IJCIT (Indonesian Journal on Computer and Information Technology)
Journal Homepage: http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit

ANALISIS SENTIMEN TERHADAP REVIEW APLIKASI MYTELKOMSEL PADA PLAY STORE DENGAN ALGORITMA NAÏVE BAYES

Mochamad Rizki Teguh Esa Pratantio[1], Sumanto[2]
[1, 2]Informatika, Universitas Bina Sarana Informatika
Jakarta, Indonesia

e-mail: 15200104@gmail.com[1], sumanto@bsi.ac.id[2],

ABSTRAK

Perusahaan telekomunikasi seluler terbesar di Indonesia, Telkomsel, didirikan pada tanggal 26 Mei 1995, Perusahaan ini selalu berusaha untuk memberikan layanan yang lebih baik kepada pelanggannya. Oleh karena itu, Telkomsel meluncurkan aplikasi mobile bernama MyTelkomsel untuk memberikan pelayanan prima kepada penggunanya. Dengan sistem self-service, pengguna MyTelkomsel dapat berinteraksi dengan bebas Teknik scrapping untuk mengumpulkan ulasan aplikasi dari Google Play Store dengan metode klasifikasi Naïve bayes untuk menganalisis sentimen ulasan. Metode klasifikasi Naïve bayes digunakan untuk mengklasifikasikan ulasan sebagai positif dan negatif berdasarkan bahasa dan kata-kata yang digunakan dalam ulasan tersebut. Analisis sentimen adalah bidang yang berkembang pesat dalam pengolahan bahasa alami dan text mining. Ini melibatkan pengidentifikasian dan klasifikasi opini, perasaan, atau sentimen yang terkandung dalam teks, sering kali dengan tujuan memahami pandangan atau reaksi publik terhadap berbagai topik, produk, atau layanan. Dalam abstrak ini, Tujuan penelitian ini adalah untuk menganalisis ulasan user terhadap mytelkomsel yang berada di Google Play Store. Metode yang digunakan dalam penelitian ini menggunakan algoritma Naïve bayes. Dataset berisi tentang ulasan mengenai aplikasi instagram di Google Play Store sebanyak 4000 dataset. Hasil penelitian ini menunjukkan accuracy 88%, precision 88%, recall 99% dan f1-score 93%.

Katakunci: Analisis sentimen, Scrapping, Google play store, Mytelkomsel, Naïve bayes

ABSTRACTS

The largest mobile telecommunications company in Indonesia, Telkomsel, was established on May 26, 1995, The company always strives to provide better services to its customers. Therefore, Telkomsel launched a mobile application called MyTelkomsel to provide excellent service to its users. With a self-service system, MyTelkomsel users can interact freely Scrapping technique to collect app reviews from Google Play Store with Naïve bayes classification method to analyze the sentiment of reviews. The Naïve bayes classification method is used to classify reviews as positive and negative based on the language and words used in the reviews. Sentiment analysis is a rapidly growing field in natural language processing and text mining. It involves identifying and classifying opinions, feelings, or sentiments contained in text, often with the goal of understanding the public's views or reactions to various topics, products, or services. In this abstract, the purpose of this study is to analyze user reviews of mytelkomsel in the Google Play Store. The method used in this study uses the Naïve bayes algorithm. The dataset contains reviews about the Instagram application on the Google Play Store as many as 4000 datasets. The results of this study show accuracy 88%, precision 88%, recall 99% and f1-score 93%.

***Keywords**: Sentiment Analysis, Scrapping, Google play store, Mytelkomsel, Naïve bayes*

1. PENDAHULUAN

Perkembangan informasi membutuhkan dukungan agar penyebaran informasi menjadi lancar. Untuk mengakses internet, operator seluler menggunakan jaringan koneksi seluler. penyedia layanan internet (ISP) tercepat di Indonesia Telkomsel. Telkomsel memiliki kecepatan mengunduh dan mengunggah yang standar. Lebih dari 200 juta orang, atau 77% dari populasi Indonesia, sudah menggunakan Internet. Dari 2020–2022, pengguna internet meningkat 20%. Internet sering digunakan oleh masyarakat Indonesia untuk berbagai tujuan, seperti video konferensi, pembelajaran *online*, *streaming* video, bisnis *online*, dan lain-lain (Indra Cahyani et al, 2023).

Aplikasi MyTelkomsel Lebih dari 100 juta pelanggan telkomsel telah mengunduh saat ini. Jumlah pelanggan Telkomsel terus meningkat seiring dengan jumlah pengguna aplikasi MyTelkomsel Aplikasi MyTelkomsel memiliki beberapa masalah. Ini termasuk kesalahan yang sering terjadi saat melakukan transaksi, penghentian paksa aplikasi, pengisian pulsa yang tidak masuk, dan harga produk yang berubah-ubah. Aplikasi MyTelkomsel sering menampilkan harga pelanggan yang berbeda, membuat banyak pelanggan Telkomsel tidak puas (Maulana et al, 2023).

Analisis sentimen, juga disebut sebagai penambangan opini, adalah teknik untuk mengekstrak, dan mengolah informasi bacaan melalui proses penggalian data, dapat digunakan untuk menganalisis ulasan pengguna di Google Play Store yang ditulis melalui proses penggalian informasi. Analisis sentimen menemukan sentimen dalam teks dengan mengolah data teks untuk memahami sentimen tersebut (Nufairi et al, 2024).

klasifikasi Naïve bayes adalah metode sederhana, tetapi cepat dan akurat. Naïve bayes tetap dapat bekerja dengan baik dalam penelitian yang dilakukan oleh. Karena kesulitan mendapatkan komentar yang tepat, jumlah data komentar yang diambil dari Google Play bahwa Penelitian ini menggunakan algoritma Naïve bayes Penerapan Naïve bayes tergolong sangat mudah dan banyak digunakan dalam penelitian. Metode Naïve bayes juga dapat diterapkan pada domain yang berbeda. Naïve bayes merupakan klasifikasi yang sangat mudah di gunakan dan probabilitas berdasarkan statistik Teorema Bayes (Ilmawan & Mude, 2020).

Penelitian sebelumnya yang menggunakan Naïve bayes. Penulis dalam penelitian ini tertarik untuk metode tersebut. dari referensi-referensi penelitian sebelumnya, penelitian ini akan menganalisis sentimen ulasan pengguna aplikasi MyTelkomsel menggunakan algoritma Naïve bayes.

Dengan alasan diatas, Penelitian ini memiliki tujuan untuk melakukan Analisa terhadap ulasan atau Penilaian pengguna dan mengklasifikasikan Ulasan pengguna aplikasi My Telkomsel tersebut menjadi 2 kelas yaitu positif dan negatif menggunakan naïve bayes sebagai topik penelitian skripsi yang berjudul "**Analisis Sentimen Terhadap Review Aplikasi MyTelkomsel Pada PlayStore Dengan Algoritma Naïve bayes."**

Analisis sentimen adalah teknik untuk memahami seseorang terhadap suatu objek, dengan hasil akhir yang dikategorikan sebagai positif dan negatif. Analisis sentimen juga merupakan proses untuk mengekstraksi data teks dengan tujuan memperoleh informasi yang dilakukan terhadap suatu objek, termasuk komentar positif dan negatif. Pendapat publik dapat digunakan sebagai referensi saat membuat keputusan tentang suatu produk (Nufairi et al, 2024).

Analisis sentimen telah banyak digunakan untuk mengkategorikan review produk. Dengan representasi yang sangat sederhana, naïve bayes, Support Vector Machines, dan K-Nearest Neighbour adalah metode yang sering digunakan dalam analisis sentimen berbasis pembelajaran mesin (Wahyudi et al, 2021).

MyTelkomsel merupakan aplikasi yang diluncurkan oleh Telkomsel yang memungkinkan pelanggan mengelola akun mereka dan mengakses layanan mereka dengan smartphone. Pengguna dapat mengunduhnya melalui Play Store untuk *Android* dan AppStore untuk *IOS*. Pengguna dapat melihat kuota internet mereka, membeli paket internet atau berbagi internet, melihat saldo pulsa mereka, membeli pulsa, dan melakukan pembayaran tagihan, membeli paket bicara, SMS, dan roaming, Hal ini Telkomsel memprioritaskan fungsi aplikasi MyTelkomsel (Devi et al, 2020).

Pengertian *Machine Learning* adalah bidang yang berkaitan dengan membangun algoritma yang bergantung pada kumpulan fenomena tertentu untuk menjadi berguna. Fenomena-fenomena ini dapat berasal dari

alam, diciptakan oleh manusia, atau diciptakan oleh algoritma lain (Luthfiansyah Dan et al, 2024).

a. *Supervised Learning*
   Dalam *Supervised learning*, set pelatihan yang disediakan untuk algoritma mencakup solusi yang diinginkan, yang disebut label.
b. *Unsupervised Learning*
   Data pelatihan tidak diberi label. Sistem mencoba belajar tanpa guru.
c. *Semi-Supervised Learning*
   Karena memberi label pada data biasanya memakan waktu dan biaya yang cukup besar, maka seringkali Anda akan memiliki banyak contoh data yang tidak memiliki label, dan hanya sedikit contoh data yang sudah diberi label.
d. *Self-Supervised Learning*
   Pendekatan lain dalam machine learning melibatkan pembuatan kumpulan data yang sepenuhnya diberi label dari kumpulan data yang sepenuhnya tidak diberi label.

Naive Bayes adalah salah satu metode klasifikasi probabilitas yang paling sederhana yang melakukan dua tahap dalam proses klasifikasi teks tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan, sampel data dipelajari, dan probabilitas untuk setiap kategori didasarkan pada sampel data tersebut. Pada tahap klasifikasi, nilai kategori ditetapkan berdasarkan *terms* yang muncul dalam data yang diklasifikasikan. Sebagai contoh, persamaan teorema Bayes sebagai berikut (Noviana & Rasal B A Jurusan, 2023):

$$P(H|X) = \frac{P(X|H) * P(H)\ A}{P(H)}$$

Keterangan:
X = Data dengan class yang belum diketahui
H = Hipotesis data menggunakan suatu class yang spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (Parteori Probabilitas)
P(H) = Probabilitas hipotesis H (Prior Probabilitas)
P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) = Probabilitas H

Text mining adalah proses menggali informasi di mana seseorang menggunakan alat analisis yang berasal dari data mining untuk berinteraksi dengan sekumpulan dokumen. Contoh alat analisis termasuk kategorisasi teks atau informasi, mengelompokkan teks, dan sebagainya. Sumber data *text mining* berasal dari sekumpulan pola bahasa yang memiliki kemampuan untuk menganalisis data teks yang tidak terstruktur dan semi-terstruktur. Salah satu komponen *text mining* adalah analisis sentimen, yang mempelajari komputasi opini, emosi, dan sentimen orang. Tujuannya adalah untuk mengkategorikan sentimen pada suatu kalimat yang berupa pendapat atau opini, baik itu positif atau negative (Noviana & Rasal B A Jurusan, 2023).

Text Preprocessing adalah tahap awal untuk mempersiapkan teks untuk pengolahan data berikutnya. Sekumpulan karakter yang bersambungan, atau teks, harus dibagi menjadi komponen yang lebih penting, yang dapat dilakukan pada berbagai tingkatan. Beberapa tahapan *preprocessing* teks termasuk (Noviana & Rasal B A Jurusan, 2023).

a.*Case Folding*
   Proses mengubah semua huruf besar (*upper case*) dalam
   dokumen menjadi huruf kecil (*lower case*).
b.*Cleansing*
   Metode pembersihan kata-kata yang tidak diperlukan untuk mengurangi noise selama proses klasifikasi.
c.*Tokenizing*
   Proses pemotongan kalimat menjadi sebuah kata, dengan memisahkan kata dan menentukan struktur dari setiap kata
d.s*temming*
   tahapan untuk mengembalikan suatu katanya atau menghilangkan imbuhan
e. *Stopwords*
   Merupakan kumpulan kata-kata yang sering muncul dalam suatu dokumen, sehingga stopword digunakan untuk menghilangkan kata penghubung yang tidak begitu penting.

Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Contoh *confusion matrix* untuk klasifikasi ditunjukkan pada Tabel 1(Normawati & Prayogi, 2021).

Tabel 1 Confussion Matrix

Fakta	Prediksi	
	positif	negatif
positif	TP	FN
negatif	FP	TN

Keterangan:

TP (*True Positive*) = jumlah dokumen dari kelas positif yang benar diklasifikasikan sebagai kelas positif

TN (*True Negative*) = jumlah dokumen dari kelas negatif yang benar diklasifikasikan sebagai kelas negatif

FP (*False Positive*) = jumlah dokumen dari kelas negatif yang salah diklasifikasikan sebagai kelas positif

FN (*False Negative*) = jumlah dokumen dari kelas positif yang salah diklasifikasikan sebagai kelas negative

Rumus confusion matrix untuk menghitung *accuracy, precision, recall* dan *f1-score* seperti berikut.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \text{x } 100\%$$

$$\text{presisi } (precision) = \frac{\text{TP}}{\text{TP} + \text{FP}} \text{x } 100\%$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{x } 100\%$$

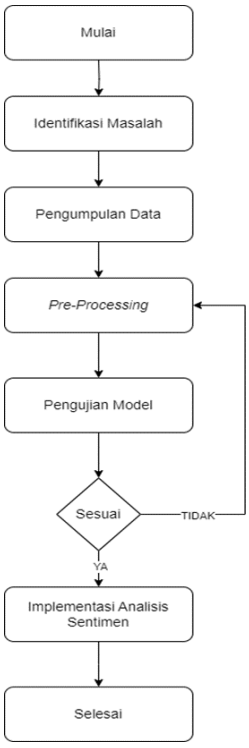
$$\text{F1 } Score = 2 \text{ x } \frac{(recall\ x\ precision)}{(recall + precision)}$$

Python adalah bahasa pemrograman yang paling umum digunakan oleh programmer untuk membuat program. Ini karena karakteristik sintaksnya yang tidak rumit, yang membuatnya menjadi salah satu bahasa pemrograman yang sangat mudah digunakan. Terdapat beberapa aturan yang harus dipenuhi saat menulis kode program menggunakan bahasa pemrograman Python. Hal ini dilakukan untuk mencegah kesalahan atau masalah dengan program yang telah dibuat sebelumnya.Penulisan statement atau perintah adalah aturan sintaks Python yang pertama (Pasek et al, 2022).

Colaboratory, atau "*Colab*" adalah produk yang dibuat oleh Google Research. Kolab berfungsi dengan baik untuk mesin pembelajaran, analisis data, dan pendidikan karena memungkinkan siapa saja menulis dan mengeksekusi kode Python secara otomatis melalui browser. Secara teknis, Colab adalah layanan notebook Jupyter yang dihosting dan dapat digunakan tanpa persiapan.. (Soen et al, 2022).

2. METODE PENELITIAN

Dalam penyusunan penelitian ini, diperlukan beberapa Langkah untuk mencapai tujuan yang telah ditetapkan sebelumnya, Adapun Langkah-langkah penyusunan yaitu



Gambar 1 Kerangka Penelitian

2.1 Identifikasi Masalah

Pada Tahapan ini digunakan untuk menentukan masalah apa yang akan dianalisis, dan hasilnya akan sesuai dengan tujuan penelitian..

2.2 Pengumpulan Data

Pada Tahapan ini digunakan untuk mengumpulkan data komentar terhadap aplikasi MyTelkomsel

1) Pengumpulan Data
   Pada penyusunan penelitian ini, peneliti mengumpulkan data yang dapat mendukung proses dalam penelitian yang berkaitan dengan proses pengumpulan data sebagai berikut:
   a. *Scrapping* Data
      Peneliti melakukan *scrapping* data dengan mengambil data komentar aplikasi MyTelkomsel di Google Play Store. Penarikan data diambil dengan Bahasa pemrograman python, *scrapping* data ini

mendapatkan total data sebanyak 4000 komentar dan akan disimpan dalam formal *Comma Separated Values* (CSV).

Table 2 sample data komentar

No.	Pengguna	Komentar
1.	Pengguna Google	Jaringan internet nya suka *down* sendiri. Dulu telkomsel rajanya jaringan internet sekarang cemen Apknya kebanyakan iklan. Saya *download* cuma pake buat cek kuota aja
2.	Pengguna Google	Tampilan bagus, cuma malah bikin bingung, dan tidak tampil semua di depan, apliaksi jg eror sudah beli kuota tetapi di my telkomsel sudah bbrpa hari masih saja anda tidak memiliki kuota aktif
3.	Pengguna Google	"Telkomsel 8.0 Buat lebih mudah" slogan dan kenyataan gak sesuai setelah update bukannya makin bagus malah makin busuk dimulai dari proses masuk ke aplikasi yang lama hingga pembelian

2.3 Pre-Processing

Pada tahap ini *pre-processing* memiliki tujuan untuk menghilangkan yang tidak diperlukan pada dataset. Tahap ini sama dengan tahap pengolahan data. Tahap pre-processing yang digunakan pada penelitian ini adalah *cleaning, tokenization, case folding, stemming,* dan *stopword removal*. Tahapan ini dilakukan dengan menggunakan aplikasi google colabs.

2.4 Pengujian Model

Setelah proses *pre-processing*, tahap selanjutnya adalah pengujian model yaitu klasifikasi dengan algoritma naïve bayes yang akan menggunakan aplikasi Google colaboratory untuk memudahkan dalam proses klasifikasi. Hasil dari klasifikasi tersebut akan masuk ke tahap perhitungan *confusion matrix*, Dimana tahap ini adalah tahap untuk menghitung keakuratan, presisi, *recall dan F1-score* dari metode yang telah diuji jika sesuai maka ke tahap implementasi analisis sentimen dan jika tidak maka akan Kembali ke tahap pre-processing.

2.5 Implementasi Analisis Sentimen

Hasil dari analisis sentimen. ini akan dianalisis untuk mempermudah dalam interpretasi data, hasil analisis akan berupa dari algoritma yang digunakan yaitu naïve bayes yang akan menjadi kesimpulan akhir penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Hasil penelitian ini adalah untuk mengetahui seberapa akurat perhitungan algoritma naïve bayes untuk ulasan komentar aplikasi myTelkomsel. Data diolah, diurutkan, dan diuji dengan Google Colabs Software. Hasilnya menunjukkan proses perhitungan yang bergantung pada model naïve bayes dan akan diimplementasikan.

3.2. Dataset

Data awal yang akan digunakan untuk analisis sentimen ini adalah data berupa teks yang dikumpulkan oleh penulis dan data teks tersebut berupa komentar aplikasi mytelkomsel Pengumpulan data dari google play store. Data awal yang akan digunakan untuk analisis sentimen ini adalah data berupa teks yang dikumpulkan oleh penulis dan data teks tersebut berupa komentar aplikasi mytelkomsel Pengumpulan data dari google play store dilakukan pada bulan april 2024 - Juli 2024.

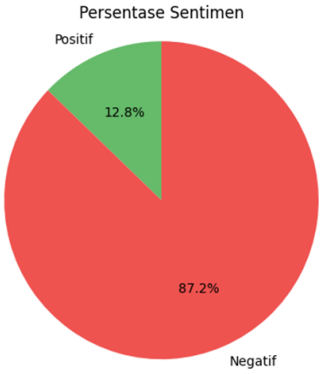
Tabel 3. Jumlah Data komentar

Kelas	Total Data
Positif	455
Negatif	3087

diketahui bahwa dari total komentar yang dipakai, 3087 data termasuk dalam komentar bersentimen negatif dan 455 data termasuk dalam komentar bersentimen positif. Berdasarkan data tersebut, didapatkan bahwa 12.8% opini pengguna aplikasi mytelkomsel pada bulan Januari 2024 – juli 2024 bersentimen positif terhadap aplikasi

mytelkomsel di playstore. Sedangkan, terdapat 87.2% opini pengguna aplikasi mytelkomsel yang bersentimen negatif. Hal ini dapat dilihat pada gambar 2

Gambar 2. Persentase Sentimen



3.3 Pelabelan Sentimen

Pelabelan sentimen pada dataset dilakukan dengan pemrograman python. Pelabel memberi label berupa positif untuk sentimen positif, negatif untuk sentimen negatif dan netral tidak digunakan. Selanjutnya, data yang dipakai hanya yang berlabel positif dan negatif. Beberapa komentar aplikasi mytelkomsel di google playstore yang telah dilabeli ditampilkan dalam Tabel 4 dan Tabel 5.

Tabel 4. Sentimen Negatif

content	score	Label
Buka aplikasi lain lancar jaya tmpa kendala. Begitu buka my telkomsel, loading nya setngah mati gk kebuka" Jga menu nya.	1	Negatif

Tabel 5. Sentimen Positif

content	score	label
Aplikasinya Oke tampilan menu utama bagus dan simple tampilan UI/UX cukup sederhana dan juga menilai pribadi aku suka simpel dan juga gampang	4	Positif

3.4 Data Preprocessing

Data *preprocessing* adalah proses yang penting dalam *sentiment analysis*, Yaitu proses untuk membersihkan data yang telah dikumpulkan (Surya Sayogo et al., 2023) Adapun tahap-tahap *preprocessing* yang akan digunakan pada kasus ini adalah sebagai berikut

1). *Cleaning*
2). *Case Folding*
3). *Stopword Removal*
4). *Tokenizing*
5). *Stemming*

Sebelum mulai pada tahap *preprocessing* terdapat beberapa *library* yang perlu di import, yang mana *library* ini akan digunakan pada tahap-tahap *preprocessing*. Adapun *library* yang akan digunakan ditampilkan pada kode program dibawah ini

Gambar 3. Kode program library

```
# Import library:
!pip install google-play-scraper
!pip install Sastrawi

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

from google_play_scraper import app
from wordcloud import WordCloud
import nltk
import re
import string
from nltk.corpus import stopwords
nltk.download('punkt')
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer


stop_words = stopwords.words()
```

Berikut penjelasan mengenai Gambar 3 yang yang di butuhkan diantaranya :

1) *Library google-play-scraper* berfungsi sebagai mengambil data ulasan instagram di google play store.
2) *Library pandas* digunakan untuk pengolahan data.
3) *Library numpy* berfungsi untuk membuat array dari struktur data.
4) *Library re(RegEx)* digunakan untuk mencocokkan,mencari, mengganti, dan memanipulasi teks berdasarkan pola tertentu.
5) *Library nltk* berfungsi untuk pengolahan data Bahasa manusia.
6) *Library sastrawi* diubakan pengurang kata-kata yang ter-infleksi dalam bahasa Indonesia ke bentuk baku
7) *Library sklearn* membantu melakukan procesing data ataupun melakukan training data untuk kebutuhan machine learning atau data science.

Tahap selanjutnya adalah preprocessing data, yaitu rangkaian proses mengubah data sesuai dengan format. Persiapan data ini menggunakan fitur text preprocessing. Tahapan data preprocessing yang dilakukan di jelaskan sebagai berikut.

a. *Cleaning*

Tahapan ini untuk menghilangkan atribut yang ada pada data seperti simbol, tanda baca, hashtag, URL, angka dan lainnya. Proses cleaning ini ditampilkan dalam Tabel 6.

Tabel 6. data hasil cleaning

Input	keren | Reply: sama sama kak Alma! Aku ga nyangka dibales ssma admin Telkomsel whehee makasih ya kak udah notice rating dariku sehat dan sukses selalu ya kak :3
Output	keren reply sama sama kak alma aku ga nyangka dibales ssma admin telkomsel whehee makasih ya kak udah notice rating dariku sehat dan sukses selalu ya kak

b. *Case Folding*

Tahap ini merupakan proses mengubah semua data huruf kapital yang ada pada dataset menjadi huruf kecil (nonkapital). contoh data yang telah diubah menjadi huruf kecil/lowercase ditampilkan pada Tabel 7.

Tabel 7. data case folding

Input	JANGAN UPDATE VERSI TERBARU DULU. SEDANG ADA MASALAH! Informasi data kuota telepon dan internet tidak akurat. Setelah update Mei 2024, opsi penambahan nomor menjadi sulit, Tidak bisa lagi menggunakan email, Padahal layanan Orbit dan Indihome bisa. TOLONG DIPERBAIKI SEGERA! Ini Sangat mengganggu.
Output	jangan update versi terbaru dulu sedang ada masalah informasi data kuota telepon dan internet tidak akurat setelah update mei opsi penambahan nomor menjadi sulit tidak bisa lagi menggunakan email padahal layanan orbit dan indihome bisa tolong diperbaiki segera ini sangat mengganggu

c. *Stopword Removal*

Tahap ini untuk menghilangkan kata-kata yang ada dalam daftar kata *stopword*, dalam kasus ini, peneliti menggunakan kata-kata yang ada dalam *library NLTK (Natural Language Toolkit)* dengan dataset bahasa Indonesia(Nurtikasari et al., 2022).dalam suatu daftar *stopword*, apabila ada kata masuk di dalam *stopword* maka kata tersebut tidak akan diproses lebih lanjut. *Stopword* juga dapat digunakan untuk menghilangkan kata yang tidak bermakna dalam dataset. Daftar beberapa kata *stopword* yang digunakan dalam penelitian ini dapat dilihat pada Tabel 8.

Tabel 8. data hasil cleaning

Input	Jelek banget, maaf ya . Karena mau beli paket data, sudah tidak ada tertera lagi . Jadi bingung mau beli paket bagaimana. Tolong diperbaikin dong, makin aneh ini aplikasi.
Output	jelek banget maaf ya beli paket data tertera bingung beli paket tolong diperbaikin aneh aplikasi

d. *Tokenizing*

Tahapan ini merupakan proses memecahkan kalimat menjadi satuan kata. Metode ini akan memotong kata berdasarkan spasi yang selanjutnya kata tersebut akan menjadi token(Surya Sayogo et al., 2023). menghasilkan token kata seperti yang ditampilkan dalam Tabel 9.

Tabel 9. data tokenizing

update versi terbaru informasi data kuota telepon internet akurat update mei opsi penambahan nomor sulit email layanan orbit indihome tolong diperbaiki mengganggu pelanggan
[update,versi,terbaru,informasi,data,kuota,telepon,internet,akurat,update,mei,opsi,penambahan,nomor,sulit,email,layanan,orbit,indihome,tolong,diperbaiki,mengganggu, pelanggan]

e. *Stemming*

Stemming merupakan tahapan untuk mengembalikan suatu katanya atau menghilangkan imbuhan (awalan dan akhiran) yang terdapat pada kata tersebut (Tirta Nugraha et al, 2023).Proses stemming ini akan menghasilkan hasil stemming seperti yang ditampilkan dalam Tabel 10.

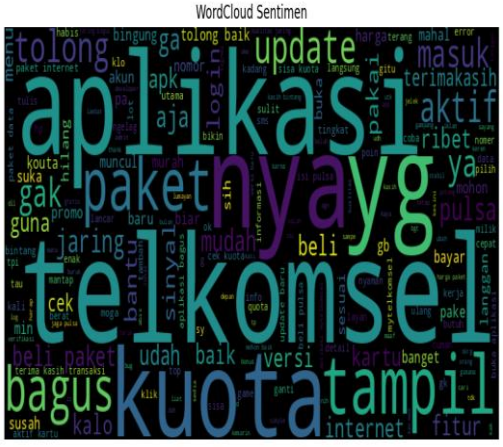
Tabel 10. data stemming

Input	Output
dilepas	lepas
bersalah	salah
dijamin	jamin
lambatnya	lambat
berbahaya	bahaya
terhambat	hambat

3.5 Wordcloud

Dataset yang sudah melewati beberapa proses *preprocessing* data, dicari frekuensi kemunculan kata yang paling sering pada tiap sentimen yang divisualisasikan dalam bentuk *wordcloud*. *Wordcloud* ini merupakan suatu gambar yang terdiri dari sekumpulan Pembuatan wordcloud ini menggunakan *library* dari *wordcloud*.

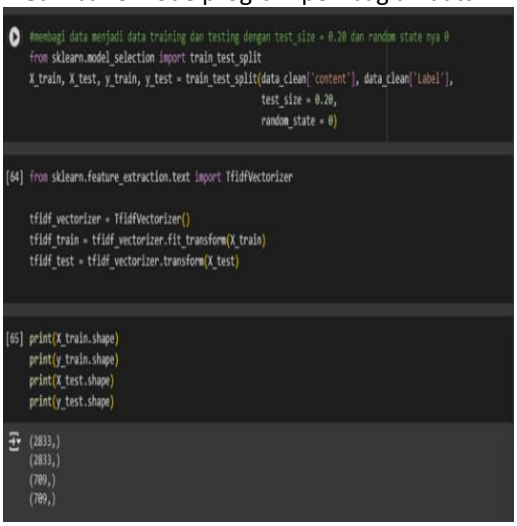
Gambar 4. wordcloud sentimen



3.6 Pembagian Data

Pembagian data dalam penelitian ini dilakukan secara acak agar pembagian dataset menjadi data latih dan data uji yang dilakukan dapat menghasilkan hasil yang optimal dengan pembagian 80% data latih dan 20% data uji. Proses pembagian data ini menggunakan *library Sklearn* untuk membagi data. Untuk menjalankan proses ini dilakukan dengan kode program pada gambar 5.

Gambar 5. kode program pembagian data

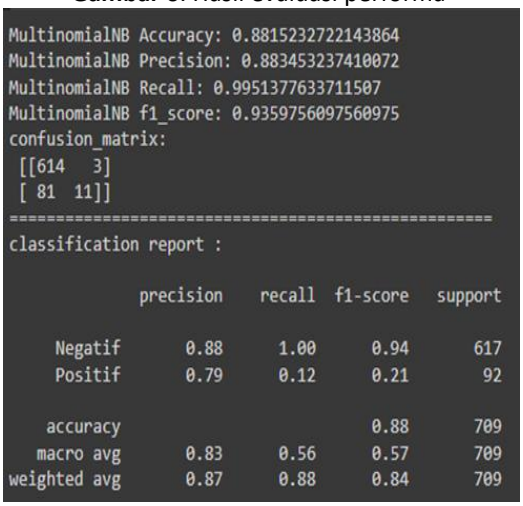


```
#membagi data menjadi data training dan testing dengan test_size = 0.20 dan random state nya 0
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_clean['content'], data_clean['Label'],
                                                    test_size = 0.20,
                                                    random_state = 0)

[64] 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)

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

    (2833,)
    (2833,)
    (709,)
    (709,)
```

3.7 Evaluasi Naïve Bayes

Evaluasi performa model menggunakan NaïveBayes beberapa metrik yangmemberikan wawasan tentang seberapa baik model dapat melakukan klasifikasi pada data. Akurasi mengukur persentase keseluruhan prediksi yang benar,

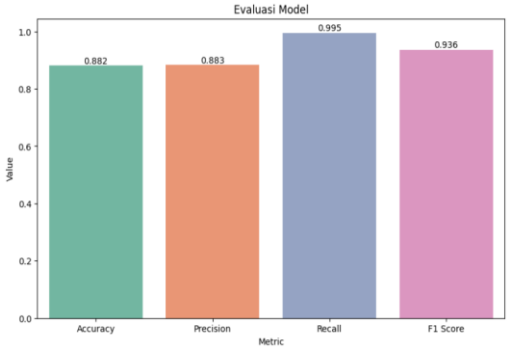
Gambar 6. Hasil evaluasi performa



```
MultinomialNB Accuracy: 0.8815232722143864
MultinomialNB Precision: 0.883453237410072
MultinomialNB Recall: 0.9951377633711507
MultinomialNB f1_score: 0.9359756097560975
confusion_matrix:
 [[614   3]
 [ 81  11]]
================================================
classification report :

              precision    recall  f1-score   support

     Negatif       0.88      1.00      0.94       617
     Positif       0.79      0.12      0.21        92

    accuracy                           0.88       709
   macro avg       0.83      0.56      0.57       709
weighted avg       0.87      0.88      0.84       709
```

Dalam pengujian menggunakan model naïve bayes, data di pecah menjadi dua subset, yaitu 80% data latih untuk melatih model dan 20% data uji untuk mengukur kinerja. Hasil pengujian ditampilkan dalam gambar 6 yang mencakup metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1 Score*, yang memberikan gambaran menyeluruh tentang kemampuan model dalam melakukan klasifikasi pada dataset.

Gambar 7. Hasil evaluasi model



3.8 Confussion Matrix

Confusion Matrix pada setiap skema dengan metode Naïve bayes. Sebanyak 614 data diprediksi secara benar sebagai data positif (*True Positive*). Sebanyak 11 data diprediksi secara benar sebagai data negatif (*True Negative*). Sebanyak 3 data diprediksi secara salah sebagai data negatif, seharusnya data tersebut adalah data positif (*False Negative*). Sebanyak 81 data diprediksi secara salah sebagai data positif, seharusnya data tersebut adalah data negatif (*False Positive*).

Tabel 10. confussion matrix

Fakta	Prediksi	
	positif	negatif
Positif	614	3
negatif	81	11

a. Akurasi

$$\text{Akurasi } \frac{614 + 11}{614 + 11 + 81 + 3} \text{x } 100\% = 88.15\%$$

b. Presisi

$$\text{presisi} = \frac{614}{614 + 81} \text{x } 100\% = 88.34\%$$

c. Recall

$$Recall = \frac{614}{614 + 3} \text{x } 100\% = 99.51\%$$

d. F1-score

$$F1\ Score = 2 \text{ x } \frac{(99.51 \text{ x } 88.34)}{(99.51 + 88.34)} = 93.59\%$$

4. KESIMPULAN

Berdasarkan rumusan masalah yang dibuat di awal bab, sebagai berikut adalah kesimpulan yang dapat diambil dari penelitian ini algorima Naïve bayes dapat meningkatkan akurasi dan memberikan solusi terhadap permasalahan klasifikasi review agar lebih akurat, akurasi yang dihasilkan dari pengujian data ini yaitu Hasil pengujian data untuk algoritma naïve bayes akurasinya adalah 88%, *recall* 99%, *precision* 88%, dan *f1-score* 93%

Beberapa saran yang dapat diterapkan untuk perbaikan dan pengembangan penelitian adalah sebagai berikut. Analisis sentimen ulasan aplikasi mytelkomsel, selain menggunakan algoritma Naïve Bayes Classifier, dapat dipertimbangkan metode klasifikasi lain seperti Support Vector Machine, Random Forest, K-Nearest Neighbor, dan Decision Tree. Sehingga hasil pengujian model dapat ditingkatkan dan menghasilkan akurasi yang lebih tinggi. Penelitian selanjutnya dapat melakukan komparasi atau perbandingan dengan menggunaan algoritma Support Vector Machine (SVM) yaitu metode yang cukup akurat untuk mengklasifikasikan komentar positif dan negatif dari aplikasi mytelkomsel sehingga dapat menghasilkan data yang lebih valid dan akurat.

5. REFERENSI

Devi, P. C., Hanafi, A., & Wardhana, A. C. (2020). Evaluasi Aplikasi My Telkomsel Menggunakan Metode Usability Testing. *Jurnal Jaring SainTek*, *5*(1), 29–38. http://ejurnal.ubharajaya.ac.id/index.php/jaring-saintek29

Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, *12*(2), 154–161. https://doi.org/10.33096/ilkom.v12i2.597.154-161

Indra Cahyani, T., Gata, W., Dwi Saputra, D., Bella Novitasari, H., & Nusa Mandiri, U. (2023). ANALISIS SENTIMEN TERHADAP TELKOMSEL DAN XL BERBASIS MACHINE LEARNING PADA DATA TWITTER SENTIMENT ANALYSIS TOWARD ISP TELKOMSEL AND XL ON TWITTER USING MACHUNE LEARNING CLASSIFICATION. *Journal of Information Technology and Computer Science (INTECOMS)*, *6*(1).

Luthfiansyah Dan, R., Wasito, B., Program, S., & Sistem, I. (2024). *ANALISIS SENTIMEN TERHADAP PARA KANDIDAT PRESIDEN 2024 BERDASARKAN NETIZEN PENGGUNA TWITTER DENGAN METODE DATA MINING DAN TEXT MINING*.

Maulana, M. P., Eka, &, Sari, P., & Kom, M. (2023). Analisa Kepuasan Pengguna Terhadap Aplikasi My Telkomsel Dengan Menerapkan Metode TAM (Technology Acceptance Model). In *Syntax: Jurnal Informatika* (Vol. 12, Issue 02).

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI* (Vol. 5, Issue 2).

Noviana, R., & Rasal B A Jurusan, I. (2023). PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER. *JTS*, *2*(2).

Nufairi, F., Pratiwi, N., & Herlando, F. (2024). ANALISIS SENTIMEN PADA ULASAN APLIKASI THREADS DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *9*(1), 339–348. https://doi.org/10.29100/jipi.v9i1.4929

Nurtikasari, Y., Syariful Alam, & Teguh Iman Hermanto. (2022). Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes. *INSOLOGI: Jurnal Sains Dan Teknologi*, *1*(4), 411–423. https://doi.org/10.55123/insologi.v1i4.770

Pasek, P., Mahawardana, O., Sasmita, G. A., Agus, P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python. In *JITTER-Jurnal Ilmiah Teknologi dan Komputer* (Vol. 3, Issue 1).

Soen, G. I. E., Marlina, M., & Renny, R. (2022). Implementasi Cloud Computing dengan Google Colaboratory pada Aplikasi Pengolah Data Zoom Participants. *JITU : Journal Informatic Technology And Communication*, *6*(1), 24–30. https://doi.org/10.36596/jitu.v6i1.781

Surya Sayogo, D., Irawan, B., & Bahtiar, A. (2023). ANALISIS SENTIMEN ULASAN INSTAGRAM DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).

Tirta Nugraha, M., Sulistiyowati, N., Enri Informatika, U., Singaperbangsa Karawang Jl HSRonggo Waluyo, U., & Timur, T. (2023). ANALISIS SENTIMEN ULASAN APLIKASI SATU SEHAT PADA GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 5).

Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, *8*(2). http://ejournal.bsi.ac.id/ejurnal/index.php/ji