

ANALISIS SENTIMEN APLIKASI GOTUBE PADA GOOGLE PLAYSTORE MENGUNAKAN ALGORITMA NAÏVE BAYES

Imam Safaat¹, Evanita², Arief Susanto³

^{1,2,3} Teknik Informatika, Universitas Muria Kudus
Email: ¹201951208@std.umk.ac.id, ²evanita@umk.ac.id, ³ariefpjl@gmail.com

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

GoTube merupakan aplikasi pemutar video yang paling banyak dicari sebagai alternatif pengganti youtube premium, karena GoTube menyuguhkan berbagai fitur yang dimiliki Youtube premium dan dapat didownload secara gratis melalui playstore dan appstore. Terhitung sampai november 2022 ini aplikasi GoTube sudah mencapai 10jt lebih unduhan, dan 230rb lebih ulasan dengan rating 4,5 di Playstore. Meskipun menduduki peringkat teratas dan rating yang sangat bagus, tentu tidak semua memberikan komentar positif, ada yang memberikan komentar negatif. Untuk menganalisis hal tersebut maka diperlukan suatu metode dan analisis untuk mengklasifikasikan komentar pengguna ke dalam beberapa kategori, dimana dalam penelitian ini berupa kategori positif dan negatif. Penelitian ini menggunakan metode Naive Bayes untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi GoTube di Playstore. Tahapan yang dilakukan adalah pengambilan data ulasan pada Google Play, pelabelan untuk mendapatkan data ulasan positif dan negatif, data preprocessing, pembobotan TF-IDF, klasifikasi menggunakan algoritma Naïve Bayes, dan evaluasi. Hasil analisis sentimen aplikasi Gotube pada Playstore menghasilkan nilai accuracy 80,89%, precision 0,7083, recall 0,6296, dan f1-score 0,6666. Mengindikasikan sentimen dari pengguna Gotube di Playstore cenderung positif.

Kata kunci: *Naïve Bayes, preprocessing, TF-IDF, GoTube.*

SENTIMENT ANALYSIS OF GOTUBE APPLICATION ON GOOGLE PLAYSTORE USING NAIVE BAYES ALGORITHM

Abstract

GoTube is the most sought-after video player application as an alternative to premium YouTube, because GoTube presents various features owned by premium Youtube and can be downloaded for free through the playstore and appstore. As of November 2022, the GoTube application has reached more than 10 million downloads, and more than 230k reviews with a rating of 4.5 on Playstore. Although it was ranked at the top and the rating was very good, of course not all gave positive comments, some gave negative comments. To research and analyze this, a method and analysis are needed to classify user comments into several categories, where in this study in the form of positive and negative categories. This research uses the Naive Bayes method to generate positive and negative sentiments towards GoTube application user comments on Playstore. The stages carried out are retrieval of review data on Google Play, labeling to obtain positive and negative review data, preprocessing data, TF-IDF weighting, classification using the Naïve Bayes algorithm, and evaluation. The sentiment analysis results of the Gotube app on Playstore yielded an accuracy score of 80.89%, precision of 0.7083, recall of 0.6296, and an f1-score of 0.6666. Indicating that user reviews of Gotube on Playstore tend to be positive.

Keywords: *Naïve Bayes, preprocessing, TF-IDF, GoTube*

1. PENDAHULUAN

Meninjau ulasan suatu produk yang tersebar di berbagai media menjadi hal yang penting. Berdasarkan penelitian pada awal 2013, ditemukan bahwa ulasan produk yang tersebar di media sosial memiliki pengaruh besar terhadap 90% keputusan

seseorang dalam membeli produk tersebut. Khusus untuk aplikasi, ulasan terhadap aplikasi tersebut dapat ditemukan di Google Play Store. Di halaman Google Play Store, ulasan diberikan dalam bentuk rating bintang dari satu hingga lima disertai dengan ulasan teks.

Aplikasi dengan rating bintang tinggi akan direkomendasikan oleh Google dan dapat muncul di halaman depan atau dalam daftar aplikasi terbaik jika diurutkan berdasarkan rating. Calon pengguna umumnya tertarik untuk mengetahui pendapat atau pengalaman pengguna lain terkait aplikasi yang akan mereka gunakan, dan biasanya mereka mencari aplikasi berdasarkan rating yang tinggi.

Namun, ada beberapa masalah dalam ulasan di Google Play Store. Terkadang, seorang pengguna memberikan rating yang rendah seperti satu atau dua bintang, namun memberikan ulasan teks yang sebenarnya positif. Di sisi lain, ada juga pengguna yang memberikan rating tinggi tetapi memberikan ulasan teks yang negatif. Hal ini mengakibatkan kesulitan bagi Google untuk membedakan antara ulasan positif dan negatif hanya melalui teks ulasan yang diberikan pengguna. Ketidaksesuaian ini dapat mempengaruhi rating sebenarnya dari aplikasi tersebut. Setelah Vanced resmi ditutup, ada banyak di antara kita yang mencari alternatif lain supaya bisa menonton YouTube dengan lebih menyenangkan hingga akhirnya menjatuhkan pilihan ke GoTube. GoTube sendiri memang tergolong sangat menarik. Selain membawa beberapa fitur unggulan di mana sebagian mirip dengan yang ditawarkan oleh Vanced, aplikasi ini juga tersedia secara resmi di halaman Google Play Store.

GoTube adalah aplikasi pemutar video yang menawarkan dukungan penayangan pada jendela melayang (pop-up) atau lebih dikenal dengan PIP (Picture-in-picture) di ekosistem Android secara umum. Aplikasi pemutar video buatan GoTube Studio ini tersedia secara gratis dengan opsi menayangkan iklan di aplikasi sebagai gantinya. Aplikasi ini pun sudah diunduh lebih dari 10 juta kali dengan rating dan ulasan yang terlihat sangat baik di Google play store. Menariknya lagi GoTube dirancang mirip sekali dengan Aplikasi YouTube dan mempunyai beberapa fitur tambahan yang sangat menarik. Fitur unggulan GoTube diantaranya adalah fitur pemblokir iklan di dalam video maupun dukungan minimized background. Meski menawarkan kemudahan dengan segala fitur unggulannya, tentunya tidak semua pengguna merasa puas dan terlayani dengan baik, oleh karena itu ada kritik yang dibutuhkan oleh pengembang untuk menjaga kualitas dan perkembangan layanan GoTube.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Sebuah penelitian terkait telah dilakukan oleh (Artanti Inez Tanggraeni & Melkior N. N, 2022), yang berjudul "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes". Penelitian tersebut

menggunakan data ulasan dari Google Play Store dan menerapkan algoritma Naïve Bayes sebagai metode klasifikasi. Hasil dari penelitian tersebut menunjukkan bahwa dengan penerapan algoritma Naïve Bayes untuk mengklasifikasikan ulasan ke dalam label positif atau negatif dari total 577 data latih dan 65 data uji, diperoleh akurasi sebesar 89%, presisi sebesar 83%, dan recall sebesar 87%.

Sebuah penelitian yang dilakukan oleh (D. Prاتمanto, 2020), berjudul "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm." Penelitian tersebut menggunakan Algoritma Naïve Bayes sebagai solusi untuk memecahkan masalah pengklasifikasian sentimen pada ulasan aplikasi Shopee, dan menghasilkan tingkat akurasi yang tinggi. Dalam penelitian tersebut, ditemukan bahwa algoritma Naïve Bayes mencapai akurasi sebesar 96,667%.

Penelitian yang dilakukan oleh (Edyt Daryfayi Putra Daulay & Ibnu Asror, 2020), berjudul "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes." Ulasan, atau yang sering disebut review, merupakan salah satu fitur penting dalam aplikasi Google Play Store. Fitur ini memungkinkan para pengguna untuk memberikan penilaian dan memberikan pendapat berupa teks ulasan terhadap aplikasi yang mereka gunakan. Namun, dalam beberapa kasus, terdapat teks ulasan yang tidak sesuai dengan nilai atau rating yang diberikan. Misalnya, pengguna memberikan rating bintang 5 namun memberikan teks ulasan yang bersifat negatif. Dalam penelitian ini, dibangun sebuah mesin klasifikasi khusus untuk membandingkan teks ulasan yang diberikan oleh pengguna dengan rating yang diberikan. Metode yang digunakan adalah Naïve Bayes, khususnya Multinomial Naïve Bayes, karena metode ini lebih cocok untuk klasifikasi teks. Dari hasil penelitian, diperoleh akurasi sebesar 78,9% untuk ulasan dengan rating bintang 5 dan bintang 1. Namun, akurasi menurun menjadi 73,7% untuk semua rating ulasan kecuali bintang 3.

Penelitian yang dilakukan oleh (Rita Apriani & Dudih Gustian, 2019), berjudul "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia." Tokopedia adalah salah satu toko online yang paling banyak dikunjungi oleh masyarakat Indonesia, dengan total pengunjung mencapai 1,2 miliar, terdiri dari 863,1 juta pengunjung dari web mobile dan 329,8 juta pengunjung dari desktop. Meskipun Tokopedia menduduki peringkat teratas dengan rating 5 mendapatkan nilai terbanyak, namun tidak semua pengguna memberikan komentar yang positif. Beberapa pengguna memberikan komentar negatif karena masalah kepercayaan menjadi faktor penting bagi kelangsungan toko online. Terkadang produk

yang dijual atau dibeli melalui situs e-commerce seperti Tokopedia tidak sesuai dengan foto di iklan atau tidak sesuai dengan apa yang diinginkan. Untuk meneliti dan menganalisis hal tersebut, penelitian ini menggunakan metode dan analisis untuk mengklasifikasikan komentar pengguna ke dalam beberapa kategori, yaitu kategori positif dan negatif. Metode Naive Bayes digunakan dalam penelitian ini untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi Tokopedia di Playstore. Pengujian dilakukan berdasarkan nilai class negative, class positive, recall, dan accuracy dalam analisis sentimen. Hasilnya menunjukkan bahwa akurasi performance mencapai 97,13%, dengan nilai precision sebesar 1. Sementara pada Class Recall, diperoleh nilai 95,49% (positive class: negative). Selain itu, nilai AUC yang diperoleh sebesar 0,980.

Penelitian yang dilakukan oleh (Adhitia Erfina & Muhamad Fani Al-shufi, 2022), berjudul "Analisa Sentimen Aplikasi Jasa Kurir Di Play Store Menggunakan Algoritma Naïve Bayes." Dengan meningkatnya penjualan melalui marketplace, industri jasa kurir juga mengalami peningkatan. Terdapat banyak pilihan jasa kurir yang dapat dipilih, dan masing-masing jasa kurir memiliki perbedaan kualitas tergantung pada layanan yang diberikan kepada pelanggan. Untuk mengetahui kualitas jasa kurir, penelitian ini dilakukan pada jasa kurir aplikasi di Play Store. Berdasarkan penelitian yang telah dilakukan, diperoleh hasil sebagai berikut: Aplikasi JNT memiliki nilai akurasi tertinggi sebesar 100%, namun hanya memiliki 50 review dengan sentimen positif. Aplikasi JNE memiliki nilai akurasi 98% dengan 46 review dengan sentimen positif. Aplikasi ninjapres memiliki nilai akurasi 97,87% dengan 49 ulasan dengan sentimen positif. Aplikasi sicepat menduduki urutan keempat dengan nilai akurasi 97,85% dan 47 review dengan sentimen positif. Aplikasi idexpress memiliki nilai akurasi 94% dengan 49 review dengan sentimen positif. Dari hasil penelitian tersebut, dapat disimpulkan bahwa aplikasi JNT memiliki nilai akurasi tertinggi, namun jumlah review dengan sentimen positifnya lebih sedikit dibandingkan dengan aplikasi JNE yang memiliki akurasi yang hampir sama tingginya.

Penelitian yang dilakukan oleh (Nuraeni Herlinawati, Yuri Yuliani, Siti Faizah, Windu Gata, dan Samudi, 2020), berjudul "Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine." Aplikasi Zoom Cloud Meetings menjadi sangat populer di era pandemi virus corona, karena mengharuskan semua kegiatan dilakukan secara virtual. Zoom Cloud Meetings merupakan aplikasi yang menawarkan berbagai fitur, termasuk konferensi video dan audio. Dalam penelitian ini,

para penulis menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) untuk menganalisis sentimen positif atau negatif dari ulasan pengguna aplikasi Zoom di Google Play Store. Setelah proses preprocessing, jumlah dataset mencapai 1.007 catatan, dan data hampir seimbang dengan 546 ulasan berlabel positif dan 461 ulasan berlabel negatif. Evaluasi model dilakukan dengan menggunakan metode 10-fold cross-validation, yang menghasilkan nilai akurasi dan nilai AUC untuk masing-masing algoritma. Penelitian ini bertujuan untuk memahami bagaimana sentimen pengguna aplikasi Zoom Cloud Meetings di Google Play Store, serta membandingkan performa metode Naïve Bayes dan Support Vector Machine dalam melakukan analisis sentimen tersebut. Hasil evaluasi akan memberikan informasi tentang efektivitas kedua metode dalam mengklasifikasikan ulasan menjadi sentimen positif atau negatif. Untuk metode Naïve Bayes, diperoleh nilai akurasi sebesar 74,37% dan nilai AUC sebesar 0,659. Sedangkan untuk metode Support Vector Machine (SVM), diperoleh nilai akurasi sebesar 81,22% dan nilai AUC sebesar 0,886. Dalam penelitian ini, dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) memberikan tingkat akurasi yang lebih tinggi sebesar 6,85% dibandingkan dengan algoritma Naïve Bayes (NB).

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis Sentimen adalah suatu proses otomatis untuk mengekstraksi, mengolah, dan memahami data teks yang biasanya berupa opini atau pandangan publik. Hasil dari analisis sentimen ini berupa kesimpulan mengenai persentase sentimen positif, negatif, atau netral dari data yang dianalisis menggambarkan analisis sentimen sebagai suatu metode yang digunakan untuk mengolah data tekstual guna mengidentifikasi dan mengevaluasi sentimen yang terkandung di dalamnya. (Indrayana et al., 2021).

2.2.2 Natural Language Processing

NLP merupakan salah satu cabang dari kecerdasan buatan yang difokuskan pada pemrosesan bahasa alami. Dengan menggunakan NLP, kita dapat mengidentifikasi sentimen teks, seperti apakah teks tersebut bersifat positif, negatif, atau netral. Referensi yang menyebutkan tentang pentingnya penggunaan NLP dalam analisis sentimen dan penerapannya dalam pengolahan teks ialah. (Parlika et al., 2020)

2.2.3 Text Mining

Anda benar, Text mining atau penambangan teks adalah teknik yang digunakan untuk mengatasi berbagai tugas seperti klasifikasi, pengelompokan (clustering), ekstraksi informasi (information extraction), dan pengambilan informasi

(information retrieval) dari data teks. Teknik ini fokus pada analisis teks yang berasal dari bahasa alami yang tidak terstruktur, seperti artikel, berita, dokumen, tweet, dan lain sebagainya.

Sementara itu, data mining adalah proses penggalian pola atau informasi yang berguna dari data yang terstruktur, biasanya disimpan dalam basis data atau tabel dengan format yang telah ditentukan. Data mining bertujuan untuk menemukan pola tersembunyi, hubungan, atau wawasan yang dapat membantu pengambilan keputusan atau prediksi masa depan. Jadi, perbedaan utama antara text mining dan data mining adalah pada sumber data yang digunakan dan jenis analisis yang dilakukan. Text mining berfokus pada analisis data teks dari bahasa alami yang tidak terstruktur, sedangkan data mining bekerja dengan data yang terstruktur dan berbentuk tabel dalam basis data. (Agustina et al, 2021).

2.2.3 Scraping

Scraping adalah sebuah proses ekstraksi informasi (teks) dari situs web dan halaman online. Ini merupakan suatu metode yang digunakan untuk mengambil informasi dan dataset guna dilakukan analisis lebih lanjut. (Widodo et al, 2021).

2.2.3 Preprocessing

Proses preprocessing dokumen dilakukan untuk menyiapkan dokumen teks mentah sehingga siap untuk dianalisis sesuai dengan metode yang digunakan. Tujuan dari proses ini adalah untuk meningkatkan akurasi klasifikasi ulasan perangkat lunak. (Astari, Divayana and Indrawan, 2020).

Ada beberapa tahapan dalam teknik preprocessing yaitu:

a. Cleansing

Cleansing adalah proses menghapus karakter non-alfabetis guna mengurangi noise dalam teks. Karakter-karakter yang dihapus termasuk tanda baca, simbol-simbol seperti tanda '@' untuk nama pengguna, hashtag (#), emotikon, dan URL dari situs web. (Tuhuteru and Iriani, 2018).

b. Case Folding

Dalam sebuah tweet, Case Folding dilakukan untuk mengatasi banyak variasi penggunaan bentuk huruf. Pada tahap ini, seluruh huruf kapital (uppercase) diubah menjadi huruf kecil (lowercase) agar formatnya seragam. (Kurniawan and Susanto, 2019).

c. Tokenizing

Pada tahapan tokenisasi, semua tanda baca dan angka yang ada dalam data ulasan positif dan negatif dihilangkan. (Hayuningtyas and Sari, 2019).

d. Stopword

Kata-kata atau konjungsi yang tidak relevan, seperti "tetapi", "dengan", "untuk", "yang", dan

kata sambung lainnya akan dihapus dari teks. (Laurensz and Eko Sedyono, 2021).

e. Stemming

Stemming adalah proses mengubah kata-kata dalam teks menjadi bentuk kata dasar sesuai dengan kaidah bahasa Indonesia yang benar. Metode yang digunakan dalam Stemming adalah Stemming Nazief dan Adriani. (Yulita, Nugroho and Algifari, 2021).

2.2.4 Metode Naïve Bayes

Algoritma Naïve Bayes Classifier merupakan algoritma yang dapat digunakan untuk mencari nilai probabilitas tertinggi pada klasifikasi data uji dikategori yang paling tepat. Naïve Bayes Classifier merupakan salah satu dari metode machine learning yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh Bayes sendiri adalah Teorema Bayes, yaitu melakukan klasifikasi dengan melakukan perhitungan nilai probabilitas (Tri Romadloni et al., 2019). Seperti pada penelitian yang dilakukan oleh (M Rizki, M Arhami, H Huzeni et al., 2021) algoritma yang digunakan untuk klasifikasi Naïve Bayes memiliki persamaan umum sebagai berikut:

$$P(Y|X) = (P(X|Y)P(Y))/P(X)$$

Keterangan:

X : Data dengan class yang belum diketahui

Y : Hipotesis data suatu class spesifik

(Y|X) : Probabilitas hipotesis berdasarkan kondisi (posteriori probability)

(Y) : Probabilitas hipotesis

(X|Y) : Probabilitas sebelumnya dari kelas kondisi hipotesis

(X) : Probabilitas Y

2.2.5 Google colab

Google Colab merupakan salah satu produk yang ditawarkan oleh Google, yang memungkinkan pengguna untuk menulis dan mengedit program secara online. Meskipun produk ini masih lebih dikenal di kalangan pengembang atau programmer, Google Colab juga dapat diakses oleh orang-orang yang tertarik untuk belajar pemrograman tanpa perlu menginstal perangkat lunak secara lokal di komputer. (Aorinka Anendya, 2023).

2.2.6 GoTube

GoTube merupakan sebuah aplikasi pemutar video yang menyediakan fitur penayangan dalam jendela melayang atau lebih dikenal sebagai PIP (Picture-in-picture) pada perangkat Android secara keseluruhan. (Sukindar, 2022).

3. METODE PENELITIAN

Tahapan yang harus ditempuh pada penelitian ini berjumlah 6 tahap, yaitu pengambilan data, pelabelan, data preprocessing, pembobotan TF-

IDF, pemisahan data dengan tujuan untuk membuat data latih dan data uji, klasifikasi menggunakan algoritma Naïve Bayes, dan evaluasi.

3.1 Pengumpulan Data

Dalam penelitian ini, informasi mengenai ulasan aplikasi GoTube diambil dari website Google Play Store. Pengambilan data ulasan aplikasi GoTube dilakukan menggunakan teknik web scraping. Data yang diambil berasal dari halaman website yang menggunakan bahasa markup. Proses web scraping berjalan secara semi-terstruktur. Hasil dari proses web scraping disimpan dalam format file CSV. Fokus pengambilan data adalah tanggal, rating, dan isi ulasan dari aplikasi GoTube. Rentang waktu data yang diambil adalah dari bulan Oktober 2022 hingga Desember 2022. Total data ulasan yang berhasil diambil sebanyak 500 ulasan.

3.2. Metode Pengembangan Sistem

Metode Pemrosesan Bahasa Alami (Natural Language Processing - NLP) ketika Anda membuat program klasifikasi dengan Google Colab. NLP adalah cabang dari kecerdasan buatan yang berfokus pada pemahaman, interpretasi, dan analisis bahasa manusia oleh komputer. Dalam konteks program klasifikasi, metode NLP digunakan untuk mengolah teks yang ada dalam dataset dan melakukan tugas klasifikasi teks, seperti analisis sentimen atau klasifikasi dokumen.

Berikut adalah penjelasan lebih lanjut tentang metode Pemrosesan Bahasa Alami (NLP) dalam program klasifikasi:

a) Preprocessing Teks:

Pada tahap ini, teks dalam dataset akan mengalami proses preprocessing untuk membersihkan teks dari karakter-karakter yang tidak relevan, seperti tanda baca, angka, atau karakter khusus. Selain itu, teks juga akan diubah menjadi lowercase dan tokenized (dipecah menjadi unit-unit teks yang lebih kecil, seperti kata-kata).

b) Representasi Vektor Teks:

Setelah preprocessing, teks perlu diubah menjadi representasi vektor numerik agar bisa digunakan dalam model machine learning. Metode NLP seperti TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengubah teks menjadi representasi vektor yang menggambarkan makna dan distribusi kata-kata dalam teks.

c) Pemilihan Algoritma Klasifikasi:

Setelah teks diubah menjadi vektor, selanjutnya Anda perlu memilih algoritma machine learning atau deep learning yang sesuai untuk tugas klasifikasi teks. Beberapa algoritma populer yang digunakan untuk klasifikasi teks adalah Naïve Bayes, Support Vector Machine (SVM), Random Forest, atau

algoritma deep learning seperti LSTM atau Transformer.

d) Pelatihan Model:

Setelah memilih algoritma klasifikasi, model akan dilatih menggunakan data latih yang telah diolah dan direpresentasikan sebagai vektor. Proses pelatihan dilakukan dengan menyesuaikan parameter model agar dapat memahami dan mengklasifikasi teks dengan tepat.

e) Evaluasi Model:

Setelah model dilatih, selanjutnya Anda harus mengukur performanya menggunakan data uji. Performa model dapat diukur dengan metrik-metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi model ini akan membantu Anda memahami seberapa baik model dalam melakukan klasifikasi teks.

Penerapan metode Pemrosesan Bahasa Alami (NLP) dalam program klasifikasi memungkinkan komputer untuk memahami dan menganalisis teks secara lebih mendalam, sehingga memungkinkan Anda untuk mengklasifikasikan teks dengan lebih tepat dan akurat.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pembahasan

Pada tahap scraping ini adalah tahap dimana pengambilan data terhadap ulasan Playstore. Untuk proses ini yang diambil adalah ulasan dari pengguna aplikasi GoTube di Playstore sebanyak 500 data ulasan. Penelitian ini menggunakan tools Google Play Scraper. Sebelum melakukan proses scraping ini peneliti harus meng-instal modul-modul seperti Google Play Scraper, Pandas, dan Numpy. Peneliti harus menggunakan jaringan internet pada saat melakukan proses scraping ini. Berikut adalah potongan gambar proses instalasi modul dan scraping :

```
[ ] from google_play_scraper import app
import pandas as pd
import numpy as np

#scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'premium.gotube.adblock.utube',
    lang='id', #disini kita mau men scrape data ulasan aplikasi shopee
    country='id', #kita setting bahasa nya menjadi bahasa indonesia
    sort=Sort.MOST_RELEVANT, # kemudian kita gunakan most_relevant unt
    count=500, # disini jumlah ulasan yang mau kita ambil
    filter_score_with=None # kemudian di filter_score kita gunakan No
```

4.1.1 Hasil Pelabelan

Setelah mendapatkan data dari proses scraping, maka di lanjutkan kedalam proses pelabelan yaitu pemberian nilai polaritas. Nilai polaritas yang digunakan yaitu, 4 (positif), 5 (positif), dan kurang dari 3 (negatif). Dibawah ini adalah potongan program pada pemberian nilai

polaritas:

```
[30] def pelabelan(score):
    if score < 3:
        return 'Negatif'
    elif score == 4:
        return 'Positif'
    elif score == 5:
        return 'Positif'
    my_df['Label'] = my_df['score'].apply(pelabelan)
    my_df.head(50)

my_df.to_csv("Labeling.csv", index = False) #kemudia
```

Gambar diatas menunjukkan perintah yang digunakan untuk pelabelan data. Hasil dari pelabelan data disimpan kedalam bentuk file CSV seperti yang ditunjukkan pada gambar dibawah ini:

	score	Label
n sampai saat ini iklan pun tidak ada	5	Positif
alagi keistimewaannya yang ditawarkan sejal	1	Negatif
ifter gmail gak bisa ,aplikasinya mutar mutar t	1	Negatif
di play store, dulu sy pake yt vanced karena s	5	Positif
kan knp Sering hilang sendiri kaya nonton tik	2	Negatif
sinya langsung layar hitam semua putih semu	3	
da yang kurang menurut saya yaitu Videonya g	5	Positif
iklan. biar bukan bentuk video	1	Negatif
punyai iklan-nya.	1	Negatif
si masuk kenapa suaranya jadi kecil dan lama l	2	Negatif
ih baik kurangi iklannya. Detik detik iklan trus	1	Negatif
h.tidak bisa cari lagu"..	1	Negatif

4.1.2 Hasil Text Pre-Processing

Setelah mendapatkan data dari proses labeling, langkah selanjutnya adalah melakukan proses pre-processing pada teks. Proses pre-processing bertujuan untuk membersihkan teks dari informasi yang tidak relevan atau yang tidak diperlukan dalam proses klasifikasi nantinya. Berikut adalah beberapa tahapan pada proses pre-processing dalam program Python:

a) Data cleaning

Data cleaning merupakan proses pembersihan data dengan nilai 3, karena nilai 3 merupakan nilai yang masih ambigu, maka tidak dibutuhkan. Potongan perintah yang digunakan beserta hasil dalam bentuk file CSV dari proses data cleaning dapat dilihat pada gambar dibawah ini:

```
import pandas as pd
pd.set_option('display.max_columns', None)
my_df = pd.read_csv('/content/scrapped_data.csv')
my_df.head(50)
```

	E	F
content		score
Setelah saya update kenapa ya loadingnya lama, padahal jaringan bagus. Dan pas buka ytb or		3
Aplikasi nya bagus bisa buat masang lagu tapi masih bisah iklan dan belum ada vitur shorts d		4
Appnya bagus bisa ngambang,dan bisa dengerin lagu sambil buka app lain,lebih praktis dari		5
GoTUBE adalah salah satu Apk yg sangat Bagus untuk di gunakan untuk menonton video dan		5
Terima kasih Aplikasi ini keren Tiap hari sy putar,tapi sayang kalau sinyal gak kuat musik berf		5
aplikasi ga jelas, kebanyakan boong, katanya (ANTI IKLAN) baru masuk aplikasi aja udah di sf		1
Sangat suka denga go tube tidak pernah ada iklan suka saat sedang mati musik masih meny		5
Wow keren bisa main di latar belakang videonya dan iklanya juga kurang munculnya, pokokr		5
Selalu terbaik.cuma 1 masalah.kalo di kembalikan dan mau masuk lagi ada iklannya.kalo yg l		5
Donwload Aplikasi ini keren banget sumpah deh kalo tidak Donwload bakal nyesel deh tidak		5
Sangat cocok...tanpa iklan,dpat di putar dilayar belakang.		5
Aplikasi nya bagus banget sumpah ,gw nyetel 10 video gada iklan sama sekali ,Nih aplikasi Ri		5
Aplikasi yg bagus...apalagi klo iklannya di awal buka aplikasi di hilangkan...lebih sempurna...		5
Setelah update sudah hilang kenyamanan saya dengan menonton tanpa iklan,, sebelum upc		2
Masih bintang 3 karena bisa play background, tapi iklan nya ngeselin tiap selesai lagu balik k		3
Sekarang bnyk sekali iklan nya, setiap buka tutup aplikasi atau layar mati pas hidup selalu ad		2
ga terlalu banyak iklan, aplikasi nya juga gratis, tapi pas masuk ke aplikasi GoTUBE kok tiba-ti		4
Ada versi premium berbayar yg bebas iklan ga? Mending berlangganan GoTUBE daripada You		5
Aplikasi nya bagus ,, emang betul gak da iklan nya ,,iklan nya klo mau keluar aja itu pon bisa		5
Assalamualaikum Alhamdulillah bagus tanpa iklan bisa di putar di layar belakang hemat bate		5
Aku sangat suka aplikasi ini karena ndak ada iklan yg mengganggu,sehingga tidak menggang		5
sekarang selalu muncul iklan adakami,iklan yang sangat sangat tidak berbobot,maf sa		1

b) Case folding

Tahapan kedua pada proses pre-processing yaitu case folding, dimana pada tahapan ini akan dilakukan pengubahan huruf besar menjadi huruf kecil semua pada data ulasan. Dibawah ini merupakan potongan dari program yang digunakan untuk proses case folding:

```
[ ] import re
def clean_text(df, text_field, new_text_field_name):
    my_df[new_text_field_name] = my_df[text_field].str.lower()
    my_df[new_text_field_name] = my_df[new_text_field_name].apply(lambda elem: re.
    # remove numbers
    my_df[new_text_field_name] = my_df[new_text_field_name].apply(lambda elem: re.
    return my_df

my_df['text_clean'] = my_df['content'].str.lower()
my_df['text_clean']
data_clean = clean_text(my_df, 'content', 'text_clean')
data_clean.head(10)
```

c) Stopword removal

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan stopwords yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya. Potongan perintah yang digunakan pada proses stopwords dapat dilihat pada gambar dibawah ini:

```
import nltk.corpus
nltk.download("stopwords")
from nltk.corpus import stopwords
stop = stopwords.words('indonesian')
data_clean['text_stopword'] = data_clean['text_clean'].apply(lambda
data_clean.head(50)
```

d) Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token

diartikan sebagai “kata” meskipun tokenize juga dapat dilakukan pada paragraf maupun kalimat. Sebagaimana potongan proses tokenizing dapat dilihat pada gambar dibawah ini:

```

+ Kode + Teks
import nltk
nltk.download('punkt')
from nltk.tokenize import sent_tokenize, word_tokenize
data_clean['text_tokens'] = data_clean['text_Stopword'].apply(lambda x: word_tokenize(x))
data_clean.head()
    
```

e) Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya. Untuk melakukan stemming bahasa Indonesia kita dapat menggunakan library Python Sastrawi yang sudah kita siapkan di awal. Library Sastrawi menerapkan Algoritma Nazief dan Adriani dalam melakukan stemming bahasa Indonesia. Potongan program yang digunakan dalam proses stemming dapat dilihat pada gambar dibawah ini:

```

[ ] from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

#-----STEMMING-----
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

# create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
    
```

Potongan hasil dari proses stemming ialah data dalam bentuk excel, dapat dilihat pada gambar dibawah ini:

content	score	text_clear	text_Stop	text_tokens
Aplikasi nya bagus bisa buat ma	4	aplikasi n\aplikasi n\	['aplikasi', 'n'	
Appnya bagus bisa ngambang,d	5	appnya ba appnya ba\	['appnya', 'ba'	
GoTUBE adalah salah satu Apk y	5	gotube adgotube sa\	['gotube', 'sa'	
Terima kasih Aplikasi ini keren	5	terima ka:terima ka\	['terima', 'ka'	
aplikasi ga jelas, kebanyakan bc	1	aplikasi ga\aplikasi ga\	['aplikasi', 'ga'	
Sangat suka denga go tube tidal	5	sangat suk\ suka deng\	['suka', 'deng\	
Wow keren bisa main di latar be	5	wow kere wow kere\	['wow', 'kere'	
Selalu terbaik.cuma 1 masalah.l	5	selalu ter\terbaikcu\	['terbaikcum'	
Donwlod Aplikasi ini keren ban	5	donwlod d\donwlod d\	['donwlod', 'd'	
Sangat cocok...tanpa iklan,dpat	5	sangat coc\cocoktanp\	['cocoktanpa'	

f) Pembobotan TF-IDF

Pada tahap perhitungan tfidf menggunakan library TfidfVectorizer. Potongan perintah yang digunakan dalam proes tfidf dapat dilihat pada gambar dibawah ini:

```

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)

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

(348,)
(348,)
(88,)
(88,)
    
```

4.1.3 Hasil Klasifikasi

Pada proses klasifikasi ini menggunakan metode klasifikasi naïve bayes yang telah disediakan oleh library python yang bernama scikit-learn. Dibawah ini adalah potongan dari program library python scikit-learn:

```

from sklearn.naive_bayes import MultinomialNB

nb = MultinomialNB()
nb.fit(tfidf_train, y_train)

MultinomialNB
MultinomialNB()
    
```

Setelah proses impor paket MultinomialNB, tahap selanjutnya dilakukan impor paket akurasi juga dari library sklearn. Sebagaimana ditampilkan pada gambar dibawah ini:

```

from sklearn.metrics import accuracy_score

accuracy = accuracy_score(y_test, y_pred)
    
```

Kemudian setelah semua library paket sudah siap, dilakukanlah proses klasifikasi pada data hasil pre-processing. Berikut merupakan potongan program yang digunakan dalam proses klasifikasi:

```

clf = MultinomialNB()
clf.fit(X_train, y_train)
predicted = clf.predict(X_test)

print("MultinomialNB Accuracy:", accuracy_score(y_test,predicted))
print("MultinomialNB Precision:", precision_score(y_test,predicted, average="binary"))
print("MultinomialNB Recall:", recall_score(y_test,predicted, average="binary"), po
print("MultinomialNB f1_score:", f1_score(y_test,predicted, average="binary"), po
    
```

Hasil dari klasifikasi naïve bayes menggunakan system yang telah dibuat ditampilkan pada gambar dibawah ini:

```

MultinomialNB Accuracy: 0.8089887640449438
MultinomialNB Precision: 0.7083333333333334
MultinomialNB Recall: 0.6296296296296297
MultinomialNB f1_score: 0.6666666666666667
confusion_matrix:
[[17 10]
 [ 7 55]]

=====
              precision    recall  f1-score   support

Negatif         0.71         0.63         0.67         27
Positif         0.85         0.89         0.87         62

accuracy         0.81         0.81         0.81         89
macro avg        0.78         0.76         0.77         89
weighted avg     0.80         0.81         0.81         89
    
```

Pada gambar diatas menunjukkan hasil dari klasifikasi Naïve Bayes, pada gambar diatas menunjukkan system sudah bisa menampilkan hasil dari accuracy, precision, recall, F1 Score, dan, confusion matrix. Perhitungan manual pada aplikasi excel bisa dilihat pada gambar dibawah ini:

no.	content	score	Label	probabilitas kelas	
1	cuma ada d	4	Positif	label	nilai
2	pp lain,leb	5	Positif	positif	0,65367
3	gunakan de	5	Positif	negatif	0,34633
4	butar,tapi	5	Positif	akurasi NBC	100%
5	ANTI IKLA	1	Negatif		
6	ng mati mu	5	Positif		
7	ya dan ikla	5	Positif		
8	hya.kalo yg	5	Positif		
9	l deh tidak	5	Positif		
10	an,dpat d	5	Positif		

actual		actual	
prediksi TP	FN	17	10
prediksi FP	TN	7	55

akurasi	80,89888
presisi	70,83333
recall	62,96296
f1 score	66,66667

Pada gambar diatas menampilkan perhitungan manual confusion matrix menggunakan formula excel.

Untuk pemaparan perhitungan pada hasil proses klasifikasi bisa dilihat pada perhitungan sebagai berikut.

a. Accuracy

Accuracy adalah sebuah penilaian prediksi benar (true) positif, negatif, dan netral dari data yang digunakan. Pada penelitian ini mendapatkan nilai akurasi sebesar 80,% diperoleh dari perhitungan:

$$Accuracy = \frac{17 + 55}{17 + 55 + 10 + 7} \times 100$$

$$Accuracy = 80,89\%$$

b. Precision

Precision adalah sebuah penilaian prediksi benar (true) pada setiap kelas yang ada dan dibagi dengan hasil keseluruhan prediksi di kelas tersebut. Hasil Precision diperoleh dari perhitungan:

$$Precision = \frac{17}{17 + 7}$$

$$Precision = 0,7083333$$

Berdasarkan hasil diatas menunjukkan nilai prediksi yang diberi label positif benar-benar relevan atau akurat dengan hasil presisi sebesar 0,7083 atau 70,83%.

c. Recall

Recall adalah sebuah penilaian prediksi benar (true) pada setiap kelas yang ada dan dibagi dengan hasil keseluruhan nilai dari kelas

tersebut. Hasil recall diperoleh dari perhitungan:

$$Recall = \frac{17}{17 + 10}$$

$$Recall = 0,6296296$$

Berdasarkan hasil diatas, model mampu mengidentifikasi semua instance positif dengan nilai sebesar 0,6296 atau 62,96%.

d. F1-Score

F1-Score adalah sebuah nilai rata-rata dari hasil perhitungan precision dan recall. Hasil F1-Score sendiri diperoleh dari perhitungan:

$$F1 - Score = 2 \times \frac{0,7083333 \times 0,6296296}{0,7083333 + 0,6296296}$$

$$F1 - Score = 0,66666$$

Hasil dari fi-score menunjukkan ukuran tentang kinerja model klasifikasi yang mempertimbangkan keseimbangan antara precision dan recall.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil sentiment tertinggi yang diperoleh dalam penelitian ini ialah sentiment positif dengan persentase 65%, berdasarkan hasil tersebut mengindikasikan ulasan dari pengguna aplikasi Gotube cenderung positif.

5.2 Saran

Pada penelitian ini hanya menggunakan satu metode yaitu naïve bayes, untuk kedepannya bisa menggunakan lebih dari satu metode untuk perbandingan.

DAFTAR PUSTAKA

Aini, Qurotul, et al. "Exploring E-Learning Challenges During the Global COVID-19 Pandemic: A Review." *Jurnal Sistem Informasi*, vol. 16, no. 2, 2020, pp. 57–65, <https://doi.org/10.21609/jsi.v16i2.1011>.

Aldio Nur Samsi, Yulian Findawati, Irwan A. Kautsar. *Analisa Sentimen Mahasiswa Terhadap Minat Berorganisasi Dengan Metode Naïve Bayes*. 2022, pp. 171–76.

Artani Inez Tangraeni, Melkior N. Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes. no. 2, 2022, pp. 785–95.

Astari, Ni Made Ayu Juli, et al. "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naïve Bayes Classifier." *Jurnal Sistem Dan Informatika (JSI)*, vol. 15, no. 1, 2020, pp. 27–29, <https://doi.org/10.30864/jsi.v15i1.332>.

Daryfayi, Edyt, et al. *Sentimen Analisis Pada*

- Ulasan Google Play Store Menggunakan Metode Naïve Bayes.* no. 2, 2020, pp. 8400–10.
- Erfina, Adhitia, and Muhamad Fani Al-shufi. *Analisis Sentimen Aplikasi Jasa Kurir Di Play Store.* no. 2, 2022, pp. 103–10.
- Fitriyah, Nur, et al. “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm).” *Jurnal Gaussian*, vol. 9, no. 3, 2020, pp. 376–90, <https://doi.org/10.14710/j.gauss.v9i3.28932>.
- Hayuningtyas, Ratih Yulia, and Retno Sari. “Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso.” *Jurnal Techno Nusa Mandiri*, vol. 16, no. 1, 2019, pp. 37–42, <https://doi.org/10.33480/techno.v16i1.115>.
- Herlinawati, Nuraeni, et al. *Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine.* no. 2, 2020, pp. 293–98.
- Pratmanto, Dany, et al. “App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm.” *Journal of Physics: Conference Series*, vol. 1641, no. 1, 2020, <https://doi.org/10.1088/1742-6596/1641/1/012043>.
- Rita Apriani, Dudih Gustian. *Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia.* no. 1, 2019.
- Rizki, Muhammad. “Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction.” *Jurnal Teknologi*, vol. 21, no. 1, 2021, p. 39, <https://doi.org/10.30811/teknologi.v21i1.2209>.
- Singgalen, Yerik Afrianto. *Pemilihan Metode Dan Algoritma Dalam Analisis Sentimen Di Media Sosial: Systematic Literature Review.* no. 2, 2021, pp. 278–302.
- Tri Romadloni, Nova, et al. “Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line.” *Jurnal IKRA-ITH Informatika*, vol. 3, no. 2, 2019, pp. 1–9.
- Tuhuteru, Hennie, and Ade Iriani. “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine Dan Naive Bayes Classifier.” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, 2018, pp. 394–401, <https://doi.org/10.30591/jpit.v3i3.977>.