
<i>Case Folding</i>	kenapa line skrng ribet banget sih mau login aja di persulit padahal aku dah punya akun lama mlh di suruh bikin akun baru kesel anj akun lama ku ga bisa di logonin mana tuh akun ku dri tahun 2015 asu lah
<i>Stopword</i>	line skrng ribet login persulit padahal dah punya akun lama mlh suruh bikin akun baru kesel anj akun lama ga logonin akun dri tahun 2015 asu
<i>Stemming</i>	line skrng ribet login sulit padahal dah punya akun lama

	suruh bikin akun baru kesel anj akun lama ga login akun dri tahun asu
<i>Tokenisasi</i>	['line', 'skrng', 'ribet', 'login', 'sulit', 'padahal', 'dah', 'punya', 'akun', 'lama', 'suruh', 'bikin', 'akun', 'baru', 'kesel', 'anj', 'akun', 'lama', 'ga', 'login', 'akun', 'dri', 'tahun', 'asu']

Hasil dari *preposeccing* data terdapat data yang dihapuskan karena berupa data duplikat dan ulasanya menjadi kosong setelah melewati proses *stopword removal*.

### C. Pelabelan Data

Langkah berikutnya yaitu pemberian label pada data ulasan yang sudah dipersiapkan. Proses penentuan label dilaksanakan secara otomatis dengan memanfaatkan metode berbasis *lexicon*, yaitu pendekatan yang mengandalkan kamus kata *lexicon* yang telah diklasifikasikan berdasarkan nilai sentiment positif, negatif, dan netral. Dalam pendekatan ini, setiap kata dalam ulasan akan dicocokkan dengan daftar kata dalam *lexicon*. Kemudian skor sentimen dari kata-kata tersebut akan diakumulasikan untuk menentukan label akhir dari ulasan.

#### a. Whatsapp

Total Pelabelan Sentimen Positif, Netral dan Negatif dapat dilihat pada tabel

**Tabel 8.** Hasil Pelabelan Data *Whatsapp*

Sentimen	Total Analisis	Presentase (%)
Positif	643	64.62
Netral	68	6.83
Negatif	284	28.54

#### b. Facebook

Total Pelabelan Sentimen Positif, Netral dan Negatif dapat dilihat pada tabel

**Tabel 9.** Hasil Pelabelan Data *Facebook*

Sentimen	Total Analisis	Presentase (%)
Positif	316	79.40
Netral	24	6.03
Negatif	58	14.57

#### c. Telegram

Total Pelabelan Sentimen Positif, Netral dan Negatif dapat dilihat pada tabel

**Tabel 10.** Hasil Pelabelan Data *Telegram*

Sentimen	Total Analisis	Presentase (%)
Positif	71	35.68
Netral	16	8.04
Negatif	112	56.28

#### d. Instagram

Total Pelabelan Sentimen Positif, Netral dan Negatif dapat dilihat pada tabel

**Tabel 11.** Hasil Pelabelan Data *Instagram*

Sentimen	Total Analisis	Presentase (%)
Positif	647	81.28
Netral	24	3.01

Negatif	125	15.70
---------	-----	-------

e. *Line*

Total Pelabelan Sentimen Positif, Netral dan Negatif dapat dilihat pada tabel

**Tabel 12.** Hasil Pelabelan Data *Line*

Sentimen	Total Analisis	Presentase (%)
Positif	224	37.52
Netral	47	7.87
Negatif	326	54.60

**D. Pemodelan Klasifikasi Sentimen**

Data yang sudah dilakukan pelabelan, selanjutnya dilakukan Klasifikasi Data dengan Metode Naïve Bayes yang berguna untuk mengetahui keakurasian model terhadap masing-masing sentimen

a. *Whatsapp*

Klasifikasi Data dengan Metode *Naïve Bayes Classifiers* pada aplikasi *Whatsapp*

**Tabel 13.** Hasil Evaluasi Model *Whatsapp*

Sentimen	<i>Precision</i>	<i>recall</i>	<i>Accuary</i>
negatif	0.54	0.35	0.6984925
netral	0.00	0.00	
positif	0.73	0.90	

Model berhasil mengklasifikasikan sekitar 69.8% ulasan dengan benar. Semua ulasan yang diprediksi negatif, 54% memang benar negative dan semua ulasan yang benar-benar negatif, hanya 35% yang berhasil diidentifikasi oleh model. Semua ulasan yang benar-benar netral, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang diprediksi positif, 73% memang benar positif dan semua ulasan yang benar-benar positif, 90% berhasil diidentifikasi oleh model.

b. *Facebook*

Klasifikasi Data dengan Metode *Naïve Bayes Classifiers* pada aplikasi *Facebook*

**Tabel 14.** Hasil Evaluasi Model *Facebook*

Sentimen	<i>Precision</i>	<i>recall</i>	<i>Accuary</i>
negatif	1.00	0.10	0.775
netral	0.00	0.00	
positif	0.77	1.00	

Model berhasil mengklasifikasikan sekitar 77.5% ulasan dengan benar. Semua ulasan yang diprediksi negatif, 100% memang benar negative dan semua ulasan yang benar-benar negatif, hanya 10% yang berhasil diidentifikasi oleh model. Semua ulasan yang benar-benar netral, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang diprediksi positif, 77% memang benar positif dan semua ulasan yang benar-benar positif, 100% berhasil diidentifikasi oleh model.



c. *Telegram*

Klasifikasi Data dengan Metode *Naïve Bayes Classifiers* pada aplikasi *Telegram*

**Tabel 15.** Hasil Evaluasi Model *Telegram*

Sentimen	<i>Precision</i>	<i>recall</i>	<i>Accuary</i>
negatif	0.69	1.00	0.75
netral	0.00	0.00	
positif	1.00	0.50	

Model berhasil mengklasifikasikan sekitar 75% ulasan dengan benar. Semua ulasan yang diprediksi negatif, 69% memang benar negative dan semua ulasan yang benar-benar negatif, dan 100% yang berhasil diidentifikasi oleh model. Semua ulasan yang benar-benar netral, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang diprediksi positif, 100% memang benar positif dan semua ulasan yang benar-benar positif, 50% berhasil diidentifikasi oleh model.

d. *Instagram*

Klasifikasi Data dengan Metode *Naïve Bayes Classifiers* pada aplikasi *Instagram*

**Tabel 16.** Hasil Evaluasi Model *Instagram*

Sentimen	<i>Precision</i>	<i>recall</i>	<i>Accuary</i>
negatif	0.00	0.00	0.7875
netral	0.00	0.00	
positif	0.79	1.00	

Model berhasil mengklasifikasikan sekitar 78.75% ulasan dengan benar. Semua ulasan yang benar-benar negatif, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang benar-benar netral, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang diprediksi positif, 79% memang benar positif dan semua ulasan yang benar-benar positif, 100% berhasil diidentifikasi oleh model.

e. *Line*

Klasifikasi Data dengan Metode *Naïve Bayes Classifiers* pada aplikasi *Line*

**Tabel 17.** Hasil Evaluasi Model *Line*

Sentimen	<i>Precision</i>	<i>recall</i>	<i>Accuary</i>
negatif	0.70	0.97	0.725
netral	0.00	0.00	
positif	0.83	0.46	

Model berhasil mengklasifikasikan sekitar 72.5% ulasan dengan benar. Semua ulasan yang diprediksi negatif, 70% memang benar negative dan semua ulasan yang benar-benar negatif, hanya 97% yang berhasil diidentifikasi oleh model. Semua ulasan yang benar-benar netral, tidak ada yang berhasil diidentifikasi oleh model. Semua ulasan yang diprediksi positif, 83% memang benar positif dan semua ulasan yang benar-benar positif, 46% berhasil diidentifikasi oleh model.

#### IV. KESIMPULAN

Berdasarkan hasil evaluasi model Naïve Bayes Classifiers yang telah dilakukan pada lima aplikasi (*WhatsApp*, *Facebook*, *Telegram*, *Instagram*, dan *Line*), secara keseluruhan dapat disimpulkan bahwa model Naïve Bayes menunjukkan performa yang beragam tergantung pada jenis aplikasi dan sentimen. Model umumnya menunjukkan performa yang lebih baik dalam mengidentifikasi ulasan positif, terutama pada aplikasi *Facebook*, *Telegram*, dan *Instagram*, di mana recall mencapai 100%. Namun, performa model dalam mengidentifikasi ulasan netral sangat buruk di semua aplikasi. Untuk meningkatkan akurasi dan kemampuan deteksi sentimen yang lebih seimbang, diperlukan peningkatan dalam preprocessing data, penanganan ketidakseimbangan data, atau penggunaan metode klasifikasi yang lebih kompleks.

#### DAFTAR PUSTAKA

- Bhatia, Surbhi, Manisha Sharma, and Komal Kumar Bhatia. 2018. "Sentiment Analysis and Mining of Opinions." *Studies in Big Data* 30(May): 503–23.
- Databoks.katadata.co.id. 2023. "Selain WhatsApp, Ini Aplikasi Chat Yang Sering Digunakan Masyarakat Indonesia." *Databoks.katadata.co.id*. <https://databoks.katadata.co.id/datapublish/2023/06/15/selain-whatsapp-ini-aplikasi-chat-yang-sering-digunakan-masyarakat-indonesia>.
- Hamzah, A. 2012. "Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademis."
- Kamber, Micheline, and Jiawei Han. 2006. *The Fundamentals of Political Science Research Data Mining: Concepts and Techniques : Concepts and Techniques*.
- Masturoh, Siti, and Achmad Baroqah Pohan. 2021. "Analisis Sentimen Terhadap E-Wallet Dana Pada Ulasan Google Play Menggunakan Algoritma K-Nearest Neighbor." *Jurnal Pilar Nusa Mandiri* 17(1): 53–58.
- Saepulrohman, Acep, Sudin Saepudin, and Dudih Gustian. 2021. "Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine." *is The Best Accounting Information Systems and Information Technology Business Enterprise this is link for OJS usf@* 6(2): 91–105.
- Septianingrum, Fitria, and Agung Susilo Yuda Irawan. 2021. "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review." *Jurnal Media Informatika Budidarma* 5(3): 799.
- Tanggaraeni, Artanti Inez, and Melkior N. N. Sitokdana. 2022. "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes." *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)* 9(2): 785–95.