**Analisis Sentimen Ulasan Platform Media Sosial Menggunakan Algoritma Naïve Bayes Classifier**

Sudin Saepudin[1]\*, Selviani Widiastuti[2], Carti Irawan[3]

Program Studi Sistem Informasi, Fakultas Teknik Komputer dan Desain[1], [2], [3]

Universitas Nusa Putra

Sukabumi, Indonesia

sudin.saepudin@nusaputra.ic.id[1], selviani.widiastuti\_si19@nusaputra.ic.id[2], carti@nusaputra.ic.id[3]

***Abstract*—** **The existence of the Covid-19 pandemic has caused lifestyle changes in society, these changes are supported by increasingly rapid technological developments, making the number of internet usage increase which encourages the spread of information very quickly through social media platforms. Not only for self-expression, social media can also be a means of communication, information, education, and even used as a marketing tool. Several social media platforms have recently been popular and widely used, the number of users is increasing from year to year, and each user can provide a rating review of the application. To find out public opinion on social media platforms, sentiment analysis will be carried out on several social media platform applications on the Google Play Store, namely Twitter, Instagram and Tiktok which will later be used as material for evaluating these applications. In this study, the dataset was taken based on ratings from user reviews on the Google Play Store using the NBC (Naïve Bayes Classifier) ​​method with the Python programming language. Based on testing of 1000 datasets, it was found that the majority gave positive sentiment (Twitter 57.2%, Instagram 74.1%, Tiktok 83.9%), and negative sentiment (Twitter 42.8%, Instagram 25.9%, Tiktok 16 .1%) with an accuracy rate of 85.6% for the Twitter application, 83.6% for the Instagram application, and 84.8% for the Tiktok application.**

**Keywords—** **Sentiment Analysis, Google Play Store, NBC (Naïve Bayes Classifier), Social Media Platform, Python.**

***Abstrak*—** **Adanya pandemi covid-19 menyebabkan terjadinya perubahan gaya hidup pada masyarakat, perubahan tersebut didukung dengan perkembangan teknologi yang semakin pesat, membuat angka penggunaan internet meningkat yang mendorong penyebaran informasi sangat cepat melalui *platform media sosial*. Tak hanya untuk mengekspresikan diri, media sosial juga dapat menjadi alat komunikasi, informasi, edukasi, bahkan dijadikan alat pemasaran. Beberapa *platform media sosial* belakangan ini sedang populer dan marak digunakan, jumlah penggunanya semakin meningkat dari tahun ke tahun, dan setiap pengguna dapat memberikan ulasan penilaian terhadap aplikasi tersebut. Untuk mengetahui opini masyarakat terhadap *platform media sosial* maka akan dilakukan analisis sentimen pada beberapa aplikasi *platform media sosial* di *google play store* yaitu Twitter, Instagram, dan Tiktok yang nantinya akan digunakan sebagai bahan evaluasi terhadap aplikasi tersebut. Dalam penelitian ini dataset diambil berdasarkan rating dari review pengguna di *google play store* menggunakan metode NBC (Naïve Bayes Classifier) dengan bahasa pemrograman python.  Berdasarkan pengujian dari 1000 dataset, didapatkan sebagian besar memberi sentimen positif (Twitter 57,2%, Instagram 74,1%, Tiktok 83,9%), dan sentimen negatif (Twitter 42,8%, Instagram 25,9%, Tiktok 16,1%) dengan nilai tingkat akurasi 85,6% untuk aplikasi Twitter, 83,6% untuk aplikasi Instagram, dan 84,8% untuk aplikasi Tiktok.**

***Kata Kunci—*** ***Analisis Sentimen, Google Play Store, NBC (Naïve Bayes Classifier), Platform Media Sosial, Python.***

#  PENDAHULUAN

Adanya pandemi covid-19 menyebabkan terjadinya perubahan gaya hidup pada masyarakat. Perubahan tersebut didukung dengan perkembangan teknologi yang semakin pesat, membuat angka penggunaan internet meningkat yang mendorong penyebaran informasi sangat cepat melalui *platform media sosial*. Pada saat ini *platform media sosial* sedang marak digunakan untuk media komunikasi, informasi, edukasi, bahkan dijadikan media pemasaran. Ada beberapa *platform media sosial* yang belakangan ini sedang populer dengan fitur serta kegunaan yang hampir sama yaitu Tiktok, Instagram, dan Twitter. Jumlah pengguna aplikasi tersebut semakin meningkat dari tahun ke tahun semenjak adanya pandemi. Setelah melakukan studi literatur dari beberapa sumber, peneliti menemukan data tiga tahun kebelakang jumlah pengguna dari aplikasi Twitter, Instagram, dan Tiktok di website Databoks – Katadata. Berikut ini grafik jumlah pengguna dari ketiga aplikasi tersebut.

Fig. 1.Grafik Jumlah Pengguna Aplikasi Platform Media Sosial

Dari Fig.1. dapat dilihat perbandingan jumlah pengguna dari masing-masing aplikasi *platform media sosial* dari tahun 2020-2022, dimana aplikasi Twitter pada tahun 2020 memiliki sebanyak 166 juta pengguna, pada tahun 2021 memiliki sebanyak 199 juta pengguna, dan pada tahun 2022 memiliki sebanyak 229 juta pengguna, selanjutnya aplikasi Instagram pada tahun 2020 memiliki sebanyak 1,734 miliar pengguna, pada tahun 2021 memiliki sebanyak 1,878 miliar pengguna, dan pada tahun 2022 memiliki sebanyak 1,961 miliar pengguna, terakhir yaitu aplikasi Tiktok pada tahun 2020 memiliki sebanyak 700 juta pengguna, pada tahun 2021 memiliki sebanyak 902 juta pengguna, dan pada tahun 2022 memiliki sebanyak 1,460 miliar pengguna.

Dari latar belakang masalah yang ada maka dilakukan analisis sentimen pada penelitian ini untuk mengetahui opini masyarakat tentang platform media sosial di *Google Play Store* [1]. Nantinya data ulasan yang sudah dikumpulkan akan dibagi menjadi tiga label yaitu negatif, netral, dan positif yang akan diproses sehingga dapat diketahui hasil akurasi dari setiap sentimen pengguna terhadap aplikasi Twitter, Instagram, dan Tiktok. dalam penelitian ini menggunakan algoritma NBC (Naïve Bayes Classifier) dengan bahasa pemrograman python.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Sola Fide, Suparti, Sudarno dengan judul “Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Store Menggunakan Metode Support Vector Machine (SVM) dan Asosiasi” Berdasarkan data sentimen sebanyak 3200 ulasan dari September 2020 hingga Februari 2021 di Google Play, pada aplikasi Tiktok ada lebih banyak ulasan positif daripada ulasan negatif, dengan ulasan positif sebanyak 1741 (54.41%), dan ulasan negatif sebanyak 1459 (45.59%). Menggunakan metode Support Vector Machine (SVM) kernel RBF dengan perbandingan data training dan data testing sebesar 80:20, klasifikasi sentimen dari hasil sentiment scoring review aplikasi TikTok di Google Play menghasilkan akurasi dan tingkat kappa terbaik sebesar 90,62% dan 81,24%, yang menandakan termasuk hasil klasifikasi yang hampir sempurna [2].

Berdasarkan penelitian sebelumnya yang dilakukan oleh Sifa Melina Salsabila, Aang Alim Murtopo, dan Nurul Fadhilah yang berjudul “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier” Data pada penelitian ini diambil dari Google Play Store. Algoritma Naïve Bayes Classifier digunakan dalam menganalisis ulasan pengguna aplikasi Tokopedia di play store. Pengujian yang dilakukan oleh rapid evaluator terhadap 143 data pengujian menghasilkan nilai akurasi 95,10%, kelas recall 91,46% (kelas positif: Positif), presisi kelas 100,00% (kelas positif: Positif), dan nilai AUC 0,999. (kelas positif: Positif) [3].

Berdasarkan penelitian sebelumnya yang dilakukan oleh Cholid Fadilah Hasri, dan Debby Alita yang berjudul “Penerapan Metode Naïve Bayes Classifier dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter” Twitter digunakan untuk mengumpulkan data pada penelitian ini. Metode yang digunakan yaitu Naive Bayes Classifier (NBC). Hasil dari penerapan algoritma Naive Bayes pada penelitian ini menunjukan nilai akurasi sebesar 81,07% dalam mengklasifikasikan sentimen tanpa penambahan fitur. Untuk memastikan hasil dari penelitian ini juga menggunakan metode Support Vector Machine, yang menghasilkan nilai akurasi 79,96% [4].

Pada beberapa penelitian yang dipaparkan sebelumnya, penelitian hanya berfokus pada satu objek saja, dalam penelitian ini akan ada tiga aplikasi sebagai objek penelitian yang ada di Google Play Store, analisis sentimen dilakukan menggunakan algoritma Naïve Bayes, tools Google Colab, dan bahasa pemrograman Python.

# METODE PENELITIAN

Penelitian ini melalui beberapa tahapan dimulai dari web scraping, preprocessing text, klasifikasi algoritma NBC, analisis hasil pengujian dan testing.

1. Web Scraping

*Web scraping* merupakan teknik yang digunakan untuk mendapatkan informasi dari suatu situs web yang diinginkan secara otomatis tanpa harus menyalin secara manual [5]. Web scraping dapat didefinisikan juga sebagai metode untuk melakukan ekstraksi informasi tertentu dari situs web yang dituju, kemudian dapat dilakukan analisis dimanfaatkan untuk berbagai tujuan [6].

1. Preprocessing Text

*Pre-processing* merupakan tahapan awal pengolahan data menjadi analisa sentimen yang bertujuan untuk menghilangkan kata atau kalimat yang tidak perlu dan menyamakan bentuk kata untuk mengurangi volume pada kosakata [7]. Langkah-langkah pada *pre-processing* yaitu sebagai berikut :

1. *Case Folding*

*Case folding* adalah suatu proses mengubah semua huruf pada suatu dokumen atau kalimat menjadi huruf kecil [8]. Hal tersebut dilakukan agar lebih mudah dan tidak memakan banyak waktu dalam proses pembacaan dalam mesin terhadap korpus [9].

1. *Tokenizing*

Tokenisasi adalah proses yang memisahkan suatu rangkaian karakter berdasarkan spasi, serta memungkinkan terjadi proses penghapusan karakter tertentu seperti tanda baca pada waktu yang bersamaan [10]. Tujuan dari dilakukannya tokenisasi ini adalah agar dalam proses selanjutnya tidak memproses berdasarkan kalimat lagi tetapi memproses kata per kata [9].

1. *Cleansing*

*Cleansing* adalah suatu proses pembersihan data teks dengan menghilangkan data yang tidak konsisten atau tidak diperlukan. Cleansing dilakukan guna menganalisa kualitas data yang memiliki format data atau file yang salah guna menghasilkan data berkualitas tinggi [11].

1. *Normalization*

*Normalization* adalah proses pengubahan kata tidak baku pada data komentar yang telah dikumpulkan menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI), hal tersebut bertujuan agar kata yang diproses menjadi bermakna seperti yang ada pada KBBI [9].

1. *Stemming*

Kata-kata dari data komentar yang telah melewati beberapa proses sebelumnya, kemudian dilakukan pengecekan kembali yang merupakan tahapan akhir dalam *preprocessing* [10]. Proses tersebut disebut stemming yang tujuannya menghilangkan semua imbuhan yang ada pada suatu kata baik yang berposisi di awal, akhir, atau kombinasi dari keduanya [9].

1. Klasifikasi Algoritma NBC

*Naïve Bayes Classifier* merupakan sebuah pengklasifikasi yang paling sederhana dan paling sering digunakan. *Naïve Bayes* adalah machine learning yang menggunakan probabilitas dengan konsep pendekatan *Bayes* sebagai perhitungannya [12]. Perhitungan probabilitas teorema *Bayes* pada *Naïve Bayes Classifier* mengasumsikan nilai atribut suatu kelas tidak ada keterkaitan dengan keberadaan atribut kelas lain, meskipun kenyataannya setiap atribut hampir memiliki ketergantungan satu sama lain [13]. Dapat disimpulkan, *Naïve Bayes* merupakan metode untuk klasifikasi teks dengan kecepatan pemrosesan yang tinggi untuk data yang besar [14]. Dengan asumsi yang dibuat seperti itu, pada akhirnya membuat *Naïve Bayes* mudah dalam melakukan perhitungan [13]. Perhitungan *Naïve Bayes* dirumuskan dalam persamaan sebagai berikut :



Fig. 2.Rumus Dasar Persamaan Teorema Bayes

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesis X merupakan suatu kelas spesifik

P(H|X) = Probabilitas H berdasarkan kondisi X

P(H) = Probabilitas pada hipotesis H (prior)

P(X|H) = Probabilitas X berdasarkan hipotesis H

P(X) = Probabilitas X (data sampel yang diamati)

Pada proses klasifikasi teks metode *Naïve Bayes Classifier* terdapat dua tahapan, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses terhadap sampel data yang nantinya diharapkan dapat menjadi representasi data tersebut. Sedangkan tahap klasifikasi merupakan penentuan probabilitas *prior* bagi setiap kategori berdasarkan sampel data pada tahap pelatihan [15].

# HASIL DAN PEMBAHASAN

Dalam bagian ini akan dijelaskan mengenai hasil penelitian yang telah dilakukan, dimulai dari proses pengumpulan data, klasifikasi naïve bayes, hasil klasifikasi (*Accuracy, Precision, Recall* dan *f1-score*), dan visualisasi data.

1. Pengumpulan Data

Pengumpulan data menggunakan metode scraping dengan data yang diperoleh sebanyak 1000 ulasan komentar dari masing-masing aplikasi pada website google play store.

TABEL 1. PROSES PRE-PROCESSING APLIKASI TWITTER

|  |  |  |  |
| --- | --- | --- | --- |
| **User Name** | **Content** | **Score** | **At** |
| jayani tinenta | Biasa | 2 | 2023-02-23 08:22:11 |
| Natan Markus | Akun yang tidak pernah posting apa-apa pun tet... | 1 | 2023-02-23 08:20:20 |
| Angga Saputra | Ga bisa buat mutar video. Tiap hari harus unin… | 1 | 2023-02-23 07:56:01 |
| Arsikin Sugeng | Saya belum bisa login Twitter  | 4 | 2023-02-23 07:50:06 |
| achmad Naufal | #Quoteshariini Jangan pantang menyerah | 1 | 2023-02-23 07:47:45  |

TABEL 2. PROSES PRE-PROCESSING APLIKASI INSTAGRAM

|  |  |  |  |
| --- | --- | --- | --- |
| **User Name** | **Content** | **Score** | **At** |
| siggit semito adi | Ok. | 5 | 2023-02-23 07:05:33 |
| Muhammad Khoirulumam | Apk aneh engak bisa buat login | 1 | 2023-02-23 07:05:25 |
| Mohamad Wahyudi | sip pokok wkwkwkwk anjay follow ig ku lurrr 😂 | 5 | 2023-02-23 07:03:30 |
| 123 GO!!! INDO  | Saya sangat senang dengan akun instagram sebag... | 5 | 2023-02-23 07:03:23 |
| Naryo Naeryl123 | Love Instagram | 5 | 2023-02-23 07:00:22 |

TABEL 3. PROSES PRE-PROCESSING APLIKASI TIKTOK

|  |  |  |  |
| --- | --- | --- | --- |
| **User Name** | **Content** | **Score** | **At** |
| KTUK Budayasa  | Aplikasi ini sangat bagus | 1 | 2023-02-23 07:15:02 |
| Ridho Saputra | Bagus keren kayak babi | 5 | 2023-02-23 07:13:50 |
| ALVARO NET | Ok | 5 | 2023-02-23 07:12:59 |
| Ilda Rasyida | apps nya bagus bangett ga nyesel deh | 5 | 2023-02-23 07:12:54 |
| Icha Chelsea | Apk ini bagus | 5 | 2023-02-23 07:12:18 |

Pada TABEL 4, TABEL 5, dan TABEL 6 terdapat proses *pre-processing* data, yaitu data mentah hasil *scraping* akan melalui tahap *case folding* (data diubah menjadi huruf kecil) dan juga dilakukan *cleaning* (menghilangkan karakter, url, hashtag dan simbol) hingga mendapatkan hasil komentar review. Selanjutnya dilakukan pemotongan kata pada kalimat komentar yang dipisahkan oleh tanda koma (,) pada tahap *tokenizing*. Lalu akan dilakukan proses *filtering* untuk menghilangkan kata-kata tidak bermanfaat. Tahapan akhir *pre-processing* yaitu *stemming* dimana akan memastikan data yang mengandung kata imbuhan akan dihilangkan agar mendapatkan kumpulan data yang berisi kata-kata dasar.

TABEL 4. HASIL PRE-PROCESSING APLIKASI TWITTER

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Folding dan Cleaning** | **Tokenizing** | **Filtering** | **Stemming** |
| Biasa | [biasa] | [] | NaN |
| akun yang tidak pernah posting apa-apa pun tet... | [akun, yang, tidak, pernah, posting, apa-apa, ... | [akun, posting, apa-apa, ditangguhkan] | akun posting apa Tangguh |
| ga bisa buat mutar video tiap hari harus unins...  | [ga, bisa, buat, mutar, video, tiap, hari, har... | [ga, mutar, video, uninstall, aplikasi, downlo... | ga mutar video uninstall aplikasi download bia... |
| saya belum bisa login twitter  | [saya, belum, bisa, login, twitter] | [login, twitter] | login twitter |
| #quoteshariini jangan pantang menyerah | [#quoteshariini, jangan, pantang, menyerah] | [#quoteshariini, pantang, menyerah] | quoteshariini pantang serah |

TABEL 5. HASIL PRE-PROCESSING APLIKASI INSTAGRAM

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Folding dan Cleaning** | **Tokenizing** | **Filtering** | **Stemming** |
| Ok | [ok] | [ok] | Ok |
| apk aneh engak bisa buat login | [apk, aneh, engak, bisa, buat, login] | [apk, aneh, engak, login] | apk aneh engak login |
| sip pokok wkwkwkwk anjay follow ig ku lurrr 😂 | [sip, pokok, wkwkwkwk, anjay, follow, ig, ku, ... | [sip, pokok, wkwkwkwk, anjay, follow, ig, ku, ... | sip pokok wkwkwkwk anjay follow ig ku lurrr |
| saya sangat senang dengan akun instagram sebag... | [saya, sangat, senang, dengan, akun, instagram... | [senang, akun, instagram, media, menyalurkan, ... | senang akun instagram media salur inspirasi ma... |
| love Instagram | [love, instagram] | [love, instagram] | love instagram |

TABEL 6. HASIL PRE-PROCESSING APLIKASI TIKTOK

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Folding dan Cleaning** | **Tokenizing** | **Filtering** | **Stemming** |
| aplikasi ini sangat bagus | [aplikasi, ini, sangat, bagus] | [aplikasi, bagus] | aplikasi bagus |
| bagus keren kayak babi | [bagus, keren, kayak, babi] | [bagus, keren, kayak, babi] | bagus keren kayak babi |
| Ok | [ok] | [ok] | Ok |
| apps nya bagus bangett ga nyesel deh | [apps, nya, bagus, bangett, ga, nyesel, deh] | [apps, nya, bagus, bangett, ga, nyesel, deh] | apps nya bagus bangett ga nyesel deh |
| apk ini bagus | [apk, ini, bagus] | [apk, bagus] | apk bagus |

1. Klasifikasi Naïve Bayes

Proses klasifikasi dilakukan menggunakan Multinomial Naïve Bayes dengan metode splitting (80:20), 80% data akan digunakan sebagai data training, dan 20% data akan digunakan sebagai data testing. Setelah melalui tahap pengujian dimana yang awalnya dari masing-masing aplikasi diambil 1000 data, beberapa komentar dihilangkan karena nilai rating adalah bias sehingga data yang akan digunakan untuk aplikasi Twitter sebanyak 940, aplikasi Instagram sebanyak 943, dan aplikasi Tiktok sebanyak 955. Data yang digunakan merupakan random dari dataset yang telah dibersihkan.

TABEL 7. MATRIKS PREDIKSI APLIKASI TWITTER

|  |  |
| --- | --- |
| Prediksi | Aktual |
| Negatif | Positif |
| Negatif | 56 | 18 |
| Positif | 9 | 105 |

TABEL 8. MATRIKS PREDIKSI APLIKASI INSTAGRAM

|  |  |
| --- | --- |
| Prediksi | Aktual |
| Negatif | Positif |
| Negatif | 25 | 30 |
| Positif | 1 | 133 |

TABEL 9. MATRIKS PREDIKSI APLIKASI TIKTOK

|  |  |
| --- | --- |
| Prediksi | Aktual |
| Negatif | Positif |
| Negatif | 2 | 29 |
| Positif | 0 | 160 |

*Confusion matrix* :

* *True Negative* (TN): Model memprediksi data terdapat pada kelas Negatif dan yang sebenarnya data memang terdapat pada kelas Negatif.
* *True Positive* (TP): Model memprediksi data terdapat pada kelas Positif dan yang sebenarnya data memang terdapat di kelas Positif.
* *False Negative* (FN): Model memprediksi data berada di kelas Negatif, tetapi data sebenarnya  ada di kelas Positif.
* *False Positive* (FP): Model memprediksi data berada di kelas Positif, tetapi data sebenarnya ada di kelas Negatif.

Dari hasil pengujian menggunakan multinomial *Naive Bayes* yang dilakukan dengan ulasan komentar aplikasi Twitter sebanyak 188 data, aplikasi Instagram sebanyak 189 data, dan aplikasi Tiktok sebanyak 191 data (20% dari dataset secara random). Dari TABEL 7, TABEL 8, dan TABEL 9 didapatkan hasil *confusion matrix* :

* TN aplikasi Twitter sebanyak 56, aplikasi Instagram sebanyak 25, dan aplikasi Tiktok sebanyak 2,
* FP aplikasi Twitter sebanyak 18, aplikasi Instagram sebanyak 30, dan aplikasi Tiktok sebanyak 29,
* TP aplikasi Twitter sebanyak 105, aplikasi Instagram sebanyak 133, dan aplikasi Tiktok sebanyak 160,
* FN aplikasi Twitter sebanyak 9, aplikasi Instagram sebanyak 1, dan aplikasi Tiktok sebanyak 0.
1. Hasil Klasifikasi (*Accuracy, Precision, Recall* dan *f1-score*)

Hasil dari klasifikasi algoritma *Naive Bayes* dengan menggunakan pembobotan TF-IDF didapatkan hasil sebagai berikut :



Fig. 3. Hasil Klasifikasi Algoritma Multinomial Naive Bayes Pada Aplikasi Twitter



Fig. 4. Hasil Klasifikasi Algoritma Multinomial Naive Bayes Pada Aplikasi Instagram



Fig. 5. Hasil Klasifikasi Algoritma Multinomial Naive Bayes Pada Aplikasi Tiktok

Berikut ini merupakan penjelasan dari Fig. 3, Fig. 4, dan Fig 5, nilai Precision (Perbandingan data antara True Positive “TP” dengan banyak data yang diprediksi Positif) dihasilkan menggunakan algoritma Multinomial Naive Bayes didapatkan :

* Aplikasi Twitter sebesar 0.86153
* Aplikasi Instagram sebesar 0.96153
* Aplikasi Tiktok sebesar 1.0

Nilai *Recall* (Perbandingan banyak data antara *True Positive* “TP” dengan banyak data yang sebenarnya positif) yang didapatkan :

* Aplikasi Twitter sebesar 0.75675
* Aplikasi Instagram sebesar 0.45454
* Aplikasi Tiktok sebesar 0.06451

F1-*Score* (Nilai *harmonic* antara *Precision* dengan *Recall*) yang didapatkan :

* Aplikasi Twitter sebesar 0.80575
* Aplikasi Instagram sebesar 0.61728
* Aplikasi Tiktok sebesar 0.12121

Dengan keakuratan algoritma yang digunakan sebesar 0.85638 untuk aplikasi Twitter, 0.83597 untuk aplikasi Instagram, dan 0.84816 untuk aplikasi Tiktok. Dapat dikatakan bahwa model klasifikasi memiliki nilai yang sangat baik.

1. Visualisasi Data

 Hasil visualisasi data dari melakukan pengklasifikasian kelas sentimen positif dan negatif divisualisasikan sebagai berikut:



Fig. 6. Diagram Pie Sentimen Positif dan Sentimen Negatif Aplikasi Twitter



Fig. 7. Diagram Pie Sentimen Positif dan Sentimen Negatif Aplikasi Instagram



Fig. 8. Diagram Pie Sentimen Positif dan Sentimen Negatif Aplikasi Tiktok

Dari Fig. 6, Fig. 7, dan Fig. 8 dapat dilihat bahwa aplikasi Twitter menghasilkan 57.2% sentimen positif, 42.8% sentimen negatif, aplikasi Instagram menghasilkan 74.1% sentimen positif, 25.9% sentimen negatif, dan aplikasi Tiktok menghasilkan 83.9% sentimen positif, 16.1% sentimen negatif.

# KESIMPULAN DAN SARAN

Dalam penelitian ini peneliti melakukan analisis sentiment pada aplikasi platform media sosial (Twitter, Instagram, dan Tiktok). Data yang dikumpulkan melalui penggunaan web scraping berupa data ulasan komentar dari tiga aplikasi platform media sosial tersebut di Google Play Store, dengan total 3000 data ulasan komentar, masing-masing aplikasi diambil 1000 ulasan komentar. Dari hasil yang didapatkan sebagian besar pengguna memberi sentimen positif terhadap masing-masing aplikasi (Twitter 57,2%, Instagram 74,1%, Tiktok 83,9%), adapun sentimen negatif (Twitter 42,8%, Instagram 25,9%, Tiktok 16,1%), dengan menghasilkan nilai tingkat akurasi 85,6% untuk aplikasi Twitter, 83,6% untuk aplikasi Instagram, dan 84,8% untuk aplikasi Tiktok. Nilai akurasi tersebut sudah cukup baik dan tinggi dalam menghasilkan sebuah model prediksi. Dalam penelitian ini, penulis menyadari bahwa masih banyak kekurangan dan keterbatasan. Oleh sebab itu berdasarkan penelitian ini, penulis menyarankan untuk penelitian sejenis selanjutnya untuk meningkatkan nilai akurasi menjadi lebih baik dengan menggunakan algoritma klasifikasi yang lain, dan diharapkan pengambilan data dari media lain seperti Twitter, Youtube, dan lain sebagainya.

##### References

[1] M. F. Al-shufi and A. Erfina, “Sentimen Analisis Mengenai Aplikasi Streaming Film Menggunakan Algoritma Support Vector Machine Di Play Store,” *J. SISMATIK*, vol. 01, no. 01, 2021, Accessed: Apr. 14, 2023. [Online]. Available: https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/22

[2] S. Fide, Suparti, and Sudarno, “Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (SVM) Dan Asosiasi,” *J. Gaussian*, vol. 10, no. 3, pp. 346–358, 2021.

[3] S. M. Salsabila, A. A. Murtopo, and N. Fadhilah, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” *J. Minfo Polgan*, vol. 11, no. 02, 2022, [Online]. Available: www.tokopedia.com

[4] C. F. Hasri and D. Alita, “Penerapan Metode Naive Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[5] S. Satriajati, S. B. Panuntun, and S. Pramana, “Implementasi Web Scraping Dalam Pengumpulan Berita Kriminal Pada Masa Pandemi Covid-19 Studi Kasus: Situs Berita detik.com,” *Semin. Nas. Off. Stat.*, no. 01, 2020.

[6] M. Djufri, “Penerapan Teknik Web Scraping Untuk Penggalian Potensi Pajak (Studi Kasus pada Online Market Place Tokopedia, Shopee dan Bukalapak),” *J. BPPK*, vol. 13, no. 02, pp. 65–75, 2020.

[7] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.

[8] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *J. TEKNO KOMPAK*, vol. 15, no. 1, 2021.

[9] R. P. Sidiq, B. A. Dermawan, and Y. Umaidah, “Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes,” *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 356, Sep. 2020, doi: 10.32493/informatika.v5i3.6571.

[10] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *JEPIN*, vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com

[11] W. Irmayani, “Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Naive Bayes,” *J. Khatulistiwa Inform.*, vol. 9, no. 1, 2021, [Online]. Available: www.bsi.ac.id

[12] W. Yulita, E. D. Nugroho, and M. H. Algifari, “Sentiment Analysis on Public Opinion About the Covid-19 Vaccine Using the Naïve Bayes Classifier Algorithm.” Jdmsi, 2021.

[13] A. E. Sari, S. Widowati, and K. M. Lhaksmana, “Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier,” *e-Proceeding Eng.*, vol. 6, no. 2, 2019.

[14] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021.

[15] D. Rusdiaman and D. Rosiyadi, “Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naive Bayes Classifier Dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 2502–7131, 2019.