UNIVERSITAS
MUHAMMADIYAH
SIDOARJO

UMSIDA
DARI SINI PENCERAHAN BERSEMI

Implementasi Arsitektur Hybrid CNN-LSTM untuk Deteksi Cyberbullying pada Komentar Berbahasa Indonesia di Media Sosial X

Oleh :

Muhammad Hafizh Fattah

Dosen Pembimbing:

Dr. Mochamad Alfan Rosid, M.Kom.

Dosen Penguji 1:

Sukma Aji, S.Kom., M.T.

Dosen Penguji 2:

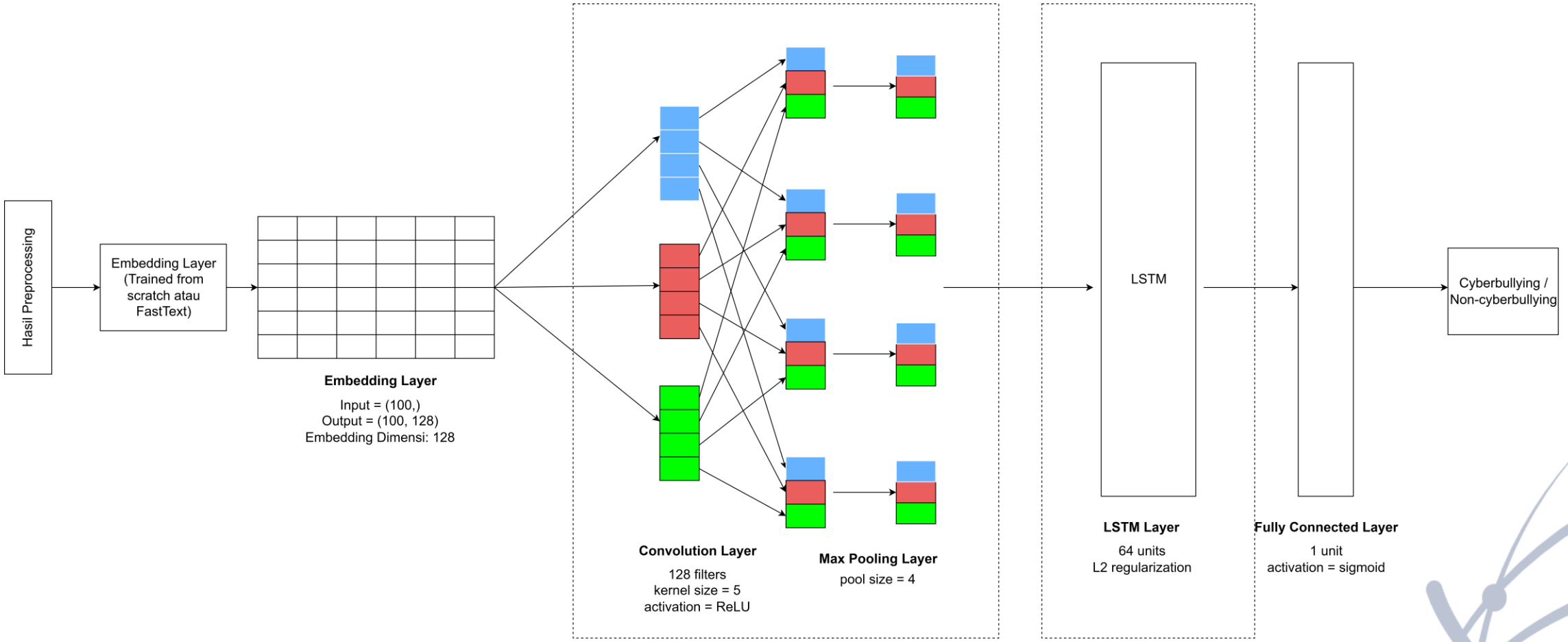
Dr.Suprianto, S.Si. M.Si.

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

2026

Arsitektur Model



Tahap Pengujian

Penelitian ini merancang tiga skenario pengujian untuk mengevaluasi kinerja model CNN–LSTM. Dataset dibagi menjadi data latih, validasi, dan uji. Skenario pertama menggunakan EarlyStopping sebagai baseline, skenario kedua tanpa EarlyStopping untuk melihat potensi overfitting, dan skenario ketiga berupa studi komparatif lanjutan dengan FastText embedding, model IndoBERTweet, serta optimasi hyperparameter menggunakan BayesianOptimization.

Evaluasi Model

Model dievaluasi menggunakan data uji untuk mengukur kemampuan generalisasi. Analisis dilakukan secara kuantitatif dengan Confusion Matrix serta metrik evaluasi (Accuracy, Precision, Recall, F1-Score, dan AUC). Selain itu, analisis kualitatif juga dilakukan dengan meninjau contoh komentar yang salah diklasifikasikan (error analysis).

Hasil & Pembahasan

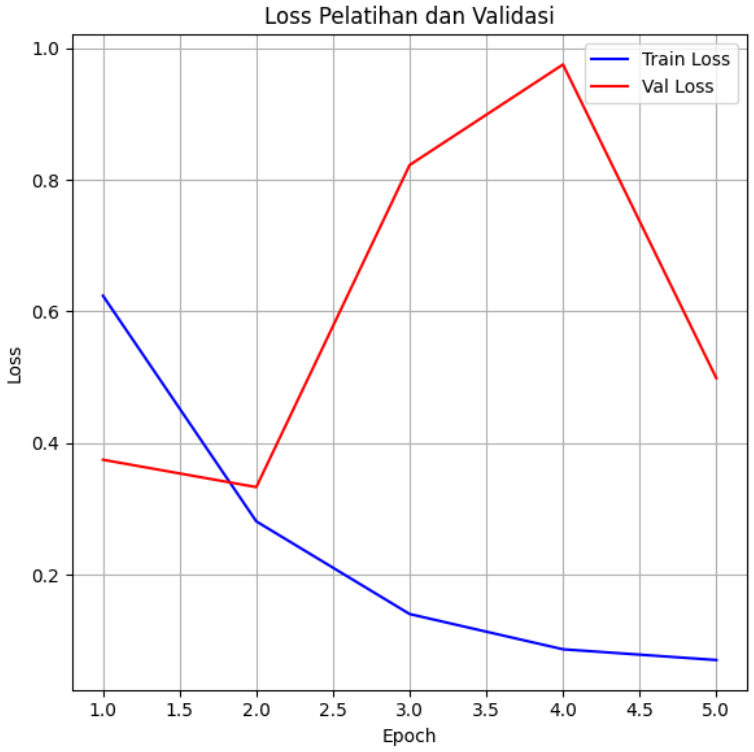
Skenario pertama

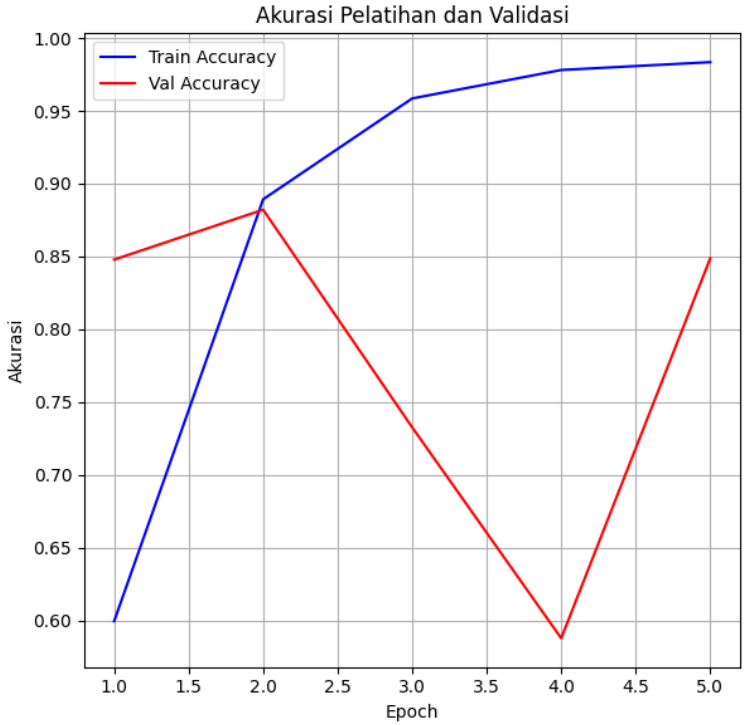
Tabel 3. Konfigurasi Model Dasar CNN-LSTM

Pengguna	Teks
Embedding Layer	Trainable, dimensi vektor kata = 128, dilatih dari nol
Conv1D Layer	128 filter, ukuran kernel = 5
BatchNormalization	Digunakan untuk stabilitas pembelajaran
Activation	ReLU

Hasil & Pembahasan

Skenario pertama

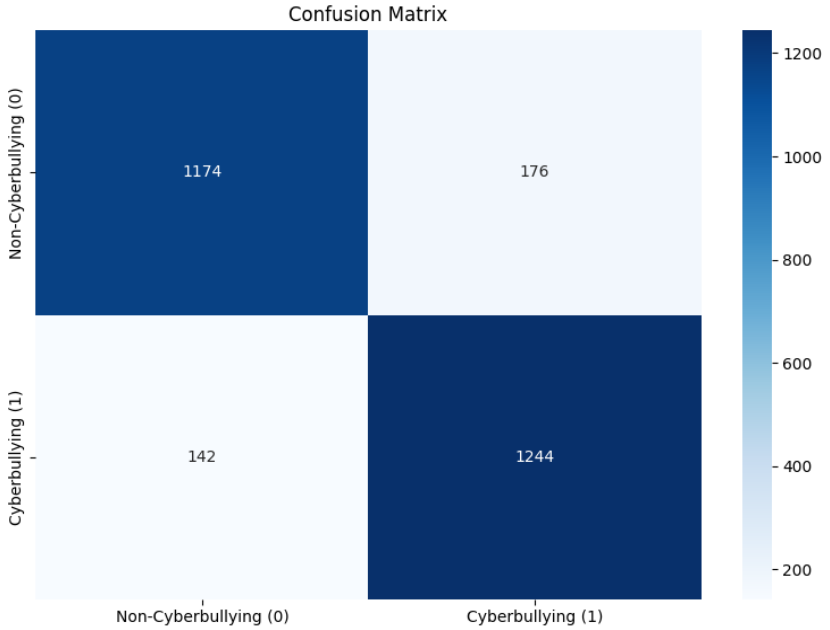


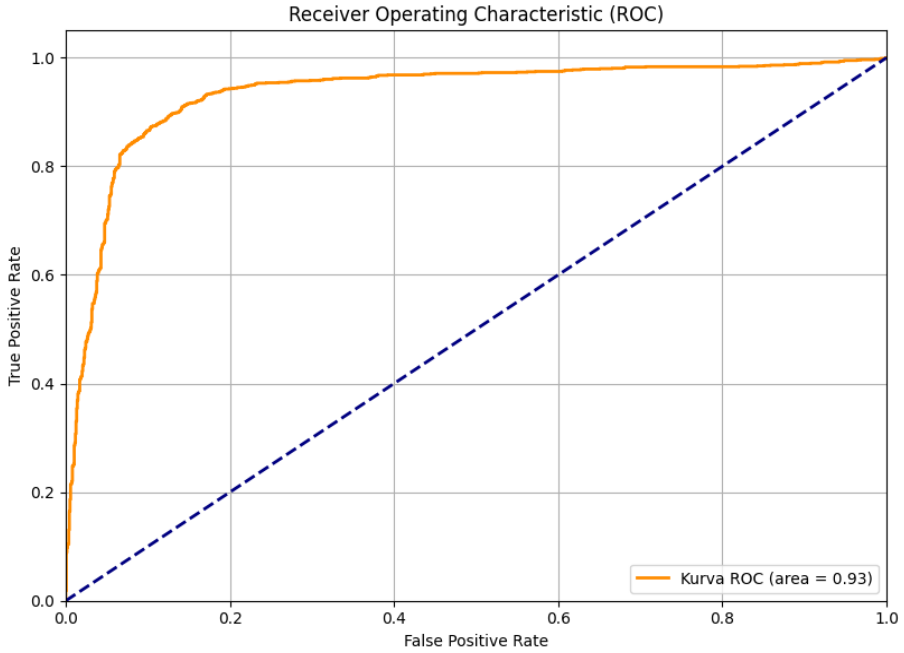


Model mencapai konvergensi optimal pada epoch ke-5 di mana mekanisme Early Stopping berhasil menghentikan pelatihan tepat sebelum validasi loss meningkat, mencegah terjadinya overfitting sejak dini dan menjaga stabilitas generalisasi model.

Hasil & Pembahasan

Skenario pertama



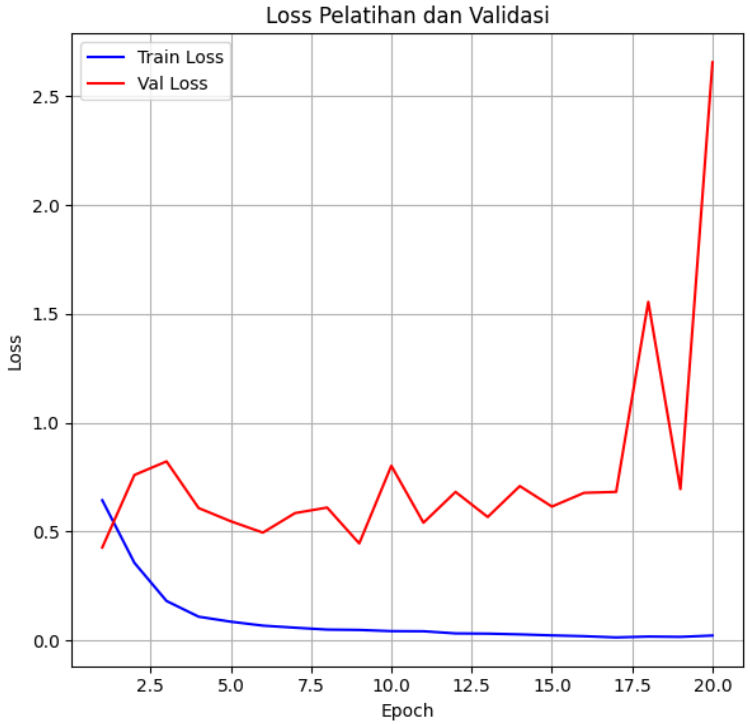


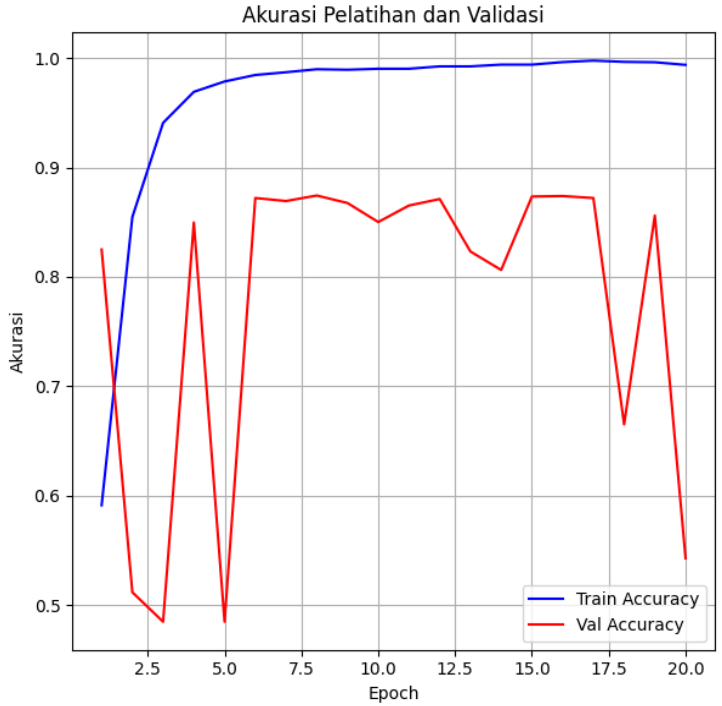
Tabel 5. Rincian Hasil Evaluasi Model CNN-LSTM

Metrik	Nilai
Akurasi	0.8838
Presisi	0.88
Recall	0.88
F1 Score	0.88
AUC	0.9346

Hasil & Pembahasan

Skenario kedua

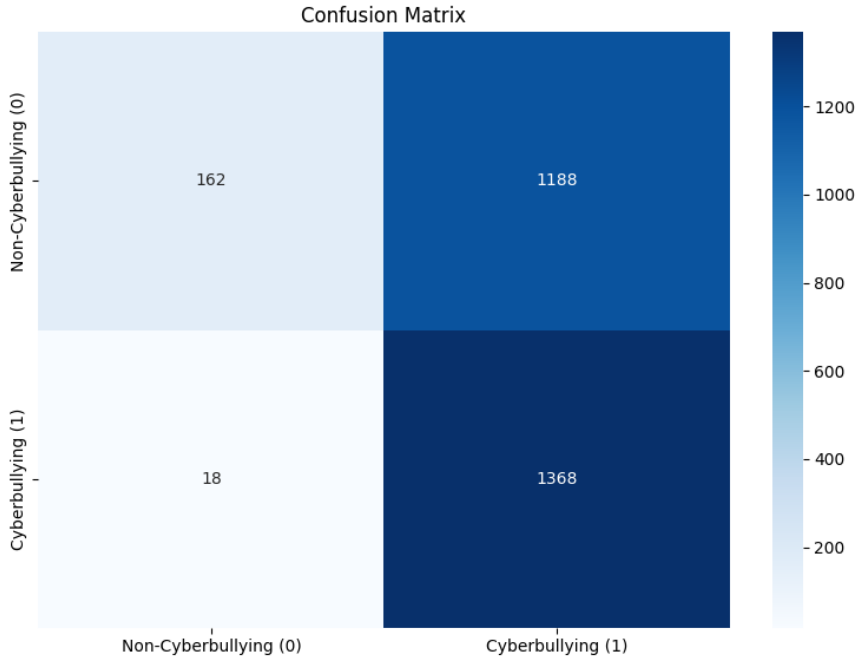


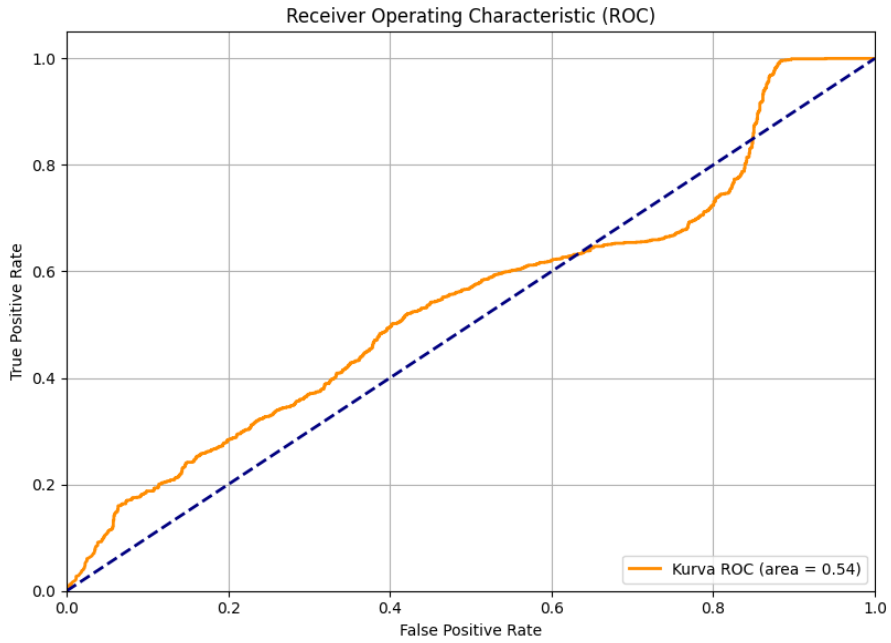


Tanpa mekanisme Early Stopping, model mengalami overfitting parah mulai epoch ke-7 yang ditandai dengan divergensi ekstrem antara akurasi pelatihan yang mendekati sempurna dan akurasi validasi yang anjlok drastis hingga 55%.

Hasil & Pembahasan

Skenario kedua





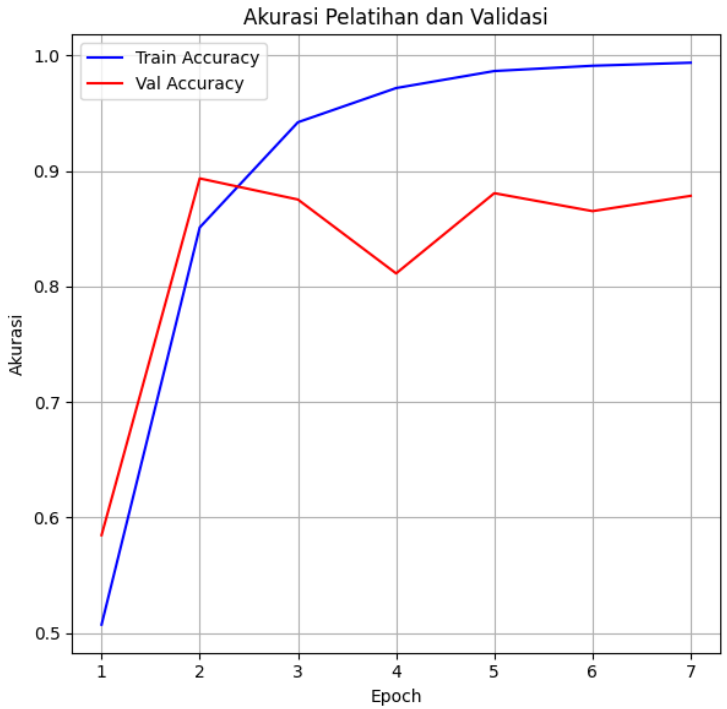
Tabel 6. Perbandingan Model Dengan Dan Tanpa Early Stopping

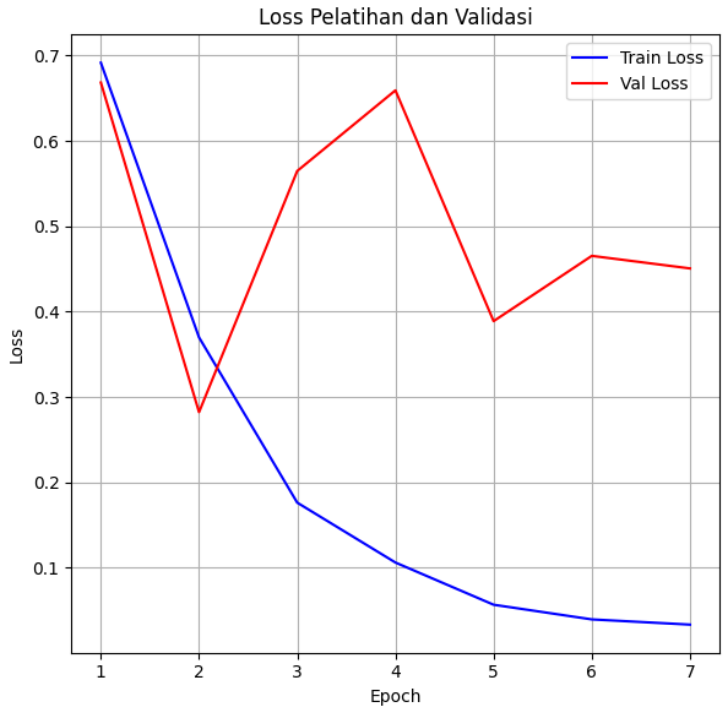
Model	Akurasi	Presisi	Recall	F1 Score	AUC
CNN-LSTM (Baseline)	0.8838	0.88	0.88	0.88	0.9346
CNN-LSTM (Tanpa Early Stopping)	0.5592	0.72	0.56	0.46	0.5266

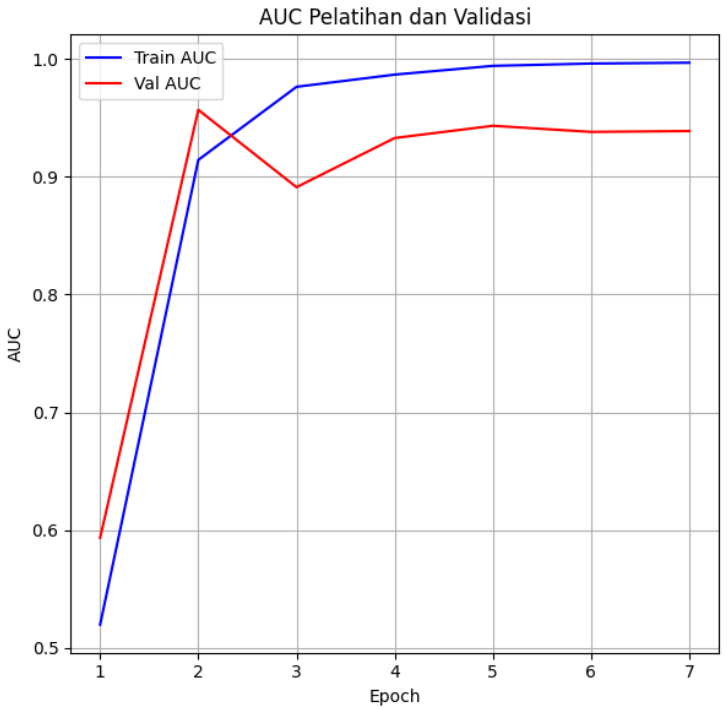
Hasil & Pembahasan

Skenario ketiga

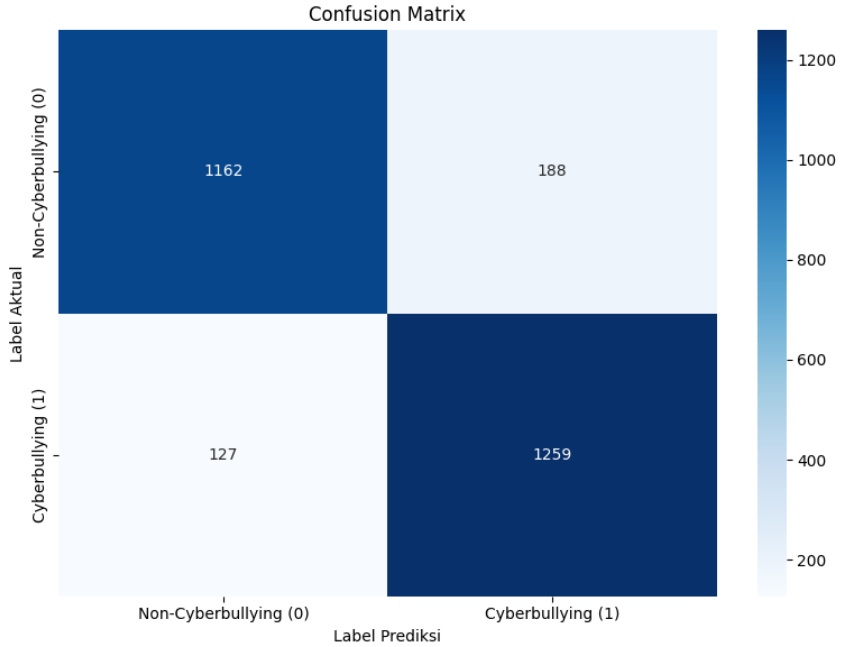
1) Optimasi Dengan Fasttext

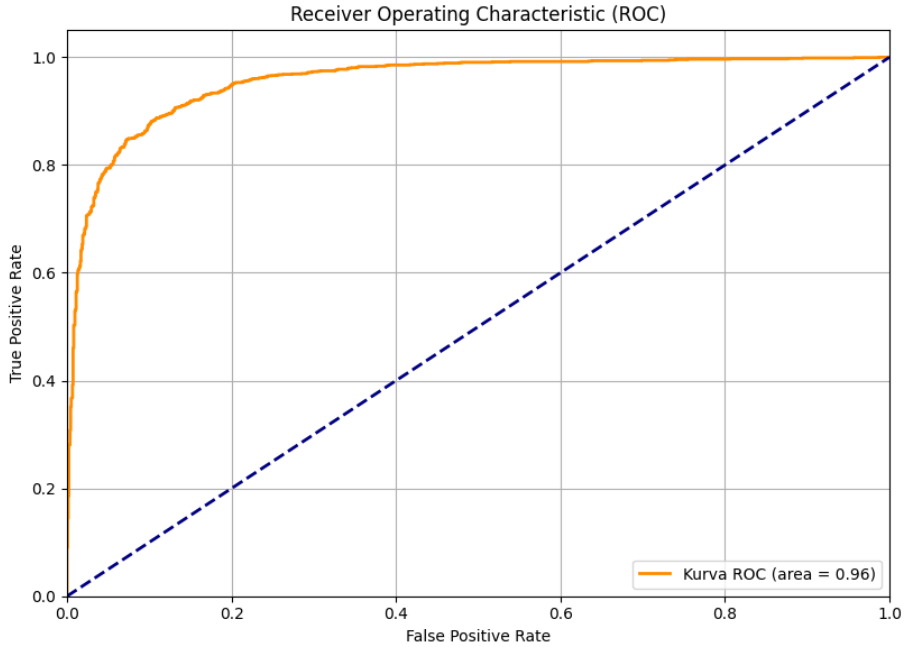






Hasil & Pembahasan





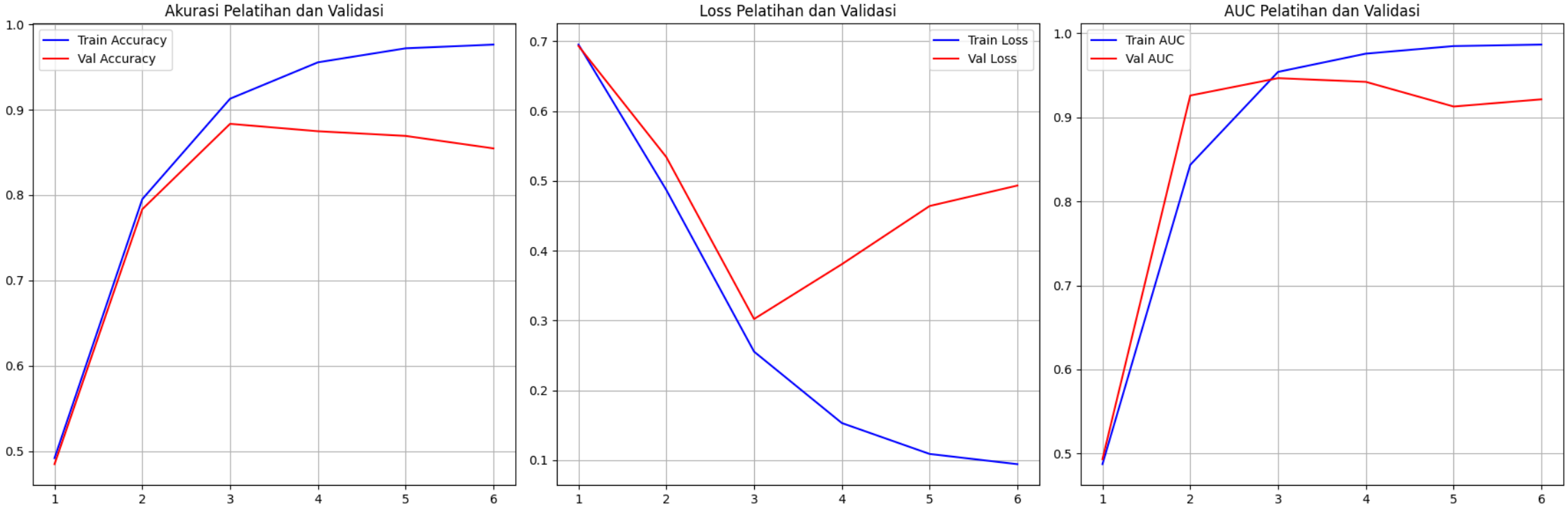

```
Evaluasi pada data uji:
Loss: 0.2867, Akurasi: 0.8849 (88.49%), AUC: 0.9559
```

Laporan Klasifikasi:

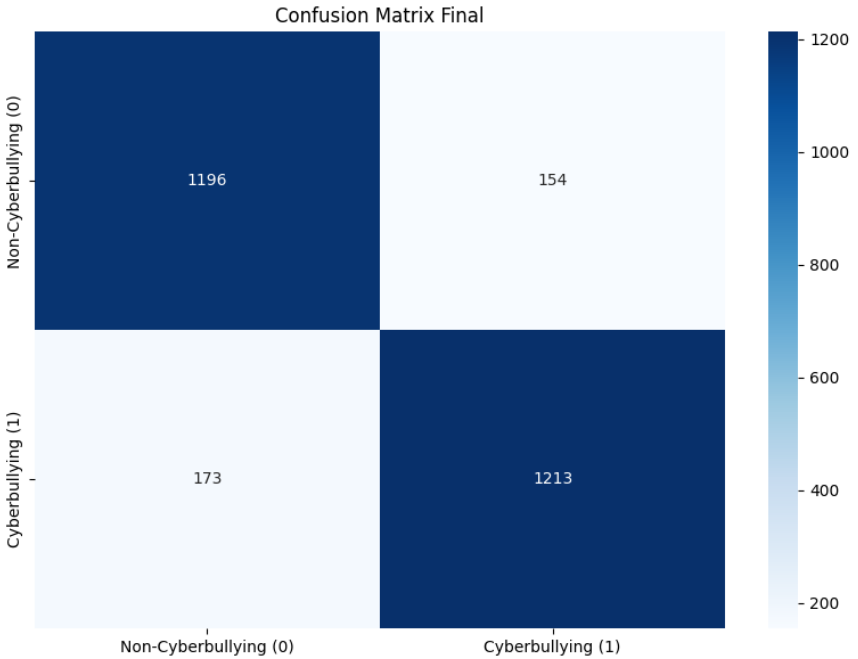
	precision	recall	f1-score	support
Non-Cyberbullying (0)	0.90	0.86	0.88	1350
Cyberbullying (1)	0.87	0.91	0.89	1386
accuracy			0.88	2736
macro avg	0.89	0.88	0.88	2736
weighted avg	0.89	0.88	0.88	2736

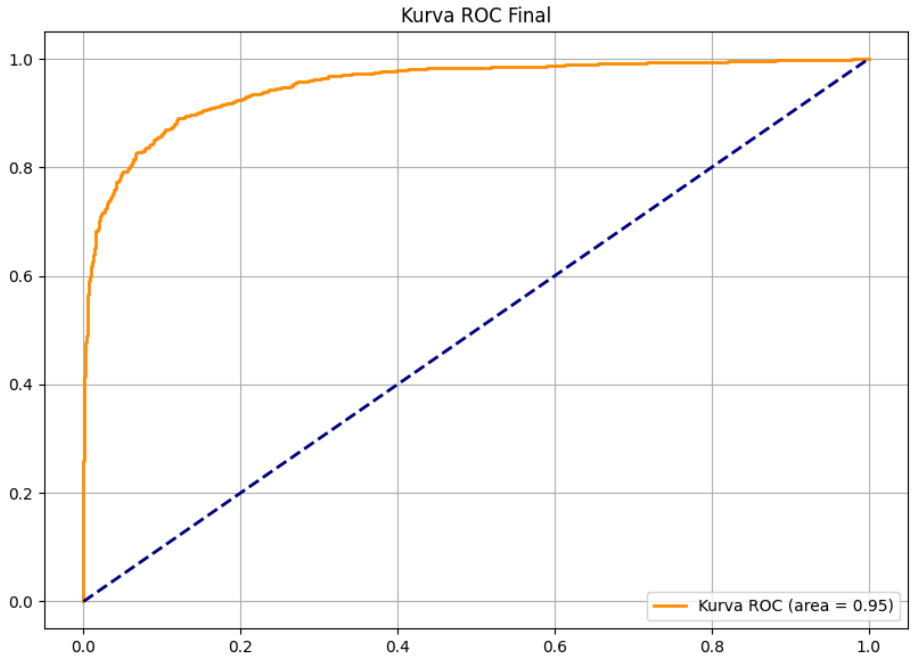
Hasil & Pembahasan

2) Optimasi hyperparameter



Hasil & Pembahasan






```
Hasil pada Data Uji:
Loss: 0.2936
Akurasi: 0.8805 (88.05%)
AUC: 0.9507
```

```
Laporan Klasifikasi Final:
                    precision    recall  f1-score   support

Non-Cyberbullying (0)      0.87      0.89      0.88      1350
    Cyberbullying (1)      0.89      0.88      0.88      1386

             accuracy                           0.88      2736
            macro avg      0.88      0.88      0.88      2736
         weighted avg      0.88      0.88      0.88      2736
```

Hasil & Pembahasan

3) Benchmarking Dengan Indobertweet

Tabel 8. Rekapitulasi Kinerja Komprehensif Seluruh Model

Model	Akurasi	Presisi	Recall	F1-Score	AUC
IndoBERTweet	**0.915**	**0.92**	**0.92**	**0.915**	-
BiLSTM + Attention	0.8871	0.8945	0.8810	0.8877	0.9577
CNN-LSTM (+ FastText)	0.8849	0.89	0.88	0.88	0.9559
CNN	0.8841	0.9071	0.8593	0.8825	0.9566
CNN-BiGRU	0.8834	0.8834	0.8834	0.8834	0.9566
LSTM	0.88	0.88	0.88	0.88	0.9596
CNN-BiLSTM	0.8816	0.8820	0.8816	0.8815	0.9548
CNN-LSTM (Base Line Early Stopping)	0.8838	0.88	0.88	0.88	0.9346
CNN-LSTM (Keras Tuner)	0.8805	0.88	0.88	0.88	0.9507
SVM + TF-IDF	0.7818	0.8266	0.7818	0.7747	-
CNN-LSTM (Tanpa EarlyStopping)	0.5592	0.72	0.56	0.46	0.5266

Analisis Kualitatif

Tabel 9. Contoh Kesalahan Prediksi Model CNN-LSTM Pada Data Uji

No	Teks Komentar (Asli)	Label Asli	Prediksi Model	Status Kesalahan
1	masuk tan anjing	Non-Bullying (0)	Bullying (1)	False Positive
2	goodness parent actually watching something apple tv brilliant yet another streaming service	Non-Bullying (0)	Bullying (1)	False Positive
3	aduh lak onok rawan disalahgunakan men	Non-Bullying (0)	Bullying (1)	False Positive
4	jomok bawa barak iya oppa kdm iya anjr panggil sayang mau pacar panggil sayang first tonton mv kambek jeoang skz ya aju main vocalnya oek	Bullying (1)	Non-Bullying (0)	False Negative
5	halo non moot ijin ikut ya selamat lolo seleksi kerja tysm moga halus pity sylus nya bandel ya	Non-Bullying (0)	Bullying (1)	False Positive

Kesimpulan

Peran Early Stopping: Mencegah overfitting fatal (Akurasi stabil di 88,38% vs 55,92%).

Dampak FastText: Meningkatkan kemampuan diskriminatif model (**AUC naik ke 0,9559**).

Efisiensi Sumber Daya: Mencapai **97% kinerja IndoBERTweet** dengan ukuran model **22x lebih kecil** (5 juta vs 110 juta parameter).

Superioritas Model: Mengungguli metode klasik (SVM) dengan selisih akurasi **>10%**.

Referensi

- Ahmad, S. R., Insani, N., & Salim, M. (2024). Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms. Jurnal Teknologi Informasi Dan Pendidikan, 17(1), 75–86. https://doi.org/10.24036/jtip.v17i1.807
- Akar, F. (2024). Performance Analysis of NLP-Based Machine Learning Algorithms in Cyberbullying Detection. Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 17(2), 445–459. https://doi.org/10.18185/erzifbed.1474112
- Akter, F., Rabbi, F., & Chowdhury, R. R. (2025). Cyberbullying Detection on Social Media Platforms Utilizing Different Machine Learning Approaches. 186(61), 40–50.
- Alqahtani, A. F., & Ilyas, M. (2024). An Ensemble-Based Multi-Classification Machine Learning Classifiers Approach to Detect Multiple Classes of Cyberbullying. Machine Learning and Knowledge Extraction, 6(1), 156–170. https://doi.org/10.3390/make6010009
- Andika, A. J., Kristian, Y., & Setiawan, E. I. (2023). Deteksi Komentar Cyberbullying Pada YouTube Dengan Metode Convolutional Neural Network – Long Short-Term Memory Network (CNN-LSTM). Teknika, 12(3), 183–188. https://doi.org/10.34148/teknika.v12i3.677
- Asqolani, I. A., & Setiawan, E. B. (2023). A Hybrid Deep Learning Approach Leveraging Word2Vec Feature Expansion for Cyberbullying Detection in Indonesian Twitter. Ingenierie Des Systemes d'Information, 28(4), 887–895. https://doi.org/10.18280/isi.280410
- El Koshiry, A. M., Eliwa, E. H. I., El-Hafeez, T. A., & Khairy, M. (2024). Detecting cyberbullying using deep learning techniques: a pre-trained glove and focal loss technique. PeerJ Computer Science, 10, 1–33. https://doi.org/10.7717/peerj-cs.1961
- Faritha Banu, J., Rajeshwari, S. B., Kallimani, J. S., Vasanthi, S., Buttar, A. M., Sangeetha, M., & Bhargava, S. (2022). Modeling of Hyperparameter Tuned Hybrid CNN and LSTM for Prediction Model. Intelligent Automation and Soft Computing, 33(3), 1393–1405. https://doi.org/10.32604/iasc.2022.024176
- Hafiza, A. A., & Setiawan, E. B. (2025). Enhancing Cyberbullying Detection on Platform "X" Using IndoBERT and Hybrid CNN-LSTM Model. Jurnal Teknik Informatika (Jutif), 6(2), 655–672. https://doi.org/10.52436/1.jutif.2025.6.2.4321
- Harish, D., Manimaran, M., Jayashakthi, V. P., & Alamelu, M. (2023). Automatic Detection of Cyberbullying on Social Media Using Machine Learning. 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2023, 13(3), 185–189. https://doi.org/10.1109/ICAECA56562.2023.10201149
- Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A Review on Deep-Learning-Based Cyberbullying Detection. Future Internet, 15(5), 1–47. https://doi.org/10.3390/fi15050179

Referensi

- Iryani, L., Junaidi, J., Paisal, P., Purba, M., Umilizah, N., Bakhtiar, B., & Ani, N. (2025). Model of Cyberbullying Detection on Social Media Using Multi-Label Deep Learning: a Comparative Study. JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer), 10(4), 742–748. https://doi.org/10.33480/jitk.v10i4.6004
- Masbadi Hatullah Nurnaryo, R., Mulaab, M., Oktavia Suzanti, I., Abdul Fatah, D., Cahyani, A. D., & Ayu Mufarroha, F. (2022). Deteksi Cyberbullying Pada Data Tweet Menggunakan Metode Random Forest Dan Seleksi Fitur Information Gain. Jurnal Simantec, 11(1), 33–40. https://doi.org/10.21107/simantec.v11i1.17256
- Maulana, M. D., & Aditya, C. S. K. (2025). Perbandingan IndoBERT dan Bi-LSTM Dalam Mendeteksi Pelanggaran Undang-Undang ITE. Science And Information Technology (SINTECH), 8(1), 52–59. https://doi.org/10.31598
- Mitra, S., Tasnim, T., Islam, M. A. R., Khan, N. I., & Majib, M. S. (2021). A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms. 6th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2021, February 2023. https://doi.org/10.1109/IC4ME253898.2021.9768450
- Nasution, M. A. S., & Setiawan, E. B. (2023). Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging FastText for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM. Revue d'Intelligence Artificielle, 37(4), 929–936. https://doi.org/10.18280/ria.370413
- Prama, T. T., Amrin, J. F., Anwar, Md. M., & Sarker, I. H. (2025). AI Enabled User-Specific Cyberbullying Severity Detection with Explainability. 1–25. http://arxiv.org/abs/2503.10650
- Purkayastha, B. S., Rahman, M., & Talukdar, T. I. (2025). Advancing Cyberbullying Detection : A Hybrid Machine Learning and Deep Learning Framework for Social Media Analysis. 2(Iceis), 348–355. https://doi.org/10.5220/0013436200003929
- Rosid, M. A., Siahaan, D., & Saikhu, A. (2024). Sarcasm Detection in Indonesian-English Code-Mixed Text Using Multihead Attention-Based Convolutional and Bi-Directional GRU. IEEE Access, June, 137063–137079. https://doi.org/10.1109/ACCESS.2024.3436107
- Santoso, H., Putri, R. A., & Sahbandi, S. (2023). Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest. Jurnal Manajemen Informatika (JAMIKA), 13(1), 62–72. https://doi.org/10.34010/jamika.v13i1.9303
- Smote, C. D. I. (2025). Analisis komparatif klasifikasi sentimen pengguna aplikasi investasi menggunakan algoritma hybrid cnn-lstm, cnn-gru dengan implementasi smote. 6, 112–118.
- Tapaskar, V., & J, M. C. T. (2024). INTERNATIONAL JOURNAL OF A Multi-Algorithm Approach for Cyberbullying Detection. 7(10). https://doi.org/10.15680/IJMRSET.2024.0710027

Sekian,
Terimakasih

UMSIDA
DARI SINI PENCERAHAN BERSEMI