

ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAIVE BAYES CLASSIFIER

Andriani Nurian¹, Muhammad Samsul Ma'arif², Indira Nur Amalia³, Chaerur Rozikin⁴

^{1,2,3,4} Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

Riwayat artikel:

Received: 10 Oktober 2022

Accepted: 29 Desember 2023

Published: 1 Januari 2024

Keywords:

Analisis Sentimen, Ulasan, Shopee, Naïve Bayes Classifier

Correspondent Email:

andrianinurian01@gmail.com

Abstrak. Perkembangan teknologi dalam era digital ini sangat pesat. Salah satu bidang yang mengalami perkembangan signifikan adalah industri keuangan (fintech). Banyak perusahaan fintech, termasuk Shopee, yang muncul dan menawarkan kemudahan kepada pengguna. Dalam konteks ini, Google Play Store merupakan platform yang memungkinkan pengguna memberikan ulasan terhadap produk yang mereka gunakan. Seiring dengan pertumbuhan jumlah pengguna, jumlah ulasan di Google Play Store juga meningkat. Ulasan dari pengguna dapat menjadi sumber informasi berharga bagi perusahaan untuk melakukan perbaikan dan peningkatan di masa depan. Penelitian ini berfokus pada analisis sentimen terhadap ulasan pengguna aplikasi Shopee di Google Play Store menggunakan algoritma Naïve Bayes. Data ulasan diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Evaluasi penelitian ini menunjukkan hasil Accuracy sebesar 87,58%, presisi sebesar 91,20%, dan recall sebesar 90,21%. Penelitian ini memberikan kontribusi dalam pemahaman sentimen pengguna terhadap aplikasi Shopee. Dengan menggunakan Naïve Bayes Classifier dan metode seleksi fitur TF-IDF, perusahaan dapat mengklasifikasikan ulasan pengguna dengan lebih efektif. Hasil penelitian ini dapat digunakan sebagai dasar untuk meningkatkan pengalaman pengguna dan kualitas layanan aplikasi Shopee.

Abstract. The rapid development of technology in the digital era has brought significant advancements in various fields, including the financial industry, commonly known as fintech. Many fintech companies, including Shopee, have emerged, offering convenience to users. In this context, Google Play Store serves as a platform where users can provide reviews for the products they have used. As the number of users grows, the quantity of reviews on Google Play Store also increases. User reviews serve as valuable sources of information for companies to make improvements and enhancements in the future. This research focuses on sentiment analysis of user reviews for the Shopee application on Google Play Store using the Naïve Bayes algorithm. The review data is classified into three categories: positive, negative, and neutral. The evaluation of this research yields an accuracy of 87.58%, precision of 91.20%, and recall of 90.21%. This study contributes to understanding user sentiment towards the Shopee application. By utilizing the Naïve Bayes Classifier and the TF-IDF feature selection method, companies can effectively classify user reviews. The findings of this research can serve as a basis for enhancing user experience and improving the quality of Shopee application services.

1. PENDAHULUAN

Perkembangan e-commerce di Pada era digital yang semakin berkembang saat ini, aplikasi mobile telah menjadi bagian penting dalam kehidupan sehari-hari pengguna smartphone. Salah satu jenis aplikasi mobile yang sangat populer adalah aplikasi e-commerce yang memungkinkan pengguna untuk melakukan berbagai transaksi online dengan mudah [1]. Dalam konteks ini, Shopee, sebuah platform e-commerce terkemuka di Asia Tenggara, telah berhasil mendapatkan perhatian yang signifikan dari pengguna smartphone.

Pada saat ini terdapat berbagai macam aplikasi yang dapat menawarkan layanan e-commerce. Shopee adalah salah satu platform e-commerce yang terkenal dan dihormati di Indonesia karena reputasinya yang baik dalam industri tersebut [2]. Berdasarkan data riset yang dilakukan oleh iPrice pada tahun 2017 bahwa Shopee masuk dalam 10 besar e-commerce yang paling sering dikunjungi dan diakses dengan data kunjungan sebanyak 25.736.356 oleh pengguna internet.

Dalam ekosistem aplikasi mobile, pengguna seringkali membagikan pengalaman mereka dengan menggunakan aplikasi melalui ulasan dan penilaian di platform seperti Google Play Store. Data yang terkandung dalam ulasan tersebut dapat menjadi sumber informasi berharga bagi perusahaan untuk memahami persepsi dan kepuasan pengguna terhadap aplikasi mereka [3]. Oleh karena itu, penting untuk menganalisis sentimen pengguna terhadap aplikasi Shopee dalam upaya untuk meningkatkan kualitas produk dan meningkatkan kepuasan pengguna.

Review atau ulasan dari pengguna aplikasi seperti halnya *Shopee* umumnya berisi saran yang bersifat positif maupun keluhan yang bersifat negatif yang dituliskan dapat secara tidak langsung baik sedikit maupun banyak, akan berpengaruh pada calon pelanggan. Untuk menyortir dan memantau ulasan tersebut bukan hal yang ringan karena jumlah ulasan yang dimuat dalam media sosial umumnya jumlahnya sangat banyak apabila diproses secara manual. Oleh karena itu dibutuhkan suatu metode yang dapat menyortir dan memantau ulasan tersebut secara cepat dan otomatis dalam mengkategorikan ulasan-ulasan tersebut baik yang positif maupun yang bersifat

negatif [4]. Salah satu fitur yang tersedia di Play Store adalah kemampuan pengguna untuk memberikan peringkat dan ulasan tentang aplikasi yang mereka gunakan. Sejauh ini, penghargaan "aplikasi terbaik" sering diberikan kepada aplikasi dengan jumlah unduhan dan peringkat bintang tertinggi di Play Store. Namun, semakin banyaknya ulasan pengguna aplikasi Shopee yang masuk ke Google Play membuat sulit bagi perusahaan untuk mendapatkan pemahaman menyeluruh dari semua ulasan tersebut. Salah satu cara yang cocok untuk mengumpulkan informasi dari ulasan tersebut adalah dengan menggunakan teknik scraping. Scraping merupakan teknik yang memungkinkan pengembangan bentuk baru dalam pengumpulan data, analisis, dan visualisasi [5].

Naïve Bayes adalah metode yang berdasarkan asumsi bahwa atribut-atribut memiliki ketergantungan yang saling bebas secara kondisional terhadap nilai output [3]. Penggunaan Naïve Bayes memiliki keuntungan karena hanya membutuhkan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter yang dibutuhkan dalam proses klasifikasi. Dalam banyak situasi dunia nyata yang kompleks, Naïve Bayes sering memberikan hasil yang lebih baik daripada yang diharapkan. Dalam proses text mining, bahasa pemrograman Python digunakan untuk melakukan langkah-langkah mulai dari preprocessing data hingga klasifikasi dalam analisis sentimen menggunakan algoritma Naïve Bayes. Bahasa pemrograman Python dipilih karena mudah dalam penggunaan dan pengaplikasiannya [6]. Oleh sebab itu diperlukan metode untuk menyortir serta menganalisis ulasan dengan cepat dan akurat serta mengkategorikan antara ulasan positif, negatif, dan netral.

Dalam konteks ini, metode analisis sentimen telah menjadi populer dalam menganalisis opini, emosi, dan sikap pengguna terhadap berbagai produk atau layanan. Metode yang telah terbukti efektif adalah klasifikasi sentimen menggunakan Naïve Bayes Classifier [8]. Naïve Bayes Classifier adalah salah satu algoritma klasifikasi yang populer dalam analisis sentimen karena kemampuannya untuk mengklasifikasikan teks ke dalam kategori sentimen yang berbeda, seperti positif, negatif, atau netral [7].

Dalam penelitian ini, kami akan memaparkan metodologi yang digunakan dalam analisis sentimen, mengumpulkan data ulasan pengguna aplikasi Shopee, melatih dan menguji Naïve Bayes Classifier, serta menginterpretasikan hasil analisis sentimen untuk mendapatkan wawasan yang berarti. Diharapkan bahwa penelitian ini dapat memberikan sumbangan yang signifikan dalam bidang analisis sentimen aplikasi mobile dan memberikan pemahaman yang lebih baik tentang persepsi pengguna terhadap aplikasi Shopee.

2. TINJAUAN PUSTAKA

Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat [8]. Analisis sentimen digunakan sebagai gambaran umum media sosial untuk mengetahui pengetahuan perasaan lebih cenderung opini positif atau opini negatif [9].

Sistem rating aplikasi Google Play Store didasarkan pada ulasan pengguna terhadap masing-masing aplikasi. Ulasan yang dimaksud berisi tanggapan pengguna terhadap aplikasi yang bersangkutan, dan pengguna dapat menggunakan ulasan suatu aplikasi sebagai peringatan awal sebelum memutuskan untuk menggunakannya [10].

Shopee kini menjadi satu-satunya aplikasi e-commerce yang banyak digunakan di Indonesia. Shopee dirancang untuk membuat transaksi bisnis online menjadi cepat, efisien, dan praktis. Dengan kemudahan dalam menggunakan transaksi online, maka masyarakat umum akan lebih mudah dalam mengoperasikan e-commerce. Memanfaatkan teknologi adalah satu-satunya faktor terpenting dalam membuat penggunaan aplikasi Shopee menjadi sederhana. Sehingga dapat memudahkan pelanggan dalam melakukan transaksi online menggunakan Shopee [11].

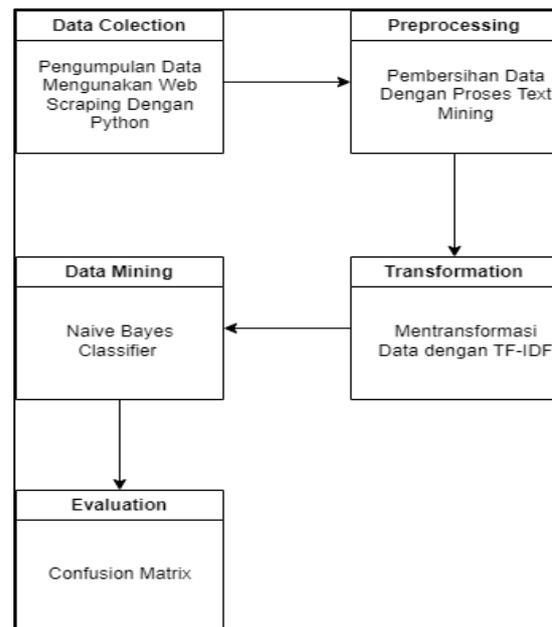
3. METODE PENELITIAN

Penelitian ini dimulai dengan melakukan web scraping pada play store dengan menggunakan python lalu memasukkan link aplikasi Shopee. Web scraping pada Play Store dengan menggunakan Python adalah proses

mengumpulkan data dari halaman-halaman aplikasi di Google Play Store secara otomatis. Dalam konteks ini, data yang ingin diambil adalah informasi tentang aplikasi, ulasan pengguna, peringkat, dan atribut lainnya yang terkait dengan aplikasi yang ada di Play Store [12].

Metode penelitian yang digunakan dalam penelitian ini adalah Knowledge Discovery from Data. Tahapan KDD atau Proses Penemuan Pengetahuan dari Data adalah serangkaian langkah-langkah sistematis yang digunakan untuk mengubah data mentah menjadi pengetahuan yang berarti dan bermanfaat [13].

KDD sendiri memiliki 5 tahapan, dengan tahapan didalamnya yakni Data Selection, Preprocessing, Transformation, Data Mining, Evaluation. Pada tahapan Preprocessing prosesnya menggunakan proses Text Mining untuk membersihkan datanya [14]. Agar dapat lebih mudah keterangan dari alur penelitian terdapat pada gambar 1.



Gambar 1. Alur Penelitian

3.1 Data Selection

Data dikumpulkan dari google play store, mengenai ulasan pengguna aplikasi Shopee. Dataset yang terkumpul 3000 data. Kemudian data dibagi menjadi negatif 2.525 data, positif 276 data, netral 199 data.

3.2 Preprocessing

Tahap *text preprocessing* adalah tahap awal dari *text mining*. *Text Preprocessing* merupakan proses seleksi data yang akan diproses pada setiap dokumen. Tahap *Preprocessing* diperlukan untuk membersihkan data dari kata yang tidak terlalu perlu. Dalam tahap *preprocessing* pada penelitian ini terdiri dari *cleaning*, normalisasi dan pemberian label (*labeling*).

3.3 Text Transformation

Text transformation merupakan tahap ketiga dalam *text mining* yang bertujuan untuk membuat data yang berupa *text* siap di proses data mining, tahap *text transformation* pada penelitian ini terdiri dari *transform case*, *Tokenizing*, *filtering*, *stop words & stemming* [15].

3.4 Data Mining

Naïve Bayes Classifier Pada tahap klasifikasi menggunakan Algoritma *Naïve Bayes*, dibagi menjadi dua proses yaitu proses *training* dan proses *testing*. Pertamadilakukan dulu proses *training* untuk pelatihan, kemudian dilakukan *prosestesting* dengan mengacu pada probabilitas dari dataset *training*. Metode *Naïve Bayes Classifier* merupakan salah satu metode yang digunakan untuk melakukan pengelompokkan suatu opini dengan baik. *Naïve Bayes Classifier* mampu mengelompokkan komentar orang-orang menjadi positif, netral, ataupun negatif terhadap suatu produk atau isu yang sedang berkembang di khalayak ramai [8].

3.5 Evaluasi

Tahap evaluasi adalah tahap yang dilakukan untuk melihat seberapa baik performansi algoritma klasifikasi yang digunakan dalam penelitian. Tolok ukur yang digunakan untuk mengukur performansi tersebut adalah *accuracy*, *precision*, *recall*, dan *f-measure* [16]. Untuk memberikan gambaran kinerja yang lebih jelas, matriks konfusi memungkinkan analisis di empat dimensi:

1. Accuracy : Untuk mengetahui tingkat kemiripan antara nilai prediksi dengan nilai aktual.
2. Precision : Untuk mengetahui seberapa tepat atau akurat model dari yang diprediksi

positif. Precision juga merupakan ukuran yang baik untuk menentukan ketika nilai False Positive pada suatu model tinggi.

3. Recall : Untuk menghitung berapa banyak dari nilai Actual Positive yang berhasil diidentifikasi oleh model melalui pelabelan sebagai True Positive. Recall juga akan menjadimetrik model yang digunakan untuk memilih model terbaik ketika ada nilai yang tinggi terkait dengan FalseNegative.

4. F-Measure : Perbandingan rata-rata nilai precision dan recall yang dibobotkan.

4. HASIL DAN PEMBAHASAN

Hasil dan pembahasan untuk analisis sentimen pada aplikasi Shopee menggunakan *Naïve Bayes Classifier* yang diolah oleh peneliti menggunakan Google Collab, contoh data *scraped*, berlabel, *preprocessing* dan nilai *Accuracy* untuk masing-masing disajikan pada sub bab selanjutnya.

4.1 Scraping data

Hasil dari *scraping* data yang telah dilakukan dengan bobot data pencarian sebanyak 3000 data pada aplikasi *Shopee* dapat dilihat pada gambar dibawah ini dengan menggunakan script.

Scraping

```
auth =
tweepy.OAuthHandler(access_key,
access_secret)
auth.set_access_token(access_token,
access_token_secret)
api = tweepy.API(auth,
wait_on_rate_limit=True)

def scraptweets(search_words,
date_since,date_untill):
```

Scraping

```

db_tweets =
pd.DataFrame(columns=['username',
'tweetcreatedts', 'text'])
tweets = tweepy.Cursor(
    api.search_tweets, q=search_words,
    lang="id", since_id=date_since,
    until=date_untill,
tweet_mode='extended').items(3000
) tweet_list = [tweet for tweet in
tweets]for tweet in tweet_list:
    username =
    tweet.user.screen_name
    tweetcreatedts =
    tweet.created_at try:
        text =
        tweet.retweeted_status.full_text
    except AttributeError:
        text = tweet.full_text

    ith_tweet = [username, tweetcreatedts,
text] db_tweets.loc[len(db_tweets)] =
    ith_tweet print('Proses Crawling Selesai
dengan jumlah
data', len(db_tweets))
    filename = 'Data Shopee.csv'
    db_tweets.to_csv(filename,
    index=False)

today = datetime.date.today().strftime("% Y-
% m-
% d")
last_week = datetime.date.today() -
timedelta(360)last_week =
last_week.strftime("% Y-% m-% d")

search_words =
"Shopee" date_since =
"2023-04-15"
date_untill = "2023-05-15"

scraptweets(search_words, date_since,
date_untill)
    
```

Berikut hasil outputnya

	username	tweetcreatedts	text
0	bbbbbb81575075	2023-05-14 17:03:55+00:00	Ayo main Shopee Candy untuk memenangkan berbag...
1	kembaranyaya41	2023-05-14 17:03:36+00:00	Temukan HJAB SEGIEMPAT VOAL MOTIF LASERCUT Ad...
2	ilyjwoooo	2023-05-14 17:03:34+00:00	wib jungwoo pink christmas \n\prefer shopee l...
3	ao4hi	2023-05-14 17:03:26+00:00	@teumemart ayo di take kakk! 🍷 tangsel, bisa ke...
4	Shopee_Link	2023-05-14 17:03:26+00:00	🔗 https://t.co/gF58z8MjTrin 🇮🇩 5in 🌊 ocean blue...
...
2995	mustha_ve	2023-05-14 11:11:38+00:00	Dispenser plus penghangat minuman\n\nLink Shop...
2996	mustha_ve	2023-05-14 11:11:30+00:00	Rak buku/baju/sepatu/ serbaguna\n\nLink Shopee...
2997	AripUdI90116809	2023-05-14 11:11:25+00:00	@yougotm03005651 Need Gestun? ke @sonianayn a...
2998	mustha_ve	2023-05-14 11:11:22+00:00	Mesin cuci + pengering baju portable \n\nLink ...
2999	hyunshioo_12	2023-05-14 11:11:18+00:00	Wts photocard mashiho dan jeongwoo full album ...

Gambar 2. Hasil Crawling Data

4.2 Pelabelan Otomatis

Setelah pengumpulan data peneliti selanjutnya melakukan pelabelan pada aplikasi *Shopee*. Berikut adalah hasil pelabelan dengan *script* yang tertera pada gambar.

Pelabelan Data

```

tweet_polarit
y = []
tweet_weigh
t = [] negasi
= False

for sentence in
df['Tweet']:
    sentence_score = 0
    sentence_weight =
    "" sentiment_count
    = 0 sentence =
    sentence.split()for
    word in sentence:
        try:
            score = lexicon[word]
            sentiment_count =
            sentiment_count + 1 except:
                score = 99

    if(score == 99):
        if (word in negative_words):
            negasi = True
            sentence_score = sentence_score + 0
            sentence_weight = sentence_weight +
            " + "+
            str(0)
        else:
            sentence_score = sentence_score +
            0 sentence_weight =
            sentence_weight + " +
            "+
            str(0)
        )
    
```


Frequency Inverse Document Frequency) dengan memberikan bobot pada masing-masing kata untuk menyeleksi fitur yang dihasilkan. Selanjutnya adalah *data mining* dengan melakukan klasifikasi data teks ke dalam sentimen positif dan negatif menggunakan *Naïve Bayes Classifier*, namun sebelumnya data teks akan dibagi berdasarkan data latih dan data uji. Setelah tahap *data mining* akan dilakukan evaluasi terhadap hasil klasifikasi dengan menghitung nilai *accuracy*, *precision*, *recall* dan *f-measure* dari *confusion matrix*. Evaluasi dilakukan dengan membandingkan hasil klasifikasi menggunakan seleksi fitur TF-IDF.

2. Gambaran atau informasi umum mengenai data ulasan pengguna aplikasi Shopee diperoleh dari visualisasi dengan menggunakan *word cloud*. Kata “Shopee”, “harga”, “dapat”, “potong”, dan “Shopeeid” menjadi kata yang paling sering muncul yang digunakan untuk ulasan aplikasi Shopee pada penelitian ini.

3. Hasil *Accuracy* tertinggi adalah klasifikasi menggunakan *Naïve Bayes Classifier* dengan seleksi fitur TF-IDF yaitu sebesar 77%, presisi 26%, *recall* 33%, dan *f1-score* 29%, data *testing* yang digunakan sebanyak 300 data atau 10% dari 3000 data, dari jumlah data yang digunakan dengan metode random pada saat *testing*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] L. D. U. Rahmat Hidayat, Siti Marlina, “Perancangan Sistem Informasi Sirkulasi Buku Pada Perpustakaan SMP Negeri 103 Jakarta,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 4, no. 2, p. 227, 2019, doi: 10.30998/string.v4i2.4179.

[2] Budi Raharjo, *Fintech Teknologi Finansial Perbankan Digital Dr. Budi Raharjo*. 2008.

[3] J. Mansur, V. C. Mawardi, and T. Sutrisno, “Analisis Pendapat Publik Terhadap Public Figure,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 2, pp.

185–190, 2020.

[4] and Z. A. O. M. Sadegh, R. Ibrahim, “A survey of opinion mining and sentiment analysis,” *Min. Text Data*, vol. 9781461432, no. 3, pp. 415–463, 2012, doi: 10.1007/978-1-4614-3223-4_13.

[5] R. Baskara and F. Rahma, “Implementasi Web Scraping Pada Media Sosial Instagram,” *Automata*, vol. 3, pp. 1–3, 2022.

[6] A. Imron, “Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,” 2019.

[7] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[8] Y. U. Titan Nugraha, Purwantoro, “Analisis Sentimen terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma Naïve Bayes,” *J. Pendidik. dan Konseling*, vol. 4, pp. 1349–1358, 2022.

[9] D. A. Wulandari, R. Rohmat Saedudin, and R. Andreswari, “Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada R UU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.

[10] P. Aditiya, U. Enri, and I. Maulana, “Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, 2022, doi: 10.30865/jurikom.v9i4.4673.

[11] Vynska Amalia Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, pp. 141–151, 2020.

[12] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, “Web scraping technologies in an API world,” *Brief. Bioinform.*, vol. 15, no. 5, pp. 788–797, 2013, doi: 10.1093/bib/bbt026.

[13] F. Septianingrum, J. H. Jaman, and U. Enri, “Analisis Sentimen Pada Isu

- Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1431, 2021, doi: 10.30865/mib.v5i4.3260.
- [14] A. Aisyah and S. Anraeni, “Analisis Penerapan Metode K-Nearest Neighbor (K-NN) pada Dataset Citra Penyakit Malaria,” *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 17–29, 2022, doi: 10.56705/ijodas.v3i1.22.
- [15] Ryan, Cooper, and Tauer, “Peningkatan Kompetensi Kelulusan,” *Pap. Knowl. . Towar. a Media Hist. Doc.*, pp. 12–26, 2013.