

Sentiment Analysis of Indonesian National Team Player Composition Using the Convolutional Neural Network

Analisis Sentimen Terhadap Komposisi Pemain Tim Nasional Indonesia Menggunakan Convolutional Neural Network

Fishilia Saqila Istanti¹, Ahmad Riyadi²

¹ Informatika, Universitas PGRI Yogyakarta, Indonesia

¹fishiliasaqila19@gmail.com, ^{2*}ahmadchasna@gmail.com,

*: Penulis korespondensi (corresponding author)

Informasi Artikel

Received: December 2025

Revised: December 2025

Accepted: January 2026

Published: February 2026

Abstract

Football is one of the most popular sports in Indonesia, especially when it comes to supporting the Indonesian National Team. The composition of the national team players is a frequently discussed topic among the public, particularly on social media platforms like YouTube. YouTube is one of the most popular social media platforms for expressing public opinions. Sentiment analysis can help identify and address issues based on public opinions shared on social media platforms such as YouTube. The classification method used in this study is Convolutional Neural Network. The dataset was obtained through data scraping, resulting in 3,200 data points. The labeling process was conducted manually by involving three annotators. The labeling results indicate 1,036 instances of "proportional," 1,416 of "not proportional," and 747 of "doubtful." Next, preprocessing was performed on the labeled data, followed by word weighting using TF-IDF. After that, modeling was conducted using Convolutional Neural Network, and the final step involved developing an interactive web application using Streamlit to analyze text sentiment based on the trained model. The accuracy result, comparing 80% training data and 20% testing data, achieved an accuracy of 89%. Meanwhile, the sentiment analysis results show that the "not proportional" sentiment appeared more frequently than the "proportional" and "doubtful" sentiments.

Abstrak

Sepak bola merupakan salah satu olahraga yang memiliki banyak penggemar di Indonesia, khususnya dalam

Keywords: Sentiment analysis, player composition, Indonesian National Team,

Convolutional Neural Network.

Kata kunci: Analisis sentimen, komposisi
pemain, Tim nasional indonesia,
Convolutional Neural Network.

mendukung Tim Nasional Indonesia. Komposisi pemain tim nasional menjadi topik yang sering diperbincangkan oleh masyarakat, terutama melalui media sosial seperti YouTube. Youtube merupakan salah satu meedia sosial yang populer untuk menyampaikan opini masyarakat. Proses analisis sentimen dapat menemukan dan menyelesaikan permasalahan berdasarkan opini masyarakat di media sosial seperti Youtube. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Convolutional Neural Network*. Dataset dapat diperoleh dari proses scraping data dan menghasilkan 3200 data. Proses pelabelan dilakukan secara manual dengan melibatkan 3 anatator. Hasil proses pelabelan menunjukan 1036 proporsional, 1416 Tidak proporsional dan 747 meragukan. Kemudian dilakukan preprocesing terhadap datayang telah diberi label sebelumnya, juga dilakukan proses pembobotan kata menggunakan TF – IDF. Setelah itu dilakukan pemodelan menggunakan *Convolutional Neural Network* dan proses terakhir adalah pembuatan sebuah web interaktif menggunakan streamlit untuk menganalisis sentimen teks dengan model yang sudah dibuat. asil akurasi dengan membandingkan 80% data training dan 20% data testing menghasilkan akurasi sebesar 89% Sedangkan hasil uji sentimen menunjukan bahwa sentiment Tidak proporsional lebih banyak dibandingkan sentiment proporsional dan meragukan

1. Pendahuluan

Indonesia seperti negara-negara lain mempunyai kecintaan yang mendalam terhadap olahraga, khususnya sepak bola karena banyaknya dukungan dari masyarakat untuk tim nasional Indonesia. Tim nasional sepak bola Indonesia merupakan simbol jati diri dan kebanggaan masyarakat Indonesia[1]. Keberhasilan tim nasional Indonesia sering kali membuat bangga negara sementara kegagalan bisa menyebabkan perdebatan Panjang mengenai strategi pertandingan dan pemain yang dipilih pelatih.

Tim nasional sepak bola Indonesia kerap menjadi perhatian publik terutama terkait struktur pemain atau komposisi pemain terutama dalam pemilihan pemain di berbagai kompetisi. Masyarakat sangat mengharapkan pemain yang dianggap layak untuk bergabung dengan timnas. Mereka mempertimbangkan berbagai hal seperti kemampuan pemain, usia, dan pengalaman bermain. Setiap kali daftar pemain diumumkan, masyarakat dan pecinta sepak bola kerap mengutarakan pendapat di media sosial.

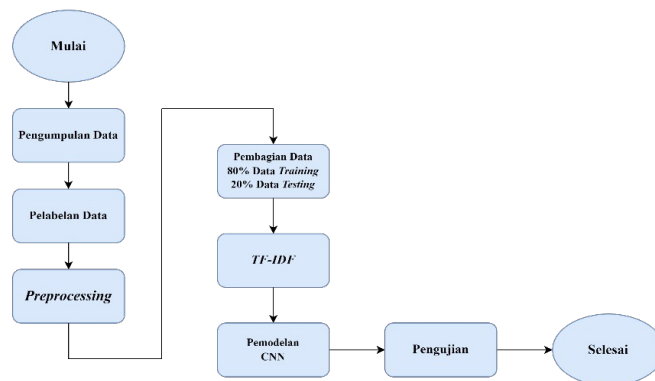
Dalam era digital ini, media sosial telah menjadi bagian penting dalam kehidupan sehari-hari[2]. Seiring dengan perkembangan media sosial, Penggunaan media sosial Youtube di Indonesia semakin meningkat. menurut riset *We Are Social* mencapai 88 % [3]. Komentar pada media sosial *Youtube* sering sekali menghasilkan opini berberbeda dari sudut pandang berbeda yang mencerminkan nilai sentimen[4]. Didalam komentar *Youtube* seringkali terdapat pendapat yang berbeda, oleh karena itu komentar *YouTube* dapat menjadi sumber untuk mencari data penting untuk menganalisis sentimen publik [5].

Analisis sentimen yang sering di sebut *opinion mining*[6]. Analisis sentimen ini penting di lakukan untuk memahami secara umum bagaimana pendapat masyarakat mengenai komposisi pemain tim nasional sepak bola indonesia dan mengklasifikasikan teks berdasarkan kalimat apakah masuk kategori proposional, tidak proposional, meragukan. *Natural Language processing* (NLP) dikenal dengan linguistik komputasi yang dapat memecahkan masalah praktis dalam bahasa manusia[7].

Convolutional Neural Network (CNN) adalah salah satu algoritma *deep learning* yang sering di gunakan untuk mengklasifikasikan gambar, untuk mengidentifikasi wajah, rambu laulitas, mengidentifikasi penyakit dan berbagai data visual lainnya namun bisa digunakan dalam kalimat atau teks. Metode *Convolutional Neural Network* juga digunakan dalam analisis sentimen dengan menggunakan *neural language processing* (NLP). Metode *Convolutional Neural Network* memiliki keunggulan di bandingkan metode *Deep learning* lainnya. *Convolutional Neural Network* memiliki jumlah koneksi yang lebih sedikit lebih, memiliki parameter yang sederhana dan mudah dalam proses pelatih (*training*) setelah dilakukan tahap *training* selanjutnya di lakukan tahap *testing*, yang bertujuan untuk mengevaluasi bagaimana *algoritma Convolutional Neural Network* mampu menghasilkan nilai evaluasi berdasarkan prediksi model yang telah dibangun[8].

2. Metode/Perancangan

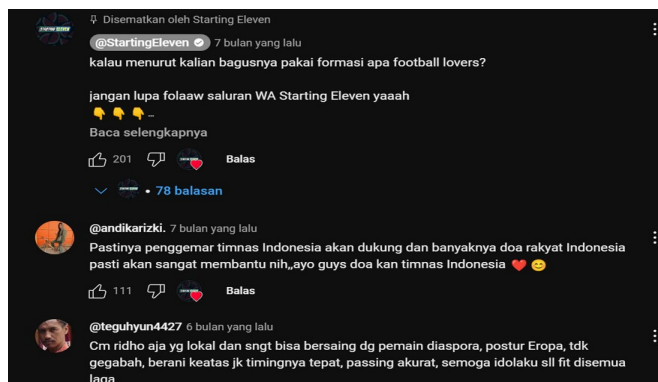
Pada Metodologi penelitian akan diawali dengan pengumpulan data, kemudian pelabelan data, *preprocessing* data, pembagian data, pembobotan data *training*, penggunaan metode *Convolutional Neural Network*, dan diakhiri dengan pengujian klasifikasi. Metodologi penelitian dapat dilihat pada **Gambar 1**.



Gambar 1. Flowchart metodologi penelitian

2.1. Pengumpulan Data

Data dikumpulkan dari komentar *Youtube* pada Channel pada *Channel starting Eleven dan starting Eleven story* pengambilan data dari 1juli 2024- 10 januari 2025. Banyaknya data komentar awal yang terkumpul 3500 data kemudian data yang akan digunakan untuk preprocessing adalah 3200 data kemudia disimpan dalam tabel dengan format (.csv). Pada tahap ini sudah berjalan 100%, artinya data telah terkumpul 3200 komentar. Berikut menunjukan komentar Youtube pada **Gambar 2**.



Gambar 2. Komentar Youtube

2.2. Pelabelan Data

Pelabelan data dilakukan secara manual untuk menilai semua komentar yang bernilai “Proposional”, “Tidak Proposional” dan “meragukan” dengan tujuan untuk mengetahui opini apa saja yang terdapat di dalam data yang sesuai dengan tema penelitian. Pembagian label dalam data seperti **Tabel 1**.

Tabel 1. Pedoman Pelabelan

Kategori	penjelasan
Proporsional	Komentar yang mengandung dukungan atau meyetujui bahwa komposisi pemain tim nasional Indonesia telah sesuai dengan performa ,kebutuhan atau syarat aturan yang berlaku.
Tidak Proporsional	Komentar yang menentang atau tidak setuju dengan komposisi pemain tim nasional Indonesia, karena tidak adil atau tidak sesuai dengan harapan.

Tabel 1. Pedomen Pelabelan

Pelabelan data dilakukan secara manual dengan memberikan nilai “Proporsional”, “Tidak

Kategori	penjelasan
meragukan	Komentar yang menunjukkan sikap ragu, seperti tidak yakin dan tidak jelas mendukung atau menolak.

Proporsional” dan “meragukan”. Tujuannya untuk mengetahui opini apa saja yang ada didalam data yang sesuai atau tidaknya dalam tema penelitian. Pelabelan dilakukan dengan cara manual dengan acuan pada pedoman pelabelan dapat memudahkan proses pelabelan pada data. Berikut contoh sampel data yang telah diberikan lebel pada **Tabel 2**

Tabel 2. Sampel data yang telah diberikan Label

No	Komentar	Label
1	Komposisi pemain kali ini sangat strategis, semoga bisa membawa hasil terbaik untuk Timnas Indonesia.	Proporsional
2	Kenapa Marselino hanya jadi cadangan? Padahal dia salah satu pemain kunci di timnas	Tidak Proporsional
3	Apakah strategi ini benar- benar akan berhasil? susunan pemainnya terlihat kurang meyakinkan	meragukan

2.3. Preprocessing

Preprocessing merupakan tahapan dimana data dipilih dan diolah sehingga menjadi lebih terstruktur. Oleh sebab itu tahapan *Preprocessing* di butuhkan untuk mengklasifikasi dokumen yang bertipe teks yaitu *Cleaning*, *Case Floding (Lowercase)*, *Normalisasi Stemming*, *Tokenezation*.

2.4. Pemodelan

Pemodelan data dibuat dengan tujuan untuk menguji keakuratan prediksi sistem berdasarkan data model yang sudah dibuat. Proses *training* untuk membangun model prediksi pada penelitian ini dibutuhkan data *training*. Selanjutnya dilakukan pengujian model klasifikasi sentimen yang dihasilkan pada proses *training* dengan menggunakan data komentar baru data *testing*. Tahap pertama adalah memasukan dataset yang berisi komentar publik dari media sosial *youtube* mengenai komposisi pemain timnas indonesia. setelah itu, dataset di beri lebel dengan tiga kategori yaitu ‘proporsional’, ‘tidak proporsional’ dan ‘meragukan’. setelah pebelan selesai data akan di proses

melalui preprocessing untuk mengubah data yang tidak terstruktur menjadi lebih terstruktur dan siap digunakan untuk proses berikutnya. Langkah berikutnya adalah membagi dataset menjadi dua bagian yaitu data training sebesar 80% dan data testing sebesar 20%. Selanjutnya dilakukan pembobotan menggunakan metode TF-IDF yang bertujuan untuk menghitung nilai bobot setiap kata dalam data training. Kemudian data akan masuk ke tahap klasifikasi menggunakan metode CNN yang akan mengenali pola-pola pada data training berdasarkan bobot kata yang telah dihitung, sehingga model dapat belajar

mengidentifikasi sentimen dari komentar yang berbeda. Pada proses pengujian, data testing tidak memerlukan pelabelan tetapi tetap harus melalui tahap *preprocessing*. Setelah data testing dimasukkan ke dalam sistem, dilakukan evaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *f1-score* untuk menilai kemampuan model dalam mengklasifikasikan komentar ke dalam kategori ‘Proporsional’, ‘Tidak Proporsional’, atau ‘Meragukan

2.5. Pengujian

Pengujian sistem dilakukan untuk mengevaluasi performa model analisis sentimen terhadap komposisi pemain timnas Indonesia yang di bangun menggunakan algoritma *Convolutional Neural Network*. *Confusion matrix* adalah cara untuk mempresentasikan hasil daripada akurasi model yang sudah dibuat. *Confusion matrix* merangkum performa dalam mengklasifikasi sesuai dengan banyaknya kategori yang diklasifikasikan dari nilai benar dari suatu kelas dari objek yang di prediksi. Berikut ini disajikan representasi tabel *Confusion matrix* dalam konteks analisis sentimen yang mencakup tiga kelas: proporsional, tidak proporsional dan meragukan. **Tabel 3** Berikut tabel yang memperlihatkan kerja dari *confusion matrix* yang ada dalam kelas prediksi dan kelas actual:

Tabel 3. Confusion Matrix

Nilai Aktual	Nilai Prediksi		
	Proporsional	Tidak Proporsional	Meragukan
Proporsional	TP (Proporsional)	FP (tidak proporsional)	FP (Meragukan)
Tidak Proporsional	FN (Proporsional)	TP (tidakproporsional)	FP (Meragukan)
Meragukan	FN (False Proporsional)	FN (tidak proporsional)	TP (Meragukan)

Keterangan :

- TP (*True Positif*) Proporsional= Jumlah data yang benar-benar berada dalam kelas Proporsional dan juga diprediksi dengan benar sebagai kelas Proporsional.
- TP (*True Positif*) Tidak Proporsional= Jumlah data yang benar-benar berada dalam kelas Tidak Proporsional dan juga diprediksi dengan benar sebagai kelas

Tidak Proporsional.

- c) TP (*True Positif*) Meragukan= Jumlah data yang benar-benar berada dalam kelas Meragukan dan juga diprediksi dengan benar sebagai kelas Meragukan.
- d) FP (*False Positif*) Tidak Proporsional= Jumlah data yang berada dalam kelas aktual Proporsional, tetapi diprediksi sebagai kelas Tidak Proporsional.
- e) FP (*False Positif*) Meragukan = Jumlah data yang berada dalam kelas aktual proporsional dan Tidak Proporsional, tetapi diprediksi sebagai kelas merakukan .
- f) FN (*False Negatif*) proporsional= Jumlah data yang berada dalam kelas aktual tidak proporsional dan merakukan, tetapi diprediksi sebagai kelas proporsional.
- g) FN (*False Negatif*) Tidak Proporsional = Jumlah data yang berada dalam kelas aktual merakukan, tetapi diprediksi sebagai Tidak Proporsional.

Dalam klasifikasi dengan tiga kelas akan memiliki enam entri utama: True Positif Proporsional (TP), True Positif Tidak Proporsional (TTP), True Positif Meragukan (TM), False Negatif Proporsional (FP), False Negatif Tidak Proporsional (FTP), False Negatif Meragukan (FM).

Nilai matriks tersebut digunakan untuk menghitung *Accuracy*, *Precision*, *Recall*, dan *F¹ score* dengan persamaan dengan memperlihatkan:

$$Akurasi = \frac{TP+TTP+TM}{TP+FP+TTP+FTP+TM+FM}$$

$$Recall = \frac{TP}{TP+FP}$$

$$Precision = \frac{TP}{TP+FTP+FM}$$

$$F^1 score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Matriks *Accuracy*, *precision*, *recall*, dan *f¹score* digunakan untuk mengukur performa model dalam melakukan prediksi pada kelas proporsional, tidak proporsional dan merakukan,

3. Hasil Dan Pembahasan

3.1 Bagi Hasil Dataset

Dataset untuk penelitian ini berasal dari proses *scraping* data di youtube. Dataset yang diperoleh diberi label secara manual oleh 3 orang yang ahli dalam bidang sastra indonesia. Berdasarkan hasil pelabelan didapatkan 1416 Tidak proporsional, 1036 Proporsional dan 747 merakukan. Sehingga jumlah keseluruhan dataset pada penelitian ini sebanyak 3200 data. Data tersebut berisi data *training* dan data

testing. Dataset untuk penelitian ini dibagi menjadi data latih dan data uji dengan perbandingan 80%:20%. Ahli bahasa telah melabeli contoh pada **Tabel 4** sebagai berikut:

Tabel 4. Contoh Pelabelan Dataset

Komentar	Label
1. Dengan susunan ini, tim terlihat lemah.	Tidak Proporsional
2. Ini adalah salah satu tim terbaik yang pernah disusun	Proporsional
3. Saya ragu apakah formasi ini bisa bekerja dengan baik....	Meragukan

3.2 Hasil Pre-processing

Pada tahap *preprocessing* dataset diawali dengan *Cleaning* proses pembersihan data yang dilakukan untuk menghilangkan kalimat atau data duplikat agar lebih terstruktur. [3]. Hasil proses *cleaning* dapat dilihat pada **Tabel 5**:

Tabel 5. Hasil *cleaning*

Komentar	Hasil <i>Cleaning</i>
1. Dengan susunan ini, tim terlihat lemah.	Dengan susunan ini, tim terlihat lemah.
2. Ini adalah salah satu tim terbaik yang pernah disusun	Ini adalah salah satu tim terbaik yang pernah disusun
3. Saya ragu apakah formasi ini bisa bekerja dengan baik	Saya ragu apakah formasi ini bisa bekerja dengan baik

Proses mengonversi data dari huruf besar ke huruf kecil (*lowercase*) dikenal sebagai *case folding*. Hasil proses *case folding* dapat dilihat pada Tabel **Tabel 6**:

Tabel 6. Hasil *Case Folding*

Komentar	Hasil <i>Case folding</i>
1. Dengan susunan ini, tim terlihat lemah.	dengan susunan ini, tim terlihat lemah.
2. Ini adalah salah satu tim terbaik	ini adalah salah satu tim terbaik yang

yang pernah disusun	pernah di susun
3. Saya ragu apakah formasi ini bisa bekerja dengan baik	saya ragu apakah formasi ini bisa bekerja dengan baik
Proses normalisasi untuk menormalisasi kata terhadap setiap kata-kata yang mengandung kata tidak baku maupun noise menjadi kata yang baku dan siap diola Tabel 7.	

Tabel 7. Hasil *normalisasi*

Komentar	Hasil Normalisasi
1. dengan susunan ini, tim terlihat lemah.	dengan susunan ini, tim terlihat lemah
2.ini adalah salah satu tim terbaik yang pernah di susun	ini adalah salah satu tim terbaik yang pernah di susun
3. saya ragu apakah formasi ini bisa bekerja dengan baik	saya ragu apakah formasi ini dapat berfungsi dengan baik

Proses Steaming untuk mencari kata dasar dari setiap kata. Pada proses ini, semua imbuhan yang ada pada data dokumen akan dihapus, termasuk awalan, akhiran, dan kombinasi keduanya . Hasil proses *steaming* dapat dilihat pada **Tabel 8:**

Tabel 8. Hasil *Steaming*

Komentar	Hasil Steaming
1. dengan susunan ini, tim terlihat lemah	dengan susun ini tim lihat lemah
2. ini adalah salah satu tim terbaik yan pernah di susun	ini ada salah satu tim baik yang pernah susun
3. saya ragu apakah formasi ini dapat berfungsi dengan baik	saya ragu apa formasi ini dapat fungsi dengan baik

Tokenizing yaitu digunakan untuk menghapus karakter-karakter tertentu seperti tanda baca. Karakter spasi berfungsi sebagai pemisah untuk membagi kalimat menjadi sekelompok kata. Hasil proses *tokenizing* dapat dilihat pada **Tabel 9:**

Tabel 8. Hasil *Tokenizing*

Komentar	Hasil Tokenizing
1.dengan susun ini tim lihat lemah	["dengan", "susun", "ini", "tim", "lihat", "lemah"]
2. ini ada salah satu tim baik yang pernah susun	["ini", "ada", "salah", "satu", "tim", "baik", "yang", "pernah", "susun"]

3. saya ragu apa formasi ini dapat fungsi ["saya", "ragu", "apa", "formasi",
dengan baik "ini", "dapat", "fungsi", "dengan",
"baik"]

Kata dalam penelitian ini dengan menggunakan fungsi *feature_extraction* pada *library scikit learn*. Sebagai contoh perhitungan proses TF-IDF dalam penelitian ini menggunakan sampel data *testing* dan *training*.

3.3 Hasil Uji Evaluasi

Proses uji evaluasi menggunakan *confusion matrix* yang berguna untuk menentukan keakuratan model yang sebenarnya. Proses perhitungan *confusion matrix* meliputi *accuracy*, *precision*, *recall* dan *f1-score*. Pada penelitian ini, penulis melakukan perbandingan 80% data *training* dan 20% data *testing*. Hasil uji evaluasi dapat dilihat pada **Tabel 10**:

Tabel 9. Hasil Uji Evaluasi

	Proporsional	Tidak Proporsional	Meragukan
<i>Accuracy</i>	89%	89%	89%
<i>Precision</i>	85%	93%	86%
<i>Recall</i>	94%	87%	85%
<i>F1-Score</i>	89%	90%	85%

4. Kesimpulan dan Saran

Disimpulkan bahwa 3200 data komentar youtube terhadap komposisi pemain tim nasional indonesia menunjukan hasil tidak proporsional yang lebih tinggi di bandingkan dengn proporsional dan meragukan. Hasil *accuracy* yang di dapatkan dengan membandingkan rasio 80:20 adalah sebesar 89%.

Saran untuk penelitian ini supaya bisa dikembangkan sistem pelabelan yang semula manual menjadi pelabelan otomatis. Dan diharapkan bisa dikembangkan dengan berbagai metode lainnya, baik metode di dalam *machine learning* atau *deep learning*, serta penelitian selanjutnya dapat menambah jumlah data yang akan dianalisis

Daftar Pustaka

- [1] E. C. Tyas Gusti, Eti Setiawati, and W. Warsiman, "Strategi Media Daring Kompas.com dalam Membentuk Identitas Sepak Bola Nasional: Analisis Wacana Model Theo van Leeuwen," *J. Onoma Pendidikan, Bahasa, dan Sastra*, vol. 10, no. 3, pp. 3099–3118, Jul. 2024, doi: 10.30605/onoma.v10i3.3919.

-
- [2] M. A. Java, Mohammad Syafrullah, W. Windarto, and P. Painem, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 2, pp. 75–80, Jan. 2024, doi: 10.70309/ticom.v12i2.112.
 - [3] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.
 - [4] M. F. Rizki, K. Auliasari, and R. Primaswara Prasetya, "Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 548–556, 2021, doi: 10.36040/jati.v5i2.3808.
 - [5] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 42–51, Feb. 2021, doi: 10.20895/DINDA.V1I1.180.
 - [6] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
 - [7] A. A. Permana, M. F. Fahrezi, D. A. Kristiyanti, and M. Sihotang, "Sentimen Analisis Opini Masyarakat Pada Media Sosial Twitter Terhadap Vaksin Berbayar Menggunakan Metode Naïve Bayes Classifier (Nbc)," *J. Tek.*, vol. 10, no. 2, pp. 84–92, 2021, doi: 10.31000/jt.v10i2.5471.
 - [8] R. Naquitasia, D. H. Fudholi, and L. Iswari, "Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning," *J. Teknoinfo*, vol. 16, no. 2, p. 156, 2022, doi: 10.33365/jti.v16i2.1516.