SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK IV - Surabaya, 27 April 2024
Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2024.5883

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Analisis Sentimen Ulasan Aplikasi Genshin Impact di Play Store Menggunakan Random Forest

Dionisius Reinaldo A.*, Yosefina Finsensia Riti

Universitas Katolik Darma Cendika

*e-mail: *dionisius951@gmail.com*

ABSTRACT

Sentiment analysis and text classification are crucial techniques in understanding users' opinions and feedback on a particular product or service. In the context of gaming, such as Genshin Impact, assessing user sentiment plays a crucial role in improving the quality and gaming experience. This research aims to apply these techniques to the reviews of Genshin Impact obtained from the application distribution platform, namely Play Store. A text classification model, Random Forest, is evaluated to classify the sentiment of reviews. Data collection is conducted using web scraping techniques to extract reviews directly from the Play Store. The review data then undergo a series of pre-processing steps, including labeling, case folding, removing stop words, tokenizing, and stemming, to clean and prepare the data. The next process involves word weighting using the TF-IDF method to evaluate the significance of words in the text. The research results show good performance from both models, with Random Forest achieving 92% accuracy.

Keywords*: Game; Genshin Impact; Random Forest; sentiment analysis; text classification.*

ABSTRAK

Analisis sentimen dan klasifikasi teks adalah teknik yang sangat penting dalam memahami opini dan tanggapan pengguna terhadap suatu produk atau layanan tertentu. Dalam konteks *game*, seperti Genshin Impact, penilaian sentimen pengguna memiliki peran yang krusial dalam meningkatkan kualitas serta pengalaman bermain. Penelitian ini bertujuan untuk menerapkan teknik-teknik tersebut pada ulasan-ulasan Genshin Impact yang diambil dari platform distribusi aplikasi, yaitu Play Store. Model klasifikasi teks, yaitu Random Forest, dievaluasi untuk mengklasifikasikan sentimen ulasan. Metode pengumpulan data dilakukan melalui teknik web scrapping untuk mengambil ulasan langsung dari Play Store. Data ulasan kemudian melalui serangkaian tahapan *pre-processing*, termasuk labeling, *case folding*, penghapusan *stop words*,

tokenizing, dan stemming, untuk membersihkan dan mempersiapkan data. Proses selanjutnya melibatkan pembobotan kata menggunakan metode TF-IDF untuk mengevaluasi signifikansi kata dalam teks. Hasil penelitian menunjukkan kinerja yang baik dari model dengan Random Forest mencapai akurasi 92%.

Kata kunci: Analisis sentimen; *game*; Genshin Impact; klasifikasi teks; Random Forest

PENDAHULUAN

Game adalah bentuk hiburan interaktif yang dimainkan melalui berbagai platform seperti komputer, konsol *game*, atau perangkat mobile. Dalam permainan, pemain berinteraksi dengan dunia virtual yang dibuat oleh pengembang untuk mencapai berbagai tujuan, seperti menyelesaikan misi, memecahkan teka-teki, atau berkompetisi dengan pemain lain. Salah satu *game* yang sangat populer dan menarik minat dari banyak pemain adalah Genshin Impact, *game* ini mempunyai ulasan yang beragam dari berbagai pemain. Genshin Impact merupakan *game* open-world action role-playing *game* (RPG) yang dikembangkan oleh miHoYo, sebuah perusahaan *game* yang berlokasi di Tiongkok. Permainan ini pertama kali dirilis pada 28 September 2020 dan dapat dimainkan di berbagai platform termasuk PC, PlayStation, iOS, dan Android.

Klasifikasi teks dan analisis sentimen adalah dua teknik yang digunakan dalam bidang pemrosesan bahasa alami (*natural language processing*). Klasifikasi teks merupakan suatu kegiatan untuk mengelompokkan teks ke dalam kategori tertentu yang telah ditentukan berdasarkan karakteristik tertentu, sementara analisis sentimen adalah proses menganalisis dan mengekstraksi sentimen atau pendapat dari teks yang tersedia, seperti ulasan atau komentar [1], [2] . Sebelum mengunduh *game* orang biasanya akan melihat ulasan dari *game* yang akan diunduh melalui halaman komentar yang ada di google play store sehingga sentimen analisis menjadi penting tidak hanya untuk konsumen tetapi juga untuk pengembang dengan mengetahui opini konsumen tentang pengalaman bermain mereka, pengembang dapat memanfaatkan opini konsumen untuk menjadi bahan evaluasi arah pengembangan dan peningkatan *game* selanjutnya [3], [4] . Dalam halaman komentar google play store data yang muncul beragam sehingga menyebabkan data tidak konsisten dan redundan sehingga memerlukan sebuah proses *pre-processing* data untuk membuat agar data yang akan diolah menjadi konsisten sehingga akan lebih mudah untuk diolah oleh peneliti

Penelitian ini bertujuan untuk menerapkan metode klasifikasi teks dan analisis sentimen pada ulasan Genshin Impact yang diperoleh dari platform distribusi aplikasi Play Store. Dalam penelitian yang dilakukan oleh Evita Fitri pada tahun 2020, yang membandingkan performa algoritma Naïve Bayes, Random Forest, dan Support Vector Machine untuk analisis sentimen aplikasi Ruang Guru, hasilnya menunjukkan bahwa Random Forest memiliki performa terbaik, yakni mencapai 97%. Oleh karena itu, dalam penelitian ini akan diuji performa Random Forest untuk menganalisis sentimen ulasan permainan Genshin Impact di Google Play Store. Tujuan dari penelitian ini adalah untuk mendapatkan pemahaman yang lebih dalam tentang tanggapan dan pengalaman pengguna terhadap Genshin Impact, serta memberikan masukan yang berharga bagi pengembang *game* untuk meningkatkan kualitas dan pengalaman bermain *game* tersebut.

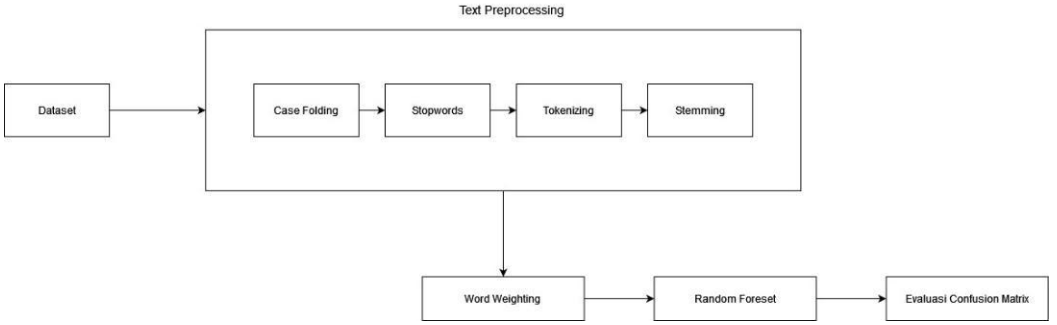
METODE

Penelitian ini dilakukan dengan langkah-langkah sesuai dengan urutan seperti pada diagram Gambar 1.

Dataset

Dalam penelitian ini *dataset* dikumpulkan dengan teknik *scrapping*. *Scrapping* adalah teknik yang digunakan untuk mengumpulkan data dari situs web (dalam penelitian ini ulasan Play Store) dengan menggunakan *script* untuk kemudian data tersebut dapat diolah dan disimpan dalam file seperti CSV. Data yang digunakan dalam penelitian ini diambil melalui link

com.miHoYo.GenshinImpact yang berisi ulasan Google Play Store *game* Genshin Impact dengan 1.221 data ulasan Google Play Store dan *filter* ulasan paling relevan.



Gambar 1. Diagram Penelitian

Text Pre-processing

1. *Case Folding*
   Merupakan teknik untuk mengubah huruf dalam teks menjadi huruf kecil sehingga dapat menjaga konsistensi dalam pengolahan kata dalam teks [5].
2. *Stop Words*
   Merupakan teknik penghapusan kata yang tidak memiliki nilai penting dalam teks.
3. *Tokenizing*
   Merupakan teknik untuk memecah-mecah kalimat menjadi bagian-bagian yang lebih kecil yaitu kata.
4. *Stemming*
   Merupakan teknik untuk menghilangkan kata imbuhan di awal atau di akhir kalimat sehingga menjadi kata dasar tanpa imbuhan.

Word Weighting

Merupakan teknik dalam analisis teks untuk mengevaluasi seberapa penting suatu kata dalam teks [6]. Dalam penelitian ini menggunakan teknik TF-IDF, metode ini digunakan dengan tujuan untuk memberikan nilai terhadap kata-kata kunci dalam teks atau dokumen. Dalam tahapan ini terdapat dua proses yang dilakukan yaitu TF (*Term-Frequency*) yaitu menghitung jumlah kemunculan kata dalam kalimat semakin banyak kata muncul maka nilai TF akan semakin tinggi dan IDF (*Inverse Document Frequency*) yang berfokus pada menghitung nilai pada kata yang jarang muncul dalam kalimat atau dokumen [7].

Berikut merupakan rumus yang digunakan untuk menghitung TF-IDF [7].

$$W(t,d) = tf(t,d) \times idf(t,d) = tf\ (t,d)\ \times\ \log\frac{N}{d\ f(t)}$$

$w\ (t,\ d\ ) = Bobot\ Nilai\ TF - IDF$
$tf\ (t,\ d\ ) = Jumlah\ Frekuensi\ Kata$
$id\ (t,\ d\ ) = Jumlah\ Inverse\ Frekuensi\ Kata$
$df\ (t\ ) = Jumlah\ frekuensi\ dokumen\ tiap\ kata$
$N = Jumlah\ total\ dokumen$

Implementasi Model

Dalam penelitian ini akan menggunakan algoritma Random Forest. Random Forest adalah sebuah metode dalam *machine learning* yang menggunakan *decission tree* untuk melakukan klasifikasi, *bootstrap sampling* digunakan untuk membentuk pohon prediksi dan masing-masing

pohon prediksi memiliki prediktor acak kemudian untuk memperoleh hasil akhir akan menggunakan *voting majority* [8], [9] . Setelah dilakukan pembersihan pada dataset yang digunakan pada tahap *pre-processing* sebelumnya kemudian dilakukan pengujian terhadap dua algoritma Random Forest untuk kemudian diteliti kemampuan model dalam menganalisis teks. Setelah dilakukan pengujian terhadap model kemudian akan dilakukan proses penghitungan akurasi dengan *confusion matrix*. *Confusion Matrix* memiliki 4 kelas yang mewakili hasil klasifikasi. TP (*True Positive)* mewakili data positif yang diklasifikasi sebagai data positif, FP (*False Positive)* mewakili data negatif namun diklasifikasikan sebagai data positif. Sedangkan TN (*True Negative*) mewakili data negatif yang diklasifikasikan sebagai data negatif dan FN (*False Negative)* mewakili data positif yang diklasifikasikan sebagai data negatif [7].

Berikut merupakan rumus penghitungan Akurasi, Presisi, Recall, dan F1-Score [10].

$$Akurasi = \frac{TP + TN}{(TP+TN+FP+FN)}$$

$$Presisi = \frac{TP}{(TP+FP)}$$

$$Recall = \frac{TP}{(TP+FN)}$$

$$F1 - score = 2\frac{(Precision * Recall)}{(Precision + Recall)}$$

$$support = TP + FN$$

HASIL DAN PEMBAHASAN

Scrapping Data

Data diambil dengan menggunakan teknik *web scrapping*. *Web scrapping* digunakan untuk mengumpulkan data dari Play Store berupa 1221 data ulasan pengguna tentang *game* Genshin Impact. Data diambil menggunakan bantuan Google Colab dengan filter ulasan paling relevan dengan atribut yang digunakan adalah username, rating, tanggal review, dan komentar. Data yang telah dikumpulkan disimpan dalam file dengan format CSV untuk kemudian dilakukan pengolahan data di tahap *pre-processing* [11].

Pre-processing

Data yang telah dikumpulkan kemudian diolah dengan memilah dan membuang kata yang tidak memiliki makna penting sehingga data yang diolah menjadi lebih konsisten. Berikut merupakan tahapan-tahapan yang dilakukan untuk mengolah data yang telah dikumpulkan.

Labeling

Pada tahapan ini dilakukan pemberian label pada data untuk mengetahui komentar dalam ulasan mengandung makna positif atau negatif [12]. Pada penelitian ini data dilabeli berdasarkan *rating*-nya. Data dengan rating kurang dari 3 diberi label negatif sedangkan rating lebih dari 3 diberi label positif. Rating dengan angka sama dengan 3 akan diberi label netral [6].

Case Folding

Pada tahap ini akan menghilangkan simbol dan mengubah huruf besar menjadi huruf kecil [12]. Pada tabel berikut ditampilkan sample hasil case folding yang telah dilakukan pada data ulasan *game* Genshin Impact.

Tabel 1. *Sample Case Folding*

Sebelum	Sesudah
Kikir impact	kikir impact
Lumayan, tpi masih belom faham alur cerita	lumayan tpi masih belom faham alur cerita

Stop Words

Pada tahap ini akan dilakukan penghapusan kata yang tidak memiliki makna penting dalam sebuah kata [13]. Berikut merupakan sample data hasil penghapusan kata yang memiliki makna tidak penting dalam kalimat ulasan *game* Genshin Impact.

Tabel 2. *Sample Stop Words*

Sebelum	Sesudah
ame nya bagus waifu nya banyak dan loli nya juga imut	game nya bagus waifu nya loli nya imut
walau gak pernah menang rate off tetapi terima kasih telah memberikan nahida untukku akan ku jaga dia	gak menang rate off terima kasih nahida untukku ku jaga

Tokenizing

Tahap ini merupakan tahapan untuk memecah kalimat menjadi beberapa kata. Berikut merupakan sample hasil tokenizing pada data ulasan *game* Genshin Impact.

Tabel 3. *Sample Tokenizing*

Sebelum	Sesudah
kikir impact	[kikir, impact]
good game	[good,game]

Stemming

Tahap ini merupakan tahapan untuk mengubah kata yang memiliki imbuhan menjadi kata dasar. Berikut merupakan sample hasil stemming pada data ulasan *game* Genshin Impact. Dalam penelitian ini menggunakan library python sastrawi, sastrawi adalah library yang memungkinkan untuk mengurangi kata berimbuhan dalam bahasa Indonesia menjadi bentuk ke dasarnya (stema).

Tabel 4. *Sample Stemming*

Sebelum	Sesudah
game kikir bug sinyal merah kuning event membosankan update makan gbgb an peminatnya reward nya dikit mengecewakan	game kikir bug sinyal merah kuning event bosan update makan gbgb an minat reward nya dikit kecewa

Weighting

Setelah data yang digunakan telah dibersihkan kemudian dilakukan pembobotan kata dengan tujuan untuk memberikan nilai kepada kata-kata kunci dalam dokumen. Dalam penelitian ini menggunakan teknik TF-IDF. Tabel di bawah ini merupakan hasil pembobotan kata dengan TF-IDF.

Tabel 5. Hasil TF-IDF

No.	Vectorize	Bobot
1	(0, 2520)	0.1154
2	(0, 2489)	0.2286
3	(0, 2400)	0.2568
4	(0, 2204)	0.1970
5	(0, 1904)	0.2568
dst.	dst.	dst.

Implementasi Model

Data ulasan yang telah dibersihkan akan dibagi ke dalam data *train* dan *test* [14]. Dalam penelitian ini menggunakan rasio 80 banding 20 sehingga data training berjumlah 976 data dan data test berjumlah 245 data. Model Random Forest dan dilatih dengan data training yang sama kemudian dilakukan pengujian dengan data test dan pada penelitian ini menggunakan matriks evaluasi seperti akurasi, presisi, recall, dan F1-Score untuk mengetahui kinerja model. Berikut merupakan hasil evaluasi dari dua model dengan *dataset* yang diberikan.

Tabel 6. Matriks Evaluasi Random Forest

	Positif	Netral	Negatif
Positif	60	2	0
Netral	0	85	6
Negatif	12	0	80

Berdasarkan hasil matriks evaluasi kemudian ditentukan nilai presisi, akurasi, *recall*, dan F1-*score* dan *support*. Berikut merupakan nilai akurasi, presisi, recall, dan F1-*score*, dan *support* pada masing-masing model.

$$Akurasi = \frac{60+85+80}{(60+2+0+0+83+8+12+0+80)} = 0.902\ atau\ 90{,}2\%$$

Kelas Positif :

$$Presisi = \frac{60}{(12+0)} = 0.833\ atau\ 83\%$$

$$Recall = \frac{60}{(60+2)} = 0.96\ atau\ 96\%$$

$$F1-score = 2\frac{(83\%*96\%)}{(83\%+\ 96\%\ )} = 89\%$$

$$support = 60+12 = 72$$

Kelas Netral :

$$Presisi = \frac{85}{(85+2\ )} = 0.97\ atau\ 97\%$$

$$Recall = \frac{85}{(85+0)} = 1\ atau\ 100\%$$

$$F1-score = 2\frac{(97\%*100\%)}{(97\%+\ 100\%\ )} = 98\%$$

$$support = 85+0 = 85$$

Kelas Negatif :

$Presisi = \frac{80}{(80+6)} = 0.93\ atau\ 93\%$

$Recall = \frac{80}{(80+12)} = 87\%$

$F1-score = 2\frac{(93\%*87\%)}{(93\%+87\%)} = 89\%$

$support = 80 + 6 = 86$

Tabel 8. Akurasi Model

Model	Akurasi
Random Forest	90,2%

Tabel 9. Hasil *Confusion* Matrix Random Forest

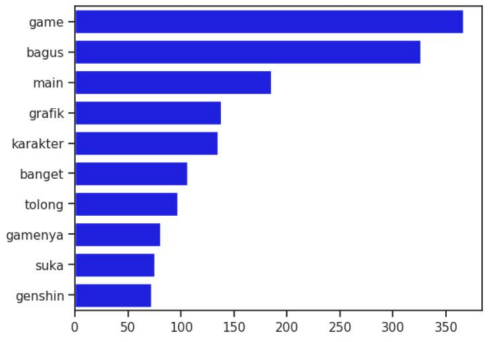
Kelas	Presisi (%)	Recall (%)	F1-Score (%)	Support
Negatif	93	87	89	86
Netral	97	100	98	85
Positif	83	96	89	72

Selain itu juga ditampilkan kata yang paling banyak muncul baik dalam label positif, label netral, maupun label negatif menggunakan tools visualisasi dari Google Colab, yaitu *word cloud* [15]. Selain itu, pada Gambar 5 ditampilkan kata yang paling banyak muncul dalam ulasan.

Gambar 2. Ulasan Positif

Gambar 3. Ulasan Negatif

Gambar 4. Ulasan Netral



Gambar 5. Sebaran kata

Pada Tabel 10 terdapat beberapa contoh ulasan dengan label yang diprediksi salah oleh model Random Forest.

Tabel 10. Data Prediksi Salah Random Forest

No.	Ulasan	Label	
		Aktual	**Prediksi**
1	saya sangat suka game ini tapi tolong untuk update size nya terlalu besar hingga 10gb	Netral	Positif
2	ping jaringan tidak stabil susah untuk mendapatkan karkater b5 limited di karenakan susah nya untuk mendapatkan primogems dan selalu dapat di pity tinggi dan sering kalah rate off	Negatif	Positif
3	game tidak se imbang game apa ini lebih baik saya unistal saja tidak ada sedikit pun asiknya malahan kayak sampah	Negatif	Positif

KESIMPULAN

Berdasarkan hasil pembahasan yang telah dilakukan dengan data ulasan *game* Genshin Impact yang dikumpulkan dengan teknik web scrapping dengan mengambil 1221 data ulasan langsung dari Google Play Store dengan kategori ulasan paling relevan. Maka diperoleh bahwa model dapat menunjukkan kinerja yang baik dalam mengenali pola sentimen pada ulasan Genshin Impact. Hasil akurasi model Random Forest menunjukkan kinerja yang bagus dengan akurasi mencapai 90,2%. Hal ini mengindikasikan bahwa model dapat melakukan tugas analisis sentimen dengan sangat baik.

DAFTAR PUSTAKA

[1] F. Noor Hasan, "Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naïve Bayes," *Media Online)*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1555.

[2] V. Arinal and B. S. Purnomo, "Optimasi Metode Decision Tree Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Review Game GTA V Roleplay," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 457–461, 2023, doi: 10.55338/saintek.v5i1.1371.

[3] R. Kusnadi, Y. Yusuf, A. Andriantony, R. Ardian Yaputra, and M. Caintan, "ANALISIS SENTIMEN TERHADAP GAME GENSHIN IMPACT MENGGUNAKAN BERT," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 6, no. 2, pp. 122–129, Jul. 2021, doi: 10.36341/rabit.v6i2.1765.

[4] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012043.

[5] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal Teknokompak*, vol. 15, no. 1, 2021.

[6] M. N. Akbar, N. Hasanahlmar'iyah Rusydi, M. Hasrul, and S. Ramadhanti, "Sentiment Analysis Terhadap Review Aplikasi Maxim di Google Play Store Menggunakan Support Vector Machine (SVM)," vol. 2, no. 2, p. 1, 2022.

[7] J. Andre Septian, T. Maulana Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION* , vol. 1, 2019, [Online]. Available: https://t.co/9WloaWpfD5

[8] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," 2022. [Online]. Available: http://j-ptiik.ub.ac.id

[9] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *TRANSFORMTIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: www.nusamandiri.ac.id,

[10] Y. Finsensia Riti and S. S. Tandjung, "Klasifikasi Covid-19 Pada Citra CT Scans Paru-Paru Menggunakan Metode Convolution Neural Network," *Jurnal Imiah Komputer*, 2022.

[11] E. Hasibuan and E. A. Heriyanto, "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER," *JTS*, vol. 1, no. 3, 2022.

[12] A. Nursalim and R. Novita, "SENTIMENT ANALYSIS OF COMMENTS ON GOOGLE PLAY STORE, TWITTER AND YOUTUBE TO THE MYPERTAMINA APPLICATION WITH SUPPORT VECTOR MACHINE," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1305–1312, 2023, doi: 10.52436/1.jutif.2023.4.6.1059.

[13] Fitriyani and T. Arifin, "PENERAPAN WORD N-GRAM UNTUK SENTIMENT ANALYSIS REVIEW," *SISTEMASI : Jurnal Sistem Informasi*, vol. 9, 2020.

[14] M. Iqbal Ahmadi *et al.*, "SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)," *Seminar Nasional Informatika*, 2020.

[15] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.