

Konstruksi Dataset Analisis Sentimen Berbahasa Indonesia Mengenai Dampak Psikologis Platform Game Roblox pada Anak

^{1*} Alif Rama Nugroho, ²Risky Aswi Ramadhani, ³Julian Sahertian

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹[*aliframanugroho@gmail.com](mailto:aliframanugroho@gmail.com), ²risky_aswi@unpkediri.ac.id,
³juliansahertian@unpkediri.ac.id

Penulis Korespondens : Alif Rama Nugroho

Abstrak—Kurangnya *dataset* teks berbahasa Indonesia yang spesifik membahas dampak psikologis teknologi pada anak menjadi hambatan dalam pengembangan sistem deteksi dini perilaku menyimpang. Penelitian ini bertujuan membangun *dataset* analisis sentimen yang dikonstruksi dari opini publik pada platform *Youtube* mengenai *game Roblox*. Di mana saat ini sedang maraknya permasalahan pada psikis anak yang diakibatkan oleh konten-konten ataupun interaksi antar pemain dalam platform Roblox. Metodologi penelitian ini meliputi akuisisi data menggunakan teknik web *scraping*, *preprocessing* teks melalui tahap normalisasi kata tidak baku, serta pelabelan secara manual. Pada penelitian ini peneliti telah berhasil mengumpulkan 2124 data teks yang bersumber dari opini publik dari komentar platform *Youtube*. Hasil penelitian menghasilkan *dataset* terstruktur yang selanjutnya akan di validasi oleh seorang pakar psikologis untuk menguji kevalidan data guna untuk menganalisis dampak *Roblox* terhadap psikis anak. *Dataset* ini diharapkan dapat menjadi sumber referensi bagi peneliti selanjutnya dalam menganalisis dampak platform *Roblox* terhadap psikis anak serta dalam mengembangkan model komputasi untuk memantau pengaruh platform digital terhadap perkembangan psikis anak.

Kata Kunci—*Analisis Sentimen, Dataset, Machine Learning, Psikologi Anak, Roblox.*

Abstract—*The lack of Indonesian text datasets specifically addressing the psychological impact of technology on children is an obstacle in developing an early detection system for deviant behavior. This study aims to develop a sentiment analysis dataset constructed from public opinion on the YouTube platform regarding the game Roblox. Currently, there is a rise in children's psychological problems caused by content and interactions between players on the Roblox platform. This research methodology includes data acquisition using web scraping techniques, text preprocessing through the normalization stage of non-standard words, and manual labeling. In this study, researchers have successfully collected 2,124 text data sources sourced from public opinion on the YouTube comment platform. The results of the study produce a structured dataset that will then be validated by a psychology expert to test the validity of the data for analyzing the impact of Roblox on children's psychology. This dataset is expected to serve as a reference source for future researchers in analyzing the impact of the Roblox platform on children's psychology and in developing computational models to incorporate the influence of digital platforms on children's psychological development.*

Keywords—*Sentiment Analysis, Dataset, Machine Learning, Child Psychology, Roblox.*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Roblox merupakan sebuah platform yang dimana memuat banyak game hasil karya para pemainnya yang dapat dimainkan juga oleh player lain di seluruh dunia. Platform ini pertama kali

diluncurkan sejak tanggal 1 September 2006 yang dimana saat ini menjadi platform terbesar dengan pengguna aktif sejumlah 85 Juta lebih di setiap harinya [1]. Platform ini menawarkan kebebasan pada playernya untuk mencurahkan kreativitasnya dalam menciptakan dunia game yang diinginkan [2].

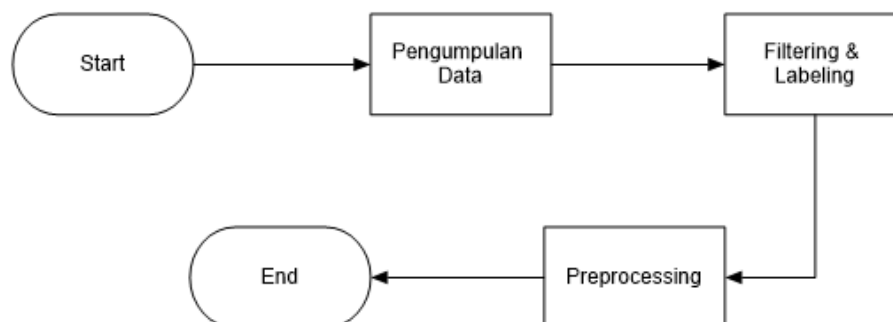
Popularitas Roblox semakin meningkat dikarenakan cepatnya media informasi mengenai game yang ada di dalamnya dan menarik minat para masyarakat terkhusus anak-anak yang merupakan target awal dari platform ini [3]. Pada awalnya target utama dari platform ini adalah anak-anak yang dimana berfungsi sebagai media dalam memberikan pengalaman sosial virtual, kreativitas, dan belajar untuk bersosialisasi [2]. Namun saat ini tidak hanya anak-anak yang memainkannya, tapi juga orang dewasa, remaja bahkan orang tua pun kini memainkannya. Bahkan kini Roblox bisa menjadi sebuah alternatif game untuk memainkan suatu game tertentu yang tidak bisa dimainkan secara bebas.

Namun platform ini tidak luput dari dampak negatif yang dimana dengan kebebasan yang diberikan banyak juga yang menyalahgunakan hal tersebut dengan membuat suatu map/game yang tidak seharusnya serta memberikan hal-hal yang kurang baik pada anak [4], sehingga dapat menimbulkan risiko psikologis pada anak seperti kecanduan [5] [6], perilaku yang buruk, dan berbagai ancaman lainnya [7] [8]. Penelitian ini berfokus pada proses pengumpulan data opini publik yang bersumber dari platform YouTube tentang dampak Roblox pada psikis anak.

Data ini akan di ambil menggunakan metode scraping yang dimana akan disimpan kedalam format database [9]. Data yang diambil merupakan komentar video YouTube yang membahas tentang dampak bermain Roblox pada psikis anak. Dari data tersebut nantinya akan diberikan label sentimen dan aspek pada tiap komentar sesuai dengan isi dari komentar tersebut yang terdiri dari sentimen positif, negatif, dan netral, serta aspek sosial, keamanan, dan kognitif pada anak. Data yang telah melalui proses labeling akan berlanjut ke proses preprocessing yang dimana berfungsi dalam membersihkan data dari noise yang dapat mengganggu proses analisis data [10] dan untuk selanjutnya data bisa digunakan dalam pembuatan model klasifikasi guna menganalisa sebuah data teks kedalam suatu sentimen dan aspek tertentu.

II. METODE

Penelitian ini mengikuti diagram alur Flowchart yang telah dibuat untuk mempermudah proses penelitian ini:



Gambar 1. Diagram Alur Pengumpulan Data

1. Pengumpulan Data: Pada proses ini, peneliti menggunakan platform *Apify* untuk mengambil komentar pada video *Youtube* yang relevan dengan topik "Dampak *Roblox*" dan "Psikologi

Anak". Proses ini menggunakan *Web scraping* yang tersedia secara publik yang dimana hanya memerlukan tautan video yang akan digunakan sebagai sumber dataset.

Tabel 1. Sumber Data

No.	Judul	Channel
1.	Game ROBLOX Memakan Banyak Korb4n & Segera Di-Blockir Pemerintah	Kamar JERI
2.	ROBLOX Sekarang Lebih Berbahaya Buat Bocil! Ada Apa?!	MainSpace
3.	20 Dosa Yang Dilakukan Player ROBLOX!! Coba Tonton Video Ini.	ROBLOX TOPIC INDONESIA

2. *Filtering dan Labeling*: Pada proses ini, peneliti akan melakukan mengelompokkan data berdasarkan relevan atau tidaknya data terhadap topik penelitian ini dan akan dikategori menjadi 2 macam yaitu sentimen yang terdiri dari positif, negatif dan netral, dan aspek yang terdiri dari sosial, keamanan, dan kognitif.

Tabel 2. Contoh *Labeling* Data

No.	Comment	Sentimen	Aspek
1.	balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang term manipulatif,, kalo anak yg berumur 15 udah bisa sih tau salah/bener,, ini menurut saya ya ges ♥	Netral	Keamanan
2.	Walaupun aku gak pernah pw2 di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox 🧑🏻 🧑🏻	Positif	Sosial
3.	blocir aja lah ini gem robox,soal nya gara gara gem ini anak gua yg di indonesia sana,jadi gak terkendali,hp gak bisa lepas dari tangan nya gara gara roblox 🤖	Negatif	Kognitif

3. *Preprocessing*: Proses ini berfungsi untuk mengolah data text mentah dengan cara membersihkan simbol, kata hubung, imbuhan dan lainnya. Pada proses ini meliputi *cleaning*, *case folding*, *stopword*, *tokenizing*, *stemming* dan pembersihan bahasa *slang* menggunakan kamus normalisasi khusus.

a) *Cleaning*

Proses ini berfungsi untuk membersihkan data teks mentah dari simbol, angka, tanda baca, dan kata lainnya yang tidak memiliki makna sentiment.

Tabel 3. Proses *Cleaning Data*

Sebelum	Sesudah
balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif,, kalo anak yg berumur 15 udah bisa sih tau salah/bener,, ini menurut saya ya ges ♥	Balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yg berumur udah bisa sih tau salah bener ini menurut saya ya ges
Walaupun aku gak pernah pw2 di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox ☹️☹️	Walaupun aku gak pernah pw di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox
blocir aja lah ini gem roblox,soal nya gara gara gem ini anak gua yg di indonesia sana,jadi gak terkendali,hp gak bisa lepas dari tangan nya gara gara roblox 🙄	blokir aja lah ini gem roblox soal nya gara gara gem ini anak gua yg di indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox

b) *Case Folding*

Proses ini berfungsi untuk menyamaratakan data text dengan cara mengubah semua kata kedalam huruf kecil.

Tabel 4. Proses *Case Folding*

Sebelum	Sesudah
Balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yg berumur udah bisa sih tau salah/bener ini menurut saya ya ges	balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yg berumur udah bisa sih tau salah bener ini menurut saya ya ges
Walaupun aku gak pernah pw di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox	walaupun aku gak pernah pw di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox
blokir aja lah ini gem roblox soal nya gara gara gem ini anak gua yg di Indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox	blokir aja lah ini gem roblox soal nya gara gara gem ini anak gua yg di indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox

c) *Normalisasi*

Proses ini berfungsi dalam mengubah kata tidak baku menjadi kata baku. Dalam penelitian ini proses ini juga berfungsi dalam memperbaiki kesalahan penulisan atau *typo* pada data serta mengubah beberapa kata slang menjadi kata baku.

Tabel 5. Proses Normalisasi

Sebelum	Sesudah
balik lagi sih bang ortunya ga ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yg berumur udah bisa sih tau salah/bener ini menurut saya ya ges	balik lagi sih bang orang tua nya gak ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yang berumur udah bisa sih tau salah bener ini menurut saya ya ges
walaupun aku gk pernah pw di roblox tp aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox	walaupun aku gak pernah pw di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox
blocir aja lah ini gem roblox soal nya gara gara gem ini anak gua yg di indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox	blokir saja lah ini game roblox soal nya gara gara game ini anak saya yang di indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox

d) *Tokenizing*

Pada proses ini, kalimat pada data teks akan di pecah menjadi beberapa bagian kecil dalam bentuk tiap kata untuk memepmudahkan pada proses pengolahan data.

Tabel 6. Proses *Tokenizing*

Sebelum	Sesudah
balik lagi sih bang orang tua nya gak ngawasin anak anak pas main sih paling logis makanya gampang termanipulatif kalo anak yang berumur udah bisa sih tau salah/bener ini menurut saya ya ges	balik, lagi, sih, bang, orang, tua, nya, gak, ngawasin, anak, anak, pas, main, sih, paling, logis, makanya, gampang, termanipulatif, kalo, anak, yang, berumur, sudah, bisa, sih, tahu, salah, bener, ini, menurut, saya, ya, ges
walaupun aku gak pernah pw di roblox tapi aku masih bahagia dan orang tua ku gak peduli kalo aku main roblox	walaupun, aku, gak, pernah, pw, di, roblox, tapi, aku, masih, bahagia, dan, orang, tua, ku, gak, peduli, kalo, aku, main, roblox
blokir saja lah ini game roblox soal nya gara gara game ini anak saya yang di indonesia sana jadi gak terkendali hp gak bisa lepas dari tangan nya gara gara roblox	blokir, saja, lah, ini, game, roblox, soal, nya, gara, gara, game, ini, anak, saya, yang, di, indonesia, sana, jadi, gak, terkendali, hp, gak, bisa, lepas, dari, tangan, nya, gara, gara, roblox

e) *Stopwords Removal*

Proses ini berfungsi untuk menghapus kata umum yang tidak memiliki makna sentimen seperti “dan”, “atau”, “yang”.

Tabel 7. Proses *Stopwords Removal*

Sebelum	Sesudah
balik, lagi, sih, bang, orang, tua, nya, gak, ngawasin, anak, anak, pas, main, sih, paling, logis, makanya, gampang, termanipulatif, kalo, anak, yang, berumur, sudah, bisa, sih, tahu, salah, bener, ini, menurut, saya, ya, ges	balik, ngawasin, anak, main, logis, gampang, termanipulatif, anak, berumur, tahu, salah, bener
walaupun, aku, gak, pernah, pw, di, roblox, tapi, aku, masih, bahagia, dan, orang, tua, ku, gak, peduli, kalo, aku, main, roblox	walaupun, pernah, pw, roblox, bahagia, orang, tua, peduli, main, roblox
blokir, saja, lah, ini, game, roblox, soal, nya, gara, gara, game, ini, anak, saya, yang, di, indonesia, sana, jadi, gak, terkendali, hp, gak, bisa, lepas, dari, tangan, nya, gara, gara, roblox	blokir, game, roblox, gara, game, anak, indonesia, terkendali, hp, lepas, tangan, gara, roblox

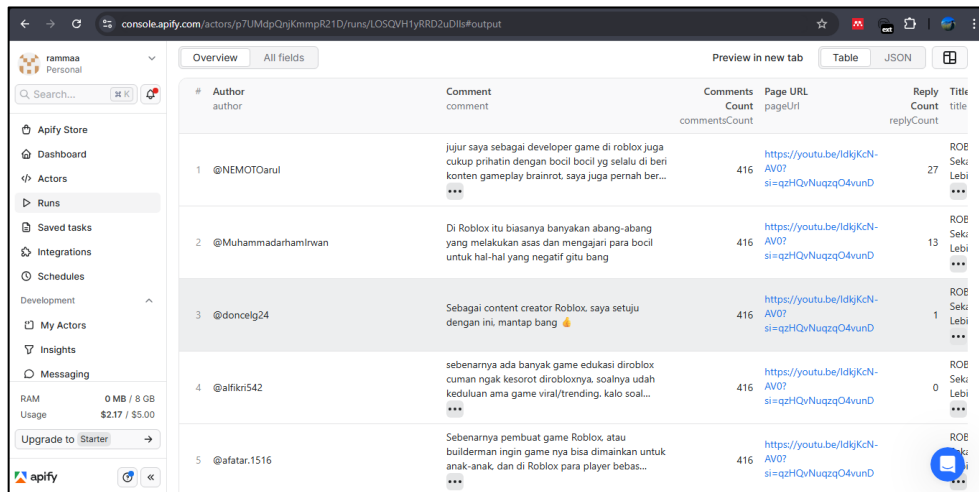
f) *Stemming*

Proses ini merupakan proses yang berfungsi untuk mengubah kata berimbuhan menjadi kata asalnya, seperti kata berimbuhan ber-, ter-, men-, dan lainnya. Proses stemming ini bekerja dengan cara memisahkan suatu kata yang memiliki imbuhan terlebih dahulu, seperti contoh kata “bermain” yang memiliki imbuhan ber-, lalu kata tersebut akan menjadi “ber”-“main” yang dimana terdiri dari kata imbuhan dan kata asalnya. Kemudian kata imbuhan di hapus sehingga hanya memisahkan kata asalnya yaitu “main”.

Tabel 8. Proses *Stemming*

Sebelum	Sesudah
balik, ngawasin, anak, main, logis, gampang, termanipulatif, anak, berumur, tahu, salah, bener	balik, awas, anak, main, logis, gampang, termanipulatif, anak, umu, tahu, salah, bener
walaupun, pernah, pw, roblox, bahagia, orang, tua, peduli, main, roblox	walaupun, pernah, pw, roblox, bahagia, orang, tua, peduli, main, roblox
blokir, game, roblox, gara, game, anak, indonesia, terkendali, hp, lepas, tangan, gara, roblox	blokir, game, roblox, gara, game, anak, indonesia, kendali, hp, lepas, tangan, gara, roblox

III. HASIL DAN PEMBAHASAN



#	Author	Comment	Comments Count	Page URL	Reply Count	Title
1	@NEMOTOarul	jujur saya sebagai developer game di roblox juga cukup prihatin dengan bocil bocil yg selalu di beri konten gameplay brainrot, saya juga pernah ber...	416	https://youtu.be/IdjKcN-AV07	27	ROB Seki Lebi
2	@Muhammadarhamlin	Di Roblox itu biasanya banyak abang-abang yang melakukan asar dan mengajari para bocil untuk hal-hal yang negatif gitu bang	416	https://youtu.be/IdjKcN-AV07	13	ROB Seki Lebi
3	@doncelg24	Sebagai content creator Roblox, saya setuju dengan ini, mantap bang 🙌	416	https://youtu.be/IdjKcN-AV07	1	ROB Seki Lebi
4	@alfikri542	sebenarnya ada banyak game edukasi di roblox cuman ngak keserot dirobloxnya, soalnya udah keduluan ama game viral/trending, kalo soal...	416	https://youtu.be/IdjKcN-AV07	0	ROB Seki Lebi
5	@afatar.1516	Sebenarnya pembuat game Roblox, atau buideman nging game nya bisa dimainin untuk anak-anak, dan di Roblox para player bebas...	416	https://youtu.be/IdjKcN-AV07		ROB Seki Lebi

Gambar 2. Hasil Pengumpulan Data

Proses konstruksi menghasilkan dataset sejumlah 2124 entri. Analisis awal menunjukkan bahwa kata-kata seperti "*toxic*", "*kasar*", dan "*candu*" mendominasi sentimen negatif, sedangkan "*kreatif*" dan "*belajar*" muncul pada sentimen positif. Dengan telah dilakukannya rangkaian proses sebelumnya, peneliti telah berhasil membangun sebuah dataset dengan validitas yang baik.

Pada penelitian ini, dataset telah diuji pada pelatihan model analisis sentimen guna mengidentifikasi data teks pada sentimen positif, negatif, atau netral. Pada proses tersebut model mendapatkan hasil evaluasi yang kurang memuaskan yang terletak pada hasil recall yang terlalu rendah. Hal ini di akibatkan oleh jumlah data yang tidak seimbang antara sentimen positif, negatif, dan juga netral. Untuk mengatasinya, dataset melakukan augmentasi data menggunakan metode SMOTE yang bekerja dengan menduplikasi dataset dengan jalur data terdekatnya. Setelah dilakukan augmentasi, hasil evaluasi model menunjukkan peningkatan yang sangat besar yang dimana seluruh hasil evaluasi berada diatas 90%.

IV. KESIMPULAN

Penelitian ini menggunakan data teks opini publik yang dikumpulkan menggunakan metode Scraping dengan mengambil komentar dari video pada platform YouTube. Data tersebut kemudian akan melalui beberapa tahapan seperti labeling, dan preprocessing. Data yang di peroleh akan dilakukan labeling dengan sentimen positif, negatif, dan netral dan data tersebut dapat digunakan dalam menganalisis sentimen platform Roblox apakah lebih dominan terhadap sentimen positif, negatif, atau netral. Setelah itu data akan melalui proses preprocessing yang dimana berfungsi untuk membersihkan data terhadap noise seperti simbol, angka, dan kata yang tidak memiliki makna terhadap sentimen.

Setelah data bersih, maka data akan di berikan bobot menggunakan metode TF-IDF yang dimana akan bekerja dengan menghitung frekuensi kemunculan kata dalam satu data dan hasilnya data akan berbentuk numerik yang bisa digunakan langsung dalam training model klasifikasi analisis sentimen. Pada tahap ini, data memang dapat langsung digunakan dalam proses training model, namun hasil yang diperoleh pada model tersebut terbilang buruk karena sistem tidak dapat membaca keseluruhan data yang diakibatkan oleh jumlah data yang tidak seimbang. Maka dari

itu akan melalui proses augmentasi data yang akan bekerja dalam meduplikasi data dengan cara mencari tetangga terdekat dari data yang dimiliki guna menyeimbangkan jumlah data.

Setelah melalui proses tersebut, model mendapatkan hasil evaluasi yang sangat baik untuk tiap aspeknya yang dimana berada di atas 90%. Maka dari itu penelitian ini telah dianggap berhasil dalam menyusun dataset analisis sentimen berbasis bahasa Indonesia yang spesifik berisi tentang dampak psikologi anak akibat game *online*. Namun data harus tetap melalui beberapa tahapan tambahan untuk mendapatkan hasil yang maksimal. Dataset ini nantinya siap digunakan dalam penelitian dan pengembangan model klasifikasi teks guna mendukung pengawasan aktivitas digital pada anak.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri dan semua pihak yang telah mendukung penelitian ini.

DAFTAR PUSTAKA

- [1]. M. P. Mahadewi, “Jejak Sejarah Muasal Gim Roblox | tempo.co.” Accessed: Nov. 05, 2025. [Online]. Available: <https://www.tempo.co/digital/jejak-sejarah-muasal-gim-roblox-2057198>
- [2]. M. Ihsan, “Sejarah Roblox dan Fakta Menarik di Balik Kesuksesannya | Dunia Games.” Accessed: Nov. 05, 2025. [Online]. Available: <https://duniagames.co.id/discover/article/fakta-menarik-asal-usul-game-roblox-dan-siapa-penciptanya>
- [3]. A. Fikri Alkindi, N. Nasution, and U. Lancang Kuning, “Analisis Sentimen Ulasan Pengguna pada Game Roblox dengan Metode Support Vector Machine dan Naïve Bayes,” *Journal of Computer online*, vol. 4, no. 2, pp. 164–177, 2024, doi: 10.33330/j-com.v4i2.3319.
- [4]. A. Maulana, “RRI.co.id - Game Online Roblox Berdampak Buruk Pada Psikologis Anak.” Accessed: Nov. 21, 2025. [Online]. Available: <https://rri.co.id/iptek/1760918/game-online-roblox-berdampak-buruk-pada-psikologis-anak>
- [5]. S. P. Subandi, N. Iman, and A. R. Syam, “Dampak Kecanduan Game Online terhadap Pendidikan Anak,” *Al-Kamal: Jurnal Kajian Islam*, pp. 243–262, 2022.
- [6]. Y. Prespi Mubaroq, P. Guru Sekolah Dasar, F. Ilmu Pendidikan, and U. Bhayangkara Jakarta Raya Alamat, “Dampak Game Online Roblox terhadap Motivasi Belajar Siswa Studi Kasus Di SD Negeri Setia Darma 04,” *Pendas : Jurnal Ilmiah Pendidikan Dasar*, vol. 10, no. 03, pp. 273–284, Aug. 2025, doi: 10.23969/JP.V10I03.31758.
- [7]. A. Kusuma, D. Sulistiani, I. Amelia, L. F. Rahma, and W. Widinia, “Representasi Gaya Bicara dan Perilaku Anak-anak SD dari Game Online Roblox,” *Karimah Tauhid*, vol. 4, no. 8, pp. 5955–5967, Aug. 2025, doi: 10.30997/KARIMAHTAUHID.V4I8.20161.
- [8]. R. Rizky, R. 1✉, and M. Zaky, “Analisis Kejahatan Siber (Cyber Child Grooming) Pada Game Online Roblox,” *INNOVATIVE: Journal Of Social Science Research*, vol. 5, pp. 1656–1669, 2025.

- [9]. Ariyo Munandar and C. Edi Widodo, “Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM,” *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 7, no. 1, pp. 77–84, 2023.
- [10]. Ma’rufudin and A. Yudhistira, “Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, pp. 845–857, Mar. 2025, doi: 10.52436/1.JPTI.717.