
EVALUASI MODEL HYBRID NAÏVE BAYES-XGBOOST UNTUK KLASIFIKASI SENTIMEN NETIZEN TERHADAP ISU FREE PALESTINE PADA PLATFORM X DI INDONESIA TAHUN 2025

Richi Andrianto¹, Mustopa Husein Lubis², Urfi Utami³, Asep Supriyanto⁴

^{1,2}Institut Teknologi dan Sains Padang Lawas Utara

^{3,4}Universitas Pasir Pengaraian

email: ¹richiandrianto28@gmail.com, ²mustopahusein1302@gmail.com,

³urfiutami@upp.ac.id, ⁴asep.tif@gmail.com

Abstract: *The Palestinian conflict has become a global issue that has triggered significant public responses, especially through social media platforms. This study aims to evaluate the performance of a hybrid Naïve Bayes–XGBoost model in classifying netizen sentiment toward the Free Palestine issue on platform X (formerly Twitter) in the year 2025. Data were collected using the X (Twitter) API with keywords such as #freepalestine and #savegaza, then processed through a series of preprocessing stages and sentiment labeling using a lexicon-based approach. The dataset was then split into 80% training data and 20% testing data to compare the performance of the baseline Naïve Bayes model and the hybrid model. The evaluation results show that the baseline Naïve Bayes model achieved an accuracy of 75.7% and an F1-score of 76%, while the hybrid Naïve Bayes–XGBoost model achieved a significantly higher accuracy of 95.5% and an F1-score of 96%. These findings indicate that integrating the two algorithms improves both accuracy and balance in sentiment classification, especially for unstructured and imbalanced data. This study recommends the use of hybrid models for public opinion analysis on social media and suggests further development using deep learning approaches.*

Keywords: *Sentiment Analysis, Free Palestine, Naïve Bayes, XGBoost, Hybrid Model*

Abstrak: Konflik Palestina menjadi isu global yang memicu respons besar dari masyarakat dunia, terutama melalui media sosial. Penelitian ini bertujuan untuk mengevaluasi performa model *hybrid Naïve Bayes–XGBoost* dalam mengklasifikasikan sentimen netizen terhadap isu *Free Palestine* di platform X (Twitter) tahun 2025. Data dikumpulkan menggunakan X (Twitter) API dengan kata kunci #freepalestine dan #savegaza, lalu diproses melalui tahapan *preprocessing*, dan pelabelan menggunakan pendekatan *lexicon-based*. Selanjutnya, data dibagi menjadi data latih (80%) dan data uji (20%) untuk membandingkan performa antara model *Naïve Bayes* dasar dan model *hybrid*. Hasil evaluasi menunjukkan bahwa model *Naïve Bayes* dasar menghasilkan akurasi 75,7% dan F1-score 76%, sedangkan model *hybrid Naïve Bayes–XGBoost* mencapai akurasi 95,5% dan F1-score 96%. Temuan ini menunjukkan bahwa integrasi kedua algoritma mampu meningkatkan akurasi dan keseimbangan klasifikasi sentimen, khususnya pada data yang tidak terstruktur dan imbalanced. Penelitian ini merekomendasikan penggunaan model *hybrid* untuk analisis opini publik di media sosial, serta pengembangan lebih lanjut menggunakan pendekatan *deep learning*.

Kata kunci: Sentimen, Free Palestine, Naïve Bayes, XGBoost, Hybrid Model

PENDAHULUAN

Perkembangan teknologi dan media sosial telah membuka ruang publik digital yang luas bagi masyarakat untuk menyuarakan opini dan ekspresi sosial. Salah satu isu global yang konsisten menjadi perhatian dunia adalah konflik antara Palestina dan Israel. Ditengah eskalasi konflik yang terjadi pada tahun 2025, media sosial seperti X (sebelumnya Twitter) menjadi kanal utama bagi netizen di berbagai negara untuk menyampaikan dukungan, kecaman, dan berbagai bentuk solidaritas melalui unggahan opini. Data opini publik ini terekam dalam bentuk data tidak terstruktur seperti tweet, komentar, dan retweet yang kaya akan muatan sentimen.

Analisis sentimen merupakan salah satu pendekatan dalam text mining yang digunakan untuk memahami persepsi masyarakat terhadap isu tertentu dengan cara mengklasifikasikan opini kedalam kategori sentimen positif, negatif, atau netral. Penelitian dibidang ini telah banyak dilakukan menggunakan pendekatan *machine learning*. Salah satu algoritma *machine learning* yang populer dalam klasifikasi sentimen adalah *Naïve Bayes* yang dikenal sederhana dan efisien dalam memproses data teks berdimensi tinggi (Gangwar & Mehta, 2022).

Namun, pendekatan tunggal seperti *Naive Bayes* memiliki keterbatasan dalam menangani data kompleks, seperti ketidakseimbangan kelas dan konteks semantik yang luas (Kowsari et al., 2019). Oleh karena itu, banyak penelitian terkini mengusulkan penggunaan metode *ensemble* untuk mengatasi kelemahan tersebut (Liu, 2023).

Salah satu metode *ensemble* yang menonjol adalah *XGBoost* (*Extreme Gradient Boosting*). *XGBoost* terbukti mampu mengatasi noise dan menangkap pola non-linear dalam data teks. Penelitian oleh Hendrawan, Utami, dan Hartanto (2022) menunjukkan bahwa model *XGBoost* dengan representasi *Word2Vec* mampu menghasilkan *F1-score* 0,941, lebih tinggi dibandingkan *Naïve Bayes* yang hanya mencapai 0,915

pada klasifikasi ulasan produk (Hendrawan et al., 2022).

Lebih lanjut, kombinasi kedua metode ini mulai banyak dikaji untuk menghasilkan model *hybrid*. Studi oleh Irwanto dan Goeirmanto (2023) membuktikan bahwa integrasi *Naïve Bayes* dan *XGBoost* dalam sentimen analisis dimedia sosial X (Twitter) menghasilkan akurasi yang lebih akurat dibanding metode tunggal (Irwanto & Goeirmanto, 2023). Bahkan dalam studi sejenis yang diterapkan pada pemilu di Nigeria, model *XGBoost* memperoleh akurasi hingga 93% dan precision sebesar 96%, menunjukkan kemampuannya dalam menangani opini publik berskala besar (Attai et al., 2024).

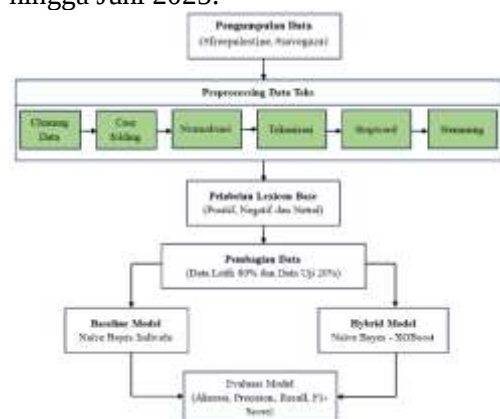
XGBoost sendiri merupakan salah satu metode *boosting* paling populer dan efektif dalam meningkatkan akurasi model klasifikasi melalui prinsip *gradient boosting decision tree*. Metode ini terbukti unggul dalam menghadapi data tidak seimbang (*imbalanced*) dan dapat mempelajari hubungan non-linear yang kompleks antar fitur. Model ini juga dikenal karena efisiensinya dalam pelatihan model dan kemampuannya untuk menangani noise dalam data teks tidak terstruktur (Burnwal & Jaiswal, 2023).

Melalui studi ini, peneliti akan mengkaji efektivitas kombinasi algoritma *Naive Bayes* dan *XGBoost* dalam menganalisis sentimen publik terhadap isu *Free Palestine* di platform X pada tahun 2025. Isu ini dipilih karena memiliki signifikansi global dan memicu respon luas dari masyarakat internasional. Evaluasi terhadap model *hybrid* ini diharapkan dapat memberikan kontribusi dalam pengembangan metode klasifikasi sentimen yang lebih akurat, khususnya untuk isu-isu sosial-politik yang dinamis dan kontroversial.

METODE

Penelitian ini mengadopsi pendekatan kuantitatif dengan metode eksperimen untuk menilai kinerja model

hybrid yaitu menggabungkan algoritma *Naive Bayes* dan *XGBoost* dalam tugas klasifikasi sentiment kemudian membandingkan kinerja model dengan algoritma *Naive Bayes* Individu. Dataset yang digunakan berasal dari unggahan publik (*tweet*) pada platform social media X, yang mengandung kata kunci #freepalestine dan #savegaza. Pengumpulan data dilakukan melalui X (Twitter) API selama periode Januari hingga Juni 2025.



Gambar 1 Alur Penelitian

Berikut ini penjelasan dari alur penelitian:

1. **Pengumpulan Data**

Pengumpulan data dilakukan dengan menggunakan teknik *crawling data* menggunakan kata kunci #FreePalestine, dan #SaveGaza. Dimana data hanya tweet berbahasa Indonesia yang diambil. Data yang dikumpulkan bersifat tidak terstruktur dan kemudian dibersihkan melalui tahapan *data preprocessing*.

2. **Preprocessing data**

Tahap *preprocessing* data dilakukan untuk membersihkan dan menyiapkan data teks agar dapat dianalisis secara efektif oleh model klasifikasi.

Proses ini diawali dengan *cleaning data*, yaitu menghapus elemen-elemen yang tidak relevan seperti tanda baca, angka, tautan URL, emoji, mention (@username), dan hashtag yang sering kali muncul dalam tweet (Cerrahoğlu & Cihan, 2023).

Selanjutnya dilakukan *case folding*, yakni mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menghindari redundansi akibat perbedaan kapitalisasi (Hafizah et al., 2025).

Setelah itu, teks melalui proses normalisasi, yaitu mengganti kata-kata tidak baku atau singkatan khas media sosial menjadi bentuk standar yang mudah dikenali oleh sistem, misalnya kata “gk” menjadi “tidak” atau “bgt” menjadi “banget” (Nugraha & Rizqullah, 2019). Kemudian, dilakukan tokenisasi, yaitu memecah kalimat menjadi potongan-potongan kata (*token*) agar dapat dianalisis secara terpisah (Jonjala et al., 2025). Token yang tidak memberikan makna penting dalam konteks sentimen, seperti “dan”, “yang”, “ke”, dan kata sambung lainnya, dihapus melalui proses *stopword removal* (Negara, 2023).

Tahapan terakhir adalah *stemming*, yaitu mengubah setiap kata ke bentuk dasarnya (*root word*) menggunakan algoritma *stemming* Bahasa Indonesia, seperti Sastrawi, misalnya, kata “menyerang”, “penyerangan”, dan “diserang” akan distem menjadi “serang” (Rianto et al., 2021).

Dengan menyelesaikan semua tahapan ini, teks menjadi lebih bersih, terstruktur, dan siap diolah lebih lanjut melalui tahapan ekstraksi fitur dan klasifikasi.

3. **Pelabelan Lexicon Based**

Setelah proses *preprocessing*, data *tweet* diberi label sentimen menggunakan pendekatan *lexicon-based*, yaitu dengan mencocokkan kata-kata dalam *tweet* terhadap kamus sentimen (*lexicon*) berbahasa Indonesia (Saputra et al., 2020). *Lexicon* yang digunakan adalah IndoLexicon yang mengklasifikasikan kata ke dalam kategori positif, negatif, dan netral. Setiap *tweet* diberi skor berdasarkan

jumlah kata bermuatan sentimen. Jika skor positif lebih tinggi, tweet dilabeli sebagai positif; jika negatif lebih tinggi, maka negatif, dan jika seimbang atau tidak ada kata sentimen, maka netral. Metode ini dipilih karena efisien, tidak membutuhkan data latih, dan cocok sebagai dasar pelabelan awal dalam penelitian klasifikasi sentiment (Mohammed & Prasad, 2023).

4. Pembagian Data

Selanjutnya membagi data menjadi dua bagian yaitu 80% sebagai data latih (*training set*) untuk membangun model klasifikasi, dan 20% sebagai data uji (*testing set*) untuk mengevaluasi kinerjanya. Pembagian ini dilakukan secara acak untuk menjaga distribusi data tetap seimbang antar kelas.

5. Pengujian *Naïve Bayes* Individu dan *Hybrid Naïve Bayes-XGBoost*

Pengujian dilakukan dengan membandingkan performa antara model *Naïve Bayes* individu dan model *hybrid Naïve Bayes-XGBoost*. Model *baseline* (*Naïve Bayes*) diterapkan langsung pada data hasil preprocessing. Sedangkan model *hybrid* dibentuk dengan pendekatan *stacking*, menggunakan *output* probabilitas *Naïve Bayes* sebagai input tambahan ke *XGBoost*.

6. Evaluasi Model

Evaluasi dilakukan untuk mengukur performa model *Naïve Bayes* dan *hybrid Naïve Bayes-XGBoost* dalam klasifikasi sentimen. Evaluasi model menggunakan metrik evaluasi berupa akurasi, precision, recall, dan F1-score (Goutte & Gaussier, 2005). Evaluasi bertujuan untuk mengetahui sejauh mana model *hybrid* mampu meningkatkan kinerja dibandingkan model individu dalam mengklasifikasikan sentimen netizen terhadap isu *Free Palestine* diplatform X.

HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dalam penelitian merupakan postingan maupun komentar yang diperoleh dengan teknik crawling data menggunakan API media sosial X (Twitter) selama periode Januari 2025 s/d Juni 2025. Selama periode tersebut diperoleh sebanyak 3.442 data tweet.



Gambar 2 Hasil Crawling Data

Preprocessing

Langkah pertama *preprocessing* data dilakukan *cleaning data*. Sehingga diperoleh data pada Tabel 1.

Tabel 1 Hasil *Cleaning Data*

Sebelum <i>Cleaning Data</i>	Sesudah <i>Cleaning Data</i>
Politisi Irlandia Menangis saat Pidato soal Gaza: Saya Berharap Netanyahu Dibakar di Neraka https://t.co/dvwNvautjI #IsraelisATerrorist #FREEMPALESTINE E #HalaMadrid	Politisi Irlandia Menangis saat Pidato soal Gaza Saya Berharap Netanyahu Dibakar di Neraka

Selanjutnya *case folding* mengubah huruf capital menjadi *lowcase*.

Tabel 2 Hasil *Case Folding*

Sebelum <i>case folding</i>	Sesudah <i>case folding</i>
MEMPERBARUI Pidato Lengkap Juru Bicara BRIGADE MUJAHIDIN ABU BILAL 9 FEBRUARI 2024 DALAM NAMA TUHAN YANG MAHA PEMURAH DAN MAHA PENYAYANG	memperbarui pidato lengkap juru bicara brigade mujahidin abu bilal 9 februari 2024 dalam nama tuhan yang maha pemurah dan maha penyayang allah akan

ALLAH AKAN MEMBANTU SIAPA PUN YANG BERJUANG DI JALANNYA DAN ALLAH PUNYA KEMENANGAN MUTLAK	membantu siapa pun yang berjuang di jalannya dan allah punya kemenangan mutlak
---	--

Setelah itu, normalisasi teks dengan mengganti kata-kata tidak baku atau singkatan menjadi bentuk standar.

Tabel 3 Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
penembakan ibu dan anaknya di lingkungan sheikh radwan oleh sniper zionis laknatullah yang juga terus menembak siapa saja yg berusaha menyelamatkan menurut aljazeera sang ibu akhirnya syahid dan anaknya berhasil bertahan hidup.	penembakan ibu dan anaknya dilingkungan sheikh radwan oleh sniper zionis laknatullah yang juga terus menembak siapa saja yang berusaha menyelamatkan menurut aljazeera sang ibu akhirnya syahid dan anaknya berhasil bertahan hidup

Kemudian tokenisasi, yaitu memecah kalimat menjadi potongan-potongan kata (token) agar dapat dianalisis secara terpisah, lihat Tabel 4.

Tabel 4 Hasil Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
penembakan ibu dan anaknya di lingkungan sheikh radwan oleh sniper zionis laknatullah yang juga terus menembak siapa saja yg berusaha menyelamatkan menurut aljazeera	[penembakan, ibu, dan, anaknya, di, lingkungan, sheikh, radwan, oleh, sniper, zionis, laknatullah, yang, juga, terus, menembak, siapa, saja, yang, berusaha,

sang ibu akhirnya syahid dan anaknya berhasil bertahan hidup.	menyelamatkan, menurut, aljazeera, sang, ibu, akhirnya, syahid, dan, anaknya, berhasil, bertahan, hidup]
---	--

Selanjutnya token yang tidak memberikan makna penting dalam konteks sentimen dihapus melalui proses *stopword removal*, lihat Tabel 5.

Tabel 5 Hasil Stopword Removal

Sebelum Stopword	Sesudah Stopword
[penembakan, ibu dan, anaknya, di lingkungan, sheikh radwan, oleh, sniper zionis, laknatullah yang, juga, terus menembak, siapa saja, yang, berusaha menyelamatkan, menurut, aljazeera sang, ibu, akhirnya syahid, dan anaknya, berhasil bertahan, hidup]	penembakan anaknya lingkungan sheikh radwan sniper zionis laknatullah berusaha menyelamatkan aljazeera sang syahid anaknya berhasil bertahan hidup

Tahapan terakhir adalah *stemming*, dengan tujuan mengubah setiap kata kedalam bentuk dasarnya, lihat Tabel 6.

Tabel 6 Hasil Stemming

Sebelum Stemming	Sesudah Stemming
penembakan anaknya lingkungan sheikh radwan sniper zionis laknatullah menembak berusaha menyelamatkan aljazeera sang syahid anaknya berhasil bertahan hidup	tembak anak lingkung sheikh radwan sniper zionis laknatullah tembak usaha selamat aljazeera sang syahid anak hasil tahan hidup

Pelabelan Lexicon Based

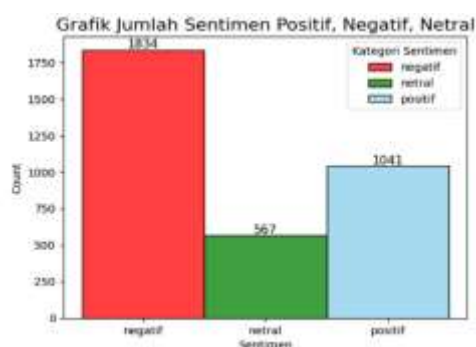
Proses pelabelan data menggunakan pendekatan *lexicon-based* berhasil mengelompokkan *tweet* terkait

isu *Free Palestine* kedalam tiga kategori sentimen yaitu positif, negatif, dan netral. *Lexicon* yang digunakan dalam penelitian ini adalah *IndoLexicon* yang berisi daftar kata berbahasa Indonesia yang telah diberi skor polaritas. Lihat hasil pada Tabel 7.

Tabel 7 Pelabelan *Lexicon Based*

No	Tweet	Label
1	semoga allah melindungi rakyat palestina doa kami selalu bersama kalian	Positif
2	penembakan warga sipil jelas pelanggaran tapi dunia diam	Negatif
3	konflik digaza terus berlanjut, pbb menyerukan gencatan senjata	Netral
4	rakyat palestina luar biasa kuat menghadapi kekejaman seperti ini	Positif
5	israel tidak punya hati nurani anak-anak dibom tanpa alasan	Negatif

Berikut visualisasi hasil pelabelan data menggunakan *Lexicon Based*.



Gambar 3 Grafik Pelabelan

Pengujian *Naïve Bayes* Individu dan *Hybrid Naïve Bayes-XGBoost*

Pengujian dilakukan untuk membandingkan kinerja antara model *Naïve Bayes* individu dan model *hybrid Naïve Bayes-XGBoost* dalam melakukan klasifikasi sentimen. Dataset telah melalui tahap *preprocessing*, pelabelan, dan pembagian data latih dan data uji dengan rasio 80:20. Hasil pengujian pada Gambar 4 dan Gambar 5.

Akurasi *Naïve Bayes* : 75.762 %

Classification Report:				
	precision	recall	f1-score	support
negatif	0.79	0.88	0.83	386
netral	0.76	0.37	0.50	111
positif	0.69	0.73	0.71	192
accuracy			0.76	689
macro avg	0.75	0.66	0.68	689
weighted avg	0.76	0.76	0.74	689

Gambar 4. Akurasi *Naïve Bayes* Individu

Akurasi *Naïve Bayes + XGBoost* : 95.536 %

Classification Report:				
	precision	recall	f1-score	support
negatif	0.78	0.94	0.85	386
netral	0.76	0.51	0.61	111
positif	0.88	0.66	0.75	192
accuracy			0.96	689
macro avg	0.90	0.91	0.94	689
weighted avg	0.90	0.90	0.89	689

Gambar 5 Akurasi *Hybrid Naïve Bayes-XGBoost*

Evaluasi Model

Tahap evaluation ini, hasil perhitungan dari *Naïve Bayes* Individu akan dibandingkan dengan mode hybrid *Naïve Bayes-XGBoost* (NB-XGBoost) berdasarkan *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Tabel 8 Perbandingan Model

Model	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
<i>Naïve Bayes</i>	75,7	75	66	76
NB-XGBoost	95,5	91	94	96

Berdasarkan hasil evaluasi, model NB-XGBoost secara konsisten menunjukkan performa yang jauh lebih unggul dibandingkan *Naïve Bayes* individu. *Naïve Bayes* menghasilkan akurasi sebesar 75,7%, dengan precision 75%, recall 66%, dan F1-score 76%. Nilai recall yang relatif rendah menunjukkan keterbatasan *Naïve Bayes* dalam mengenali seluruh data positif secara efektif, terutama pada kelas minoritas.

Sebaliknya, model *hybrid NB-*

XGBoost mencatat peningkatan signifikan dengan akurasi mencapai 95,5%, precision 91%, recall 94%, dan F1-score 96%. Ini membuktikan bahwa integrasi probabilitas *Naïve Bayes* sebagai fitur tambahan pada XGBoost mampu memperbaiki kelemahan model dasar dalam menangkap pola non-linear dan ketidakseimbangan data.

Secara keseluruhan, model hybrid tidak hanya meningkatkan akurasi, tetapi juga memberikan keseimbangan antara precision dan recall yang lebih baik. Hal ini menunjukkan bahwa NB-XGBoost lebih andal untuk klasifikasi sentimen pada data teks, khususnya dalam konteks isu *Free Palestine* yang cenderung memuat opini dengan ekspresi beragam.

SIMPULAN

Penelitian ini menyimpulkan bahwa model *hybrid Naïve Bayes-XGBoost* memiliki performa yang lebih baik dibandingkan model *Naïve Bayes* individu (dasar) dalam klasifikasi sentimen netizen terhadap isu *Free Palestine* di platform X. Model *hybrid* mencapai akurasi 95,5% dan *F1-score* 96%, menunjukkan efektivitas integrasi dua algoritma dalam meningkatkan hasil klasifikasi pada data tidak terstruktur. Pendekatan *lexicon-based* juga terbukti efisien untuk pelabelan awal data sentimen.

Sebagai saran, penelitian selanjutnya dapat menguji model ini pada data lintas *platform* dan menerapkan metode *deep learning* untuk perbandingan lebih lanjut. Selain itu, pengembangan sistem monitoring sentimen secara *real-time* juga dapat menjadi inovasi aplikasi dari penelitian ini.

DAFTAR PUSTAKA

Attai, K., Asuquo, D., Okonny, K. E., Johnson, E. A., Bassey, A., John, A., Bardi, I., Iroanwusi, C., & Michael, O. (2024). Sentiment

Analysis of Twitter Discourse on the 2023 Nigerian General Elections. *European Journal of Computer Science and Information Technology*, 12(4), 18–35. <https://doi.org/10.37745/ejcsit.2013/vol12n41835>

Burnwal, Y., & Jaiswal, Dr. R. C. (2023). A Comprehensive Survey on Prediction Models and the Impact of XGBoost. *International Journal for Research in Applied Science and Engineering Technology*, 11(12), 1552–1556. <https://doi.org/10.22214/ijraset.2023.57625>

Cerrahoğlu, E., & Cihan, P. (2023). Sentiment Analysis and Emojification of Tweets. *International Conference on Pioneer and Innovative Studies*, 481–486. <https://doi.org/10.59287/icpis.876>

Gangwar, A., & Mehta, T. (2022). Sentiment Analysis of Political Tweets for Israel using Machine Learning. *Springer Proceedings in Mathematics and Statistics*, 1–10. <https://doi.org/https://doi.org/10.48550/arXiv.2204.06515>

Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in Information Retrieval* (pp. 345–359). Springer Berlin Heidelberg.

Hafizah, R., Saragih, T. H., Muliadi, M., Indriani, F., & Mazdadi, M. I. (2025). Machine Learning Implementation for Sentiment Analysis on X/Twitter: Case Study of Class Of Champions Event in Indonesia. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(2), 370–386. <https://doi.org/10.35882/ijeemi.v7i2.81>

Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text

- Classification. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 143–149.
<https://doi.org/10.29408/edumatic.v6i1.5613>
- Irwanto, A., & Goeirmanto, L. (2023). Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia using Naïve Bayes and XGboost Classifier Algorithm. *Sinergi*, 27(2), 145–152.
<https://doi.org/10.22441/sinergi.2023.2.001>
- Jonnala, N. S., Ram Teja, A. V. S., Rajeswari, S. R., Jakeer, S., Dheeraj, A., Bansal, S., Prakash, K., Singh, S., Faruque, M. R. I., & Al-mugren, K. S. (2025). Leveraging hybrid model for accurate sentiment analysis of Twitter data. *Scientific Reports*, 15(1).
<https://doi.org/10.1038/s41598-025-09794-2>
- Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. In *Information (Switzerland)* (Vol. 10, Issue 4). MDPI AG.
<https://doi.org/10.3390/info10040150>
- Liu, D. (2023). Improvement of Naive Bayes Text Classifier Based on Ensemble Technology and Feature Engineering. *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence*, 557–563.
https://doi.org/10.2991/978-94-6463-300-9_57
- Mohammed, I., & Prasad, R. (2023). Building lexicon-based sentiment analysis model for low-resource languages. *MethodsX*, 11.
<https://doi.org/10.24843/LKJITI.2023.v14.i03.p05>
- Negara, A. B. P. (2023). The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification. *LONTAR KOMPUTER*, 14(3), 172–185.
<https://doi.org/10.24843/LKJITI.2023.v14.i03.p05>
- Nugraha, I. G. B. B., & Rizqullah, R. D. (2019). Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(3), 218–224.
- Rianto, Mutiara, A. B., Wibowo, E. P., & Santosa, P. I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, 8(1).
<https://doi.org/10.1186/s40537-021-00413-1>
- Saputra, F. T., Wijaya, S. H., Nurhadryani, Y., & Defina. (2020). Lexicon Addition Effect on Lexicon-Based of Indonesian Sentiment Analysis on Twitter. *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, 136–141.
<https://doi.org/10.1109/ICIMCIS51567.2020.9354269>