

Implementasi Sistem Prediksi Saham IDX80 Berbasis Berita Ekonomi Berbahasa Indonesia Menggunakan LSTM dan Framework Streamlit

Muhamad Hernan Fauzan¹, Syariful Alam², Chandra Dewi Lestari³

¹Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Indonesia

^{2,3}Program Studi Teknik Informatika, STT Wastukencana Purwakarta, Indonesia

Email: ¹hfauzan47@gmail.com, ²syarifulalam@wastukencana.ac.id, ³candradewi@wastukencana.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah Masuk, 9 Juli 2026

Direvisi, 14 Juli 2026

Diterima, 31 Juli 2026

ABSTRAK

Abstract- Manual sentiment analysis of economic news for stocks included in the IDX80 index poses a significant challenge for retail investors due to the massive volume of information and delayed market responses. This study aims to build a sentiment classification system for Indonesian-language economic news articles related to IDX80 stocks using the Long Short-Term Memory (LSTM) method to provide an objective and agile decision-making tool. The methodology follows a supervised learning approach based on the CRISP-DM framework. The dataset consists of 3,000 news articles processed through text preprocessing stages such as case folding, cleansing, stopword removal, tokenizing, and stemming using the Sastrawi library to handle complex Indonesian morphology. The model architecture includes a 128-dimensional embedding layer, a 128-unit LSTM layer, 0.5 dropout to prevent overfitting, and a Softmax activation function for three-class classification. Test results demonstrate that the LSTM model successfully achieved a peak validation accuracy of 92.67%, with the training process automatically stopping at the 7th epoch via EarlyStopping. The research contribution is realized through the implementation of a web-based application using Streamlit, which supports batch prediction features via CSV files. This system is proven effective in capturing long-term dependencies in financial text word sequences, providing a responsive investment strategy support instrument.

Kata Kunci:

Long Short-Term Memory (LSTM)

Analisis Sentimen

IDX80

Berita Ekonomi

Deep Learning

Abstrak- Analisis sentimen manual terhadap berita ekonomi saham yang tergabung dalam indeks IDX80 menjadi tantangan besar bagi investor ritel karena volume informasinya yang sangat masif serta keterlambatan respon pasar. Penelitian ini bertujuan membangun sistem klasifikasi sentimen berita ekonomi berbahasa Indonesia terkait saham IDX80 menggunakan metode Long Short-Term Memory (LSTM) guna menyediakan alat bantu pengambilan keputusan yang objektif dan lincah. Metodologi yang digunakan adalah pendekatan *supervised learning* berdasarkan kerangka kerja CRISP-DM. Dataset terdiri dari 3000 artikel berita yang diproses melalui tahapan pra-pemrosesan teks seperti *case folding*, *cleansing*, *stopword removal*, *tokenizing*, dan *stemming* menggunakan pustaka Sastrawi untuk menangani morfologi bahasa Indonesia yang kompleks. Arsitektur model mencakup *embedding layer* berdimensi 128, lapisan LSTM 128 unit, *dropout* 0,5 untuk mencegah *overfitting*, serta fungsi aktivasi *Softmax* untuk klasifikasi tiga kelas sentimen. Hasil pengujian menunjukkan bahwa model LSTM berhasil mencapai akurasi validasi tertinggi sebesar 92,67% dengan penghentian otomatis pada *epoch* ke-7 menggunakan fitur *EarlyStopping*. Kontribusi penelitian ini diwujudkan melalui implementasi aplikasi berbasis web menggunakan *Streamlit* yang mendukung fitur prediksi *batch* via file *CSV*. Sistem ini terbukti efektif dalam menangkap

ketergantungan jangka panjang pada urutan kata teks keuangan sehingga mampu menjadi instrumen pendukung strategi investasi yang responsif di tengah ketidakpastian pasar modal.

Copyright © 2026 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Muhamad Hernan Fauzan
Program Studi Teknik Informatika,
STT Wastukencana Purwakarta
Jl. Cikopak No. 53, Purwakarta 41151, Jawa Barat, Indonesia
Email: hfauzan47@gmail.com

1. Pendahuluan

Investasi saham di Indonesia saat ini telah menunjukkan kemajuan yang sangat pesat, tercermin dari lonjakan jumlah investor ritel yang signifikan [1]. Berdasarkan data, jumlah investor meningkat tajam dari sekitar 2,48 juta orang pada tahun 2019 menjadi lebih dari 7,48 juta investor pada tahun 2021 [2]. Tren positif ini menandakan peningkatan kesadaran masyarakat dalam mengelola kekayaan melalui instrumen pasar modal guna mencapai tujuan keuangan jangka panjang. Untuk membantu investor memilih emiten di tengah ratusan pilihan, Bursa Efek Indonesia (BEI) meluncurkan indeks IDX80 pada 1 Februari 2019, yang terdiri dari 80 saham dengan likuiditas tinggi, kapitalisasi pasar besar, serta fundamental perusahaan yang kuat [3]. Indeks ini memiliki peran krusial karena mewakili sekitar 80% hingga 90% dari nilai dan volume transaksi di seluruh pasar saham Indonesia [4].

Meskipun saham-saham dalam IDX80 dikategorikan sebagai aset yang stabil, harganya tetap rentan terhadap volatilitas tajam yang dipicu oleh arus informasi dan berita ekonomi [5]. Berita ekonomi membentuk sentimen pasar, suasana hati kolektif investor yang sering kali menjadi faktor eksternal penggerak harga lebih cepat dibandingkan angka statistik murni, karena reaksi berbasis sentimen berkorelasi langsung dengan variasi harga saham [6]. Tantangan utama yang dihadapi investor ritel saat ini adalah volume berita ekonomi yang masif dan terbit secara *real-time*, sehingga sangat sulit untuk dianalisis secara manual. Evaluasi kinerja konvensional menggunakan metode kuantitatif tradisional seperti rasio *Sharpe*, *Treynor*, atau indeks *Jensen* dinilai kurang lincah karena bersifat *backward-looking* atau hanya mengevaluasi data historis masa lalu, sehingga

sering kali terlambat dalam merespons perubahan sentimen pasar yang terjadi dalam hitungan detik.

Beberapa penelitian terdahulu telah membuktikan efektivitas teknologi kecerdasan buatan dalam menangani data teks sekuensial. Nugroho dkk. (2025) menerapkan algoritma LSTM untuk analisis sentimen platform pendidikan dan mencapai akurasi sebesar 93,58% [7]. Di pasar global, Huang (2023) menunjukkan bahwa model LSTM sangat efektif dalam menangkap tren harga saham dengan akurasi mencapai 94% [8]. Namun, terdapat analisis *gap* di mana mayoritas penelitian tersebut masih menggunakan dataset berbahasa Inggris dan kurang memberikan fokus pada pasar saham domestik, khususnya indeks IDX80 yang memiliki karakteristik morfologi bahasa Indonesia yang unik dan formal di sektor keuangan.

Oleh karena itu, diperlukan teknologi cerdas yang mampu memproses informasi berita secara otomatis dan akurat. Metode *Deep Learning*, khususnya *Long Short-Term Memory* (LSTM), menjadi solusi yang sangat efektif karena kemampuan uniknya dalam mengingat urutan informasi jangka panjang dan mengatasi masalah *vanishing gradient* pada jaringan saraf sekuensial [9]. Penelitian ini bertujuan untuk membangun sistem klasifikasi sentimen berita ekonomi berbahasa Indonesia terkait saham IDX80 menggunakan metode LSTM. Kebaruan penelitian ini terletak pada penggunaan dataset spesifik berita ekonomi emiten IDX80 berbahasa Indonesia yang diintegrasikan ke dalam aplikasi berbasis *Streamlit*. Hasil penelitian ini diharapkan dapat menjadi "kompas" digital bagi investor untuk memantau polaritas sentimen berita secara objektif, sehingga mendukung pengambilan keputusan investasi yang lebih efektif dan responsif di tengah ketidakpastian pasar modal.

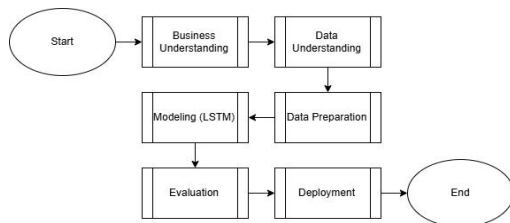
2. Metode

Penelitian ini menggunakan pendekatan *Supervised Learning* dengan kerangka kerja sistematis yang diadaptasi dari metodologi CRISP-DM, meliputi tahapan *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [10]. Penggunaan model ini menjamin alur penelitian yang terukur mulai dari identifikasi masalah investasi hingga implementasi sistem prediksi berbasis web.

2.1 Kerangka Penelitian

Alur penelitian dimulai dari pengumpulan data tekstual berita ekonomi hingga tahap implementasi model ke dalam antarmuka aplikasi web. Secara visual, tahapan sistematis ini digambarkan dalam kerangka berpikir penelitian pada Gambar 1.

Gambar 1. Diagram Alur Penelitian



2.2 Dataset dan Pelabelan

Data yang digunakan dalam penelitian ini adalah data tekstual berupa 3000 artikel berita ekonomi berbahasa Indonesia terkait emiten dalam indeks IDX80 yang dikumpulkan melalui teknik *web scraping* dari portal berita terpercaya. Proses pelabelan sentimen dilakukan berdasarkan tren perubahan harga saham pasca-berita diterbitkan atau divalidasi menggunakan kamus leksikon ekonomi yang kredibel. Dataset dibagi menjadi dua bagian: 80% untuk data latih (2400 data) dan 20% untuk data uji (600 data) guna memastikan validasi kinerja model yang objektif.

2.3 Pra-pemrosesan Teks (*Preprocessing*)

Tahap ini merupakan proses yang sangat krusial dalam analisis sentimen karena bertujuan untuk mengubah data teks yang masih bersifat tidak terstruktur menjadi representasi numerik yang dapat dipahami dan diproses oleh algoritma Deep Learning. Selain itu, proses ini juga berfungsi untuk mengurangi noise, meningkatkan kualitas data, serta memastikan informasi penting dalam teks tetap dipertahankan sehingga model dapat mempelajari pola dengan lebih efektif. Dengan melakukan text preprocessing yang baik, diharapkan performa model dalam mengklasifikasikan sentimen menjadi lebih akurat. Tahapan text preprocessing yang dilakukan dalam penelitian ini meliputi:

Tabel 1. Tahapan *preprocessing*

Tahapan	Penjelasan
<i>Case Folding</i>	Mengubah seluruh teks menjadi huruf kecil
<i>Cleansing</i>	Menghapus URL, angka, tanda baca, simbol, dan karakter khusus
<i>Stopword Removal</i>	Menghapus kata umum yang kurang memiliki makna penting
<i>Stemming</i>	Mengubah kata ke bentuk dasar menggunakan Sastrawi
<i>Label Encoding</i>	Mengubah label sentimen menjadi angka
<i>Tokenizing</i>	Mengubah teks menjadi urutan token
<i>Padding</i>	Menyamakan panjang <i>sequence</i> agar sesuai dengan <i>input</i> model

Setelah pembersihan, teks diubah menjadi urutan indeks kata (*sequence*) dan diseragamkan panjangnya menggunakan teknik *padding* dengan batas maksimum 100 token agar konsisten sebagai masukan model. Contoh transformasi data disajikan pada Tabel 2.

Tabel 2. Hasil Pra-pemrosesan Data Teks

Tahapan	Hasil
Data Mentah	"Emiten IDX80 mencatatkan kenaikan laba bersih yang signifikan di kuartal III."
<i>Case Folding</i>	"emiten idx80 mencatatkan kenaikan laba bersih yang signifikan di kuartal iii."
<i>Cleansing</i>	"emiten idx80 mencatatkan kenaikan laba bersih yang signifikan di kuartal iii"
<i>Tokenizing</i>	["emiten", "idx80", "mencatatkan", "kenaikan", "laba", "bersih", "yang", "signifikan", "di", "kuartal", "iii"]
<i>Stopword Removal</i>	["emiten", "idx80", "mencatatkan", "kenaikan", "laba", "bersih", "signifikan", "kuartal", "iii"]
<i>Stemming</i>	["emiten", "idx80", "catat", "naik", "laba", "bersih", "signifikan", "kuartal", "iii"]

2.4 Arsitektur Model LSTM

Model klasifikasi dibangun menggunakan algoritma LSTM karena kemampuannya yang unggul dalam menangkap ketergantungan jangka panjang (*long-term dependencies*) dan mengatasi masalah *vanishing gradient* pada jaringan saraf sekuensial tradisional [11]. Arsitektur model terdiri dari *Embedding Layer* (dimensi 128), lapisan

LSTM (128 unit), *Dropout* (0,5) untuk mencegah *overfitting*, dan *Dense Layer* dengan aktivasi *Softmax* untuk klasifikasi tiga kelas. Detail konfigurasi disajikan pada Tabel 3.

Tabel 3. Konfigurasi Hiperparameter LSTM

Parameter	Nilai
<i>Algoritma</i>	<i>Long Short-Term Memory</i> (LSTM)
<i>Embedding Dimension</i>	128
<i>LSTM Units</i>	128
<i>Dropout Rate</i>	0.5
<i>Optimizer</i>	Adam
<i>Loss Function</i>	<i>Sparse Categorical Crossentropy</i>
<i>Epochs</i>	20 (menggunakan <i>Early Stopping</i>)

Alur komputasi pada unit LSTM diatur oleh mekanisme gerbang (*gate*) yang didefinisikan dalam persamaan matematika berikut[8]:

- **Forget Gate** (f_t): Memutuskan informasi mana yang akan dibuang.

$$f_t = \sigma(W_f \cdot [h_{t-1} x_t] + b_f) \quad (1)$$

- **Input Gate** (i_t): Memutuskan informasi baru yang akan disimpan.

$$i_t = \sigma(W_i \cdot [h_{t-1} x_t] + b_i) \quad (2)$$

- **Output Gate** (o_t): Menentukan bagian sel yang akan menjadi *output*.

$$O_t = \sigma(W_o \cdot [h_{t-1} x_t] + b_o) \quad (3)$$

2.5 Metrik Evaluasi

Kinerja model diukur menggunakan metrik klasifikasi standar berbasis *Confusion Matrix* untuk melihat tingkat ketepatan model dalam mengenali sentimen positif, negatif, dan netral [12]. Metrik yang digunakan meliputi:

- **Accuracy**: Persentase total prediksi yang benar.

- **Precision**: Ketepatan antara data yang diprediksi positif dengan data aktual.
- **Recall**: Kemampuan model menemukan kembali semua data aktual pada kelas tertentu.
- **F1-Score**: Rata-rata harmonik antara *Precision* dan *Recall*.

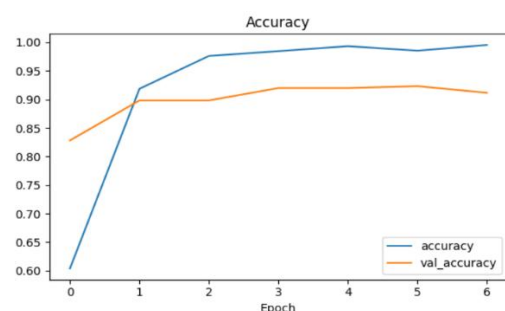
Evaluasi dilakukan pada 600 data uji untuk memastikan model memiliki kemampuan generalisasi yang baik sebelum diimplementasikan ke dalam aplikasi berbasis Streamlit.

3. Hasil dan Pembahasan

Tahap ini menyajikan temuan ilmiah dari pengujian model *Long Short-Term Memory* (LSTM) pada 3000 artikel berita ekonomi terkait indeks IDX80. Pembahasan mencakup hasil pelatihan, evaluasi metrik sebaran data, serta implementasi sistem ke dalam aplikasi prediksi sentimen.

3.1 Hasil Pelatihan Model

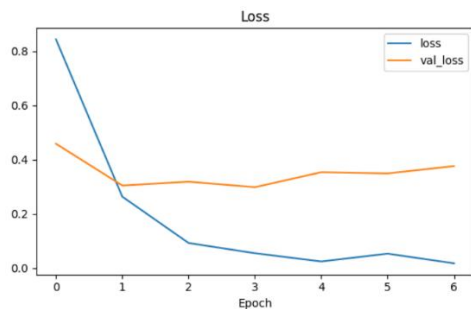
Eksperimen dilakukan menggunakan dataset yang dibagi dengan rasio 80% data latih (2400 artikel) dan 20% data uji (600 artikel). Model dilatih menggunakan *optimizer* Adam dengan *learning rate* 0.001. Meskipun skenario awal direncanakan untuk 20 *epoch*, proses pelatihan dihentikan secara otomatis pada *epoch* ke-7 melalui mekanisme *EarlyStopping* guna mencegah risiko *overfitting* dan memastikan efisiensi komputasi. Hasil pelatihan menunjukkan bahwa model mencapai nilai akurasi validasi sebesar 92,67%.



Gambar 2. Grafik Akurasi pada tahap pelatihan dan validasi

Berdasarkan Gambar 2, akurasi pelatihan mengalami peningkatan yang signifikan pada beberapa *epoch* pertama, kemudian cenderung stabil hingga proses pelatihan dihentikan. Sementara itu, akurasi validasi juga menunjukkan tren yang stabil di atas 90%, yang menandakan bahwa model mampu memberikan performa yang konsisten pada data yang belum pernah dilihat sebelumnya. Hasil ini menunjukkan bahwa

konfigurasi parameter yang digunakan telah menghasilkan model dengan tingkat akurasi yang baik untuk proses analisis sentimen.



Gambar 3. Grafik Loss pada tahap pelatihan dan validasi

Berdasarkan Gambar 3, nilai *training loss* mengalami penurunan yang signifikan pada setiap *epoch*, menunjukkan bahwa model semakin baik dalam mempelajari pola data pelatihan. Sementara itu, *validation loss* sempat menurun pada awal pelatihan, kemudian cenderung stabil dengan sedikit peningkatan pada *epoch* terakhir. Hal ini menunjukkan bahwa model telah mencapai konvergensi dan mekanisme *Early Stopping* berhasil mencegah terjadinya *overfitting* yang berlebihan.

3.2 Evaluasi Metrik Klasifikasi

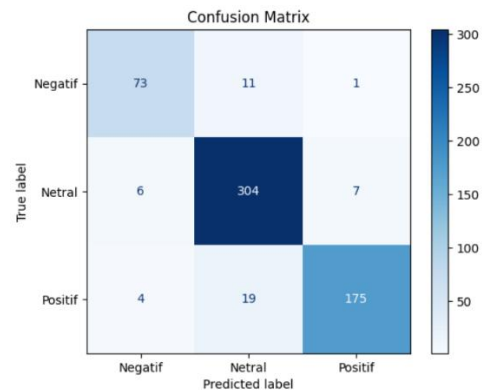
Kinerja model dievaluasi secara mendalam menggunakan *classification report* dan *confusion matrix* untuk melihat sebaran ketepatan prediksi pada tiap kelas sentimen (Negatif, Netral, dan Positif). Detail performa model disajikan pada Tabel 4.

Tabel 4. Hasil Classification Report

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Negatif	87,36%	89,41%	88,37%	85
Netral	93,73%	94,32%	94,03%	317
Positif	95,36%	93,44%	94,39%	198
<i>accuracy</i>	92,67%	92,67%	92,67%	600

Berdasarkan hasil evaluasi, model LSTM memperoleh nilai akurasi sebesar 92,67%, yang menunjukkan kemampuan klasifikasi sentimen yang sangat baik. Kelas Positif memiliki nilai *precision* tertinggi sebesar 95,36%, sedangkan kelas Netral memperoleh nilai *recall* tertinggi sebesar 94,32%. Selain itu, seluruh kelas memiliki nilai *F1-score* di atas 88%, yang mengindikasikan bahwa model mampu mengklasifikasikan sentimen negatif, netral, dan positif secara konsisten dan seimbang. Sebaran kesalahan prediksi

divisualisasikan melalui *Confusion Matrix* pada Gambar 4.



Gambar 4. Confusion Matrix Prediksi Sentimen Saham IDX80

Berdasarkan Gambar 4, model LSTM mampu mengklasifikasikan sebagian besar data dengan benar, yang terlihat dari dominasi nilai pada diagonal utama *confusion matrix*. Kesalahan klasifikasi hanya terjadi pada sebagian kecil data, terutama antara kelas negatif, netral, dan positif. Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam membedakan ketiga kelas sentimen, yang didukung oleh akurasi pengujian sebesar 92,67%.

4. Kesimpulan

Berdasarkan hasil penelitian dan eksperimen yang telah dilakukan, dapat disimpulkan bahwa sistem prediksi sentimen saham IDX80 berbasis berita ekonomi berhasil dirancang dan diimplementasikan menggunakan metode *Long Short-Term Memory* (LSTM). Model LSTM terbukti efektif dalam menangani data tekstual sekuensial berbahasa Indonesia dengan mencapai akurasi validasi tertinggi sebesar 92,67%. Penggunaan mekanisme *EarlyStopping* pada *epoch* ke-7 menjamin efisiensi proses pelatihan sekaligus meminimalkan risiko *overfitting*.

Penelitian ini memberikan kontribusi ilmiah dalam memperkaya literatur mengenai efektivitas algoritma *Deep Learning* untuk klasifikasi teks keuangan formal dalam bahasa Indonesia yang memiliki kompleksitas morfologi unik. Tahapan pra-pemrosesan teks menggunakan pustaka Sastrawi terbukti krusial dalam meningkatkan kualitas representasi kata dasar sebelum masuk ke tahap *vector embedding*.

Secara implikasi praktis, aplikasi berbasis Streamlit yang dikembangkan berfungsi sebagai "kompas" digital bagi investor ritel untuk melakukan analisis sentimen secara otomatis dan objektif terhadap volume berita ekonomi yang

masif. Fitur prediksi *batch* melalui file CSV memungkinkan respons strategis yang lebih cepat dibandingkan metode evaluasi kinerja tradisional yang bersifat *backward-looking*.

Ucapan Terima kasih

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Bapak Syariful Alam, M.Kom. selaku pembimbing utama dan Ibu Chandra Dewi Lestari, S.Pd, M.T. selaku pembimbing pendamping atas arahan dan bimbingannya. Apresiasi juga ditujukan kepada STT Wastukencana Purwakarta yang telah memfasilitasi sarana akademik selama penyusunan penelitian ini.

Daftar Pustaka

- [1] E. Fitria *et al.*, “Indonesian Journal of Economics, Management, and Accounting Analisis Perbandingan Pembentukan Portofolio Optimal Menggunakan Fama-French Three Factor Model Dan Five Factor Model (Studi Pada Saham Indeks IDX80 Di Bursa Efek Indonesia),” vol. 1, no. 8, pp. 894–903, 2024.
- [2] L. Dionosius, M. Hendrily, and R. Kuswanto, “ANALYSIS OF STOCK PORTFOLIO PERFORMANCE WITH SHARPE, TREYNOR, AND JENSEN METHODS (STUDY ON IDX80 STOCKS) ANALISIS KINERJA PORTOFOLIO SAHAM DENGAN METODE SHARPE, PENDAHULUAN Pasar modal Indonesia telah menunjukkan kemajuan pesat dalam beberapa tahun t,” vol. 8, pp. 1662–1679, 2025.
- [3] A. Fernando and K. Hidayat, “Kinerja IDX80 lebih tinggi dari LQ45 dan IDX30, ini kata analis,” p. 1, 2019. [Online]. Available: <https://investasi.kontan.co.id/news/kinerja-idx80-lebih-tinggi-dari-lq45-dan-idx30-ini-kata-analis>
- [4] D. N. Dewi, E. Hendri, and N. N. Ayu Salmah, “Pembentukan Portofolio Optimal Saham IDX 80 Menggunakan Single Index Model Yang Terdaftar Di Bursa Efek Indonesia (BEI),” vol. 9, no. 24, pp. 257–276, 2023.
- [5] Y. Ma, R. Mao, Q. Lin, P. Wu, and E. Cambria, “Multi-source aggregated classification for stock price movement prediction,” *Information Fusion*, vol. 91, pp. 515–528, Mar. 2023, doi: 10.1016/j.inffus.2022.10.025.
- [6] S. T. P. Rice and M. O. P. Rediction, “I MPACT OF LLM S NEWS S ENTIMENT A NALYSIS,” no. 3, pp. 1–6, 2026.
- [7] A. S. Nugroho, E. Prasetyo, A. Priyanto, and D. A. Puryono, “IMPLEMENTASI ALGORITMA DEEP LEARNING UNTUK ANALISIS SENTIMEN PENGGUNA PLATFORM PENDIDIKAN IMPLEMENTATION OF DEEP LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS OF EDUCATIONAL PLATFORM USERS,” vol. 8, no. 2, 2025.
- J. Xiao, S. Bi, and T. Deng, “Comparative Analysis of LSTM, GRU, and Transformer Models for Stock Price Prediction.”
- W. Huang, “Enhancing Stock Market Prediction Through LSTM Modeling and Analysis,” European Alliance for Innovation n.o., Aug. 2023. doi: 10.4108/eai.2-6-2023.2334692.
-] B. Mahesh, “Machine Learning Algorithms - A Review,” vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.
-] D. W. Otter, J. R. Medina, and J. K. Kalita, “A Survey of the Usages of Deep Learning for Natural Language Processing,” vol. XX, no. X, pp. 1–22, 2019.
-] M. Taufik Anugrah, W. Astuti, and A. Firman Ihsan, “Deteksi Anomali Data Time Series pada Operasional Industri Pipa Gas Alam menggunakan Metode Bidirectional Long Short-Term Memory (Bi-LSTM),” *Jurnal Penelitian Informatika*, vol. 3, pp. 90–97, 2025, doi: 10.25124/logic.v3i2.6467.