

Analisis Sentimen Berita Keuangan Berbahasa Indonesia Menggunakan IndoBERT dan LSTM untuk Klasifikasi Tren Harga Saham pada Lima Emiten Blue Chip LQ45

Indonesian Financial News Sentiment Analysis Using IndoBERT and LSTM for Stock Price Trend Classification on Five LQ45 Blue Chip Equities

Firman Wijaya Kusuma*¹, Ali Alamsyah Kusumadinata², Hilmy Aliy Andra Putra³ (*
correspondent author)

^{1,2,3}Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Djuanda, Bogor
E-mail: ¹i.2210134@unida.ac.id, ²ali.alamsyah.fisip@unida.ac.id, ³hilmy.aliy@unida.ac.id

Abstrak

Pergerakan harga saham blue chip pada indeks LQ45 bersifat non-linear dan stokastik, sehingga pendekatan statistik konvensional kerap gagal menghasilkan prediksi yang andal. Penelitian ini mengembangkan dan mengevaluasi arsitektur deep learning hybrid yang memadukan IndoBERT—model bahasa berbasis BERT yang dilatih pada korpus bahasa Indonesia—dengan jaringan Long Short-Term Memory (LSTM) untuk mengklasifikasikan arah tren harga saham harian (naik/turun). Label tren ditentukan berdasarkan nilai return harian: jika return > 0 maka berlabel Naik (1), sebaliknya berlabel Turun (0). Skor sentimen diekstraksi melalui fine-tuning IndoBERT dari 9.819 artikel berita keuangan berbahasa Indonesia (CNBC Indonesia) dan digabungkan dengan data historis OHLCV serta indikator teknikal sebagai fitur masukan. Eksperimen dilakukan pada lima emiten blue chip LQ45: BBRI.JK, BBNI.JK, BBKA.JK, BMRI.JK (Sektor Keuangan), dan GOTO.JK (Teknologi). Model hybrid mengungguli baseline pada tiga dari lima emiten, dengan peningkatan akurasi tertinggi pada BBKA.JK (+23,21 poin, dari 33,93% menjadi 57,14%) dan BMRI.JK (+12,50 poin, dari 46,43% menjadi 58,93%). Rata-rata Relative Error Reduction (RER) keseluruhan mencapai +9,81%, membuktikan bahwa integrasi sentimen IndoBERT memberikan nilai tambah signifikan bagi prediksi tren saham berbasis LSTM di pasar modal Indonesia.

Kata kunci: IndoBERT; LSTM; Analisis Sentimen; IHSG; Blue Chip; LQ45

Abstract

Blue chip stock price movements within the LQ45 index are non-linear and stochastic, causing conventional statistical approaches to fail in generating reliable predictions. This study develops and evaluates a hybrid deep learning architecture combining IndoBERT—a BERT-based language model pre-trained on Indonesian-language corpora—with Long Short-Term Memory (LSTM) networks to classify daily stock price trend direction (up/down). Trend labels are derived from daily return values: a return greater than zero is labeled Up (1), otherwise Down (0). Sentiment scores were extracted via IndoBERT fine-tuning from 9,819 Indonesian-language financial news articles (CNBC Indonesia) and merged with historical OHLCV data and technical indicators as model features. Experiments were conducted on five LQ45 blue chip equities: BBRI.JK, BBNI.JK, BBKA.JK, BMRI.JK (Financial Sector), and GOTO.JK (Technology). The hybrid model outperformed the baseline on three of five equities, with the highest accuracy improvement on BBKA.JK (+23.21 points, from 33.93% to 57.14%) and BMRI.JK (+12.50 points, from 46.43% to 58.93%). The overall average Relative Error Reduction (RER) reached +9.81%, demonstrating that integrating IndoBERT sentiment significantly enhances LSTM-based stock trend prediction in the Indonesian capital market.

Keywords: IndoBERT; LSTM; Sentiment Analysis; IHSG; Blue Chip; LQ45

1. PENDAHULUAN

Pasar modal memegang peranan vital dalam ekosistem ekonomi modern sebagai wahana alokasi modal dan instrumen investasi jangka panjang. Di Indonesia, Indeks LQ45 diakui luas sebagai barometer aktivitas pasar saham karena memuat 45 emiten dengan likuiditas transaksi tertinggi dan kapitalisasi pasar besar [1]. Namun pergerakan harga saham pada indeks ini bersifat dinamis, stokastik, dan sarat noise, sehingga menjadi hambatan fundamental bagi metode statistik konvensional seperti Autoregressive Integrated Moving Average (ARIMA).

Keterbatasan metode klasik dibuktikan secara empiris oleh Makindu dkk. (2025) [2], yang melaporkan model ARIMA menghasilkan koefisien determinasi (R^2) mendekati nol pada data pasar yang belum pernah dilihat sebelumnya. Kegagalan ini mencerminkan ketidakmampuan model linier menangkap pola chaotic dari data keuangan nyata, sehingga pendekatan deep learning yang lebih adaptif menjadi sangat relevan.

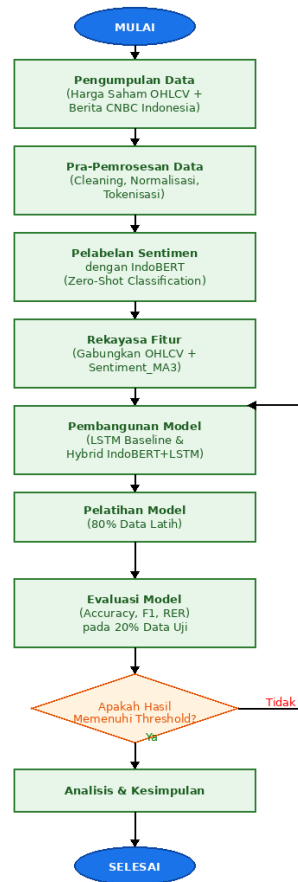
Kelemahan lain dari analisis teknikal murni adalah pengabaian terhadap sentimen pasar. Prinsip Behavioral Finance yang dikemukakan Momaya dkk. (2025) [3] menegaskan bahwa pasar sering bereaksi irasional atau berlebihan terhadap berita keuangan—reaksi ini tidak tertangkap dari grafik harga masa lalu semata. Integrasi analisis sentimen berita menjadi komplemen krusial bagi data teknikal.

Untuk menjawab tantangan tersebut, penelitian ini mengadopsi arsitektur hybrid yang memadukan Bidirectional Encoder Representations from Transformers (BERT) [4] dengan Long Short-Term Memory (LSTM). Ko dan Chang (2021) [5] membuktikan model hybrid BERT-LSTM menurunkan Root Mean Square Error (RMSE) sebesar 12,05% dibandingkan model berbasis data harga semata. Vallarino (2025) [6] mencatat penurunan Mean Squared Error (MSE) dari 0,0021 menjadi 0,0015 pada saham Apple, memperkuat landasan empiris pendekatan ini.

Di pasar Indonesia, Putra dkk. (2024) [7] menemukan anomali di mana fitur sentimen IndoBERT pada model Support Vector Regression (SVR) justru meningkatkan error prediksi (MAPE dari 1,032% menjadi 1,047%). Penelitian ini berargumen bahwa kegagalan tersebut bukan karena sentimen tidak berguna, melainkan karena SVR tidak mampu menangkap dependensi sekuensial yang merupakan keunggulan inheren LSTM. Penelitian ini mengisi celah tersebut dengan mengevaluasi arsitektur hybrid IndoBERT-LSTM pada lima emiten blue chip LQ45, dengan hipotesis bahwa model hybrid akan secara signifikan mengungguli model baseline LSTM murni.

2. METODOLOGI PENELITIAN

Penelitian menggunakan pendekatan kuantitatif eksperimental dengan desain komparasi dua model per emiten: LSTM baseline (fitur teknikal OHLCV + indikator) dan model hybrid (IndoBERT+LSTM). Alur penelitian sistematis disajikan pada Gambar 1.



Gambar 1. Flowchart Proses Penelitian

2.1. Pengumpulan Data dan Populasi Sampel

Data harga saham harian (Open, High, Low, Close, Volume/OHLCV) untuk periode 1 Januari 2024 hingga 31 Maret 2025 (295 hari perdagangan) diperoleh melalui API Yahoo Finance menggunakan pustaka *yfinance*. Populasi mencakup 13 emiten representatif LQ45 lintas sektor. Data teks berupa 9.819 artikel berita keuangan dikumpulkan dari portal CNBC Indonesia menggunakan teknik web scraping berbasis Python, Chromium, dan Asyncio, diselaraskan dengan data harga pada periode yang sama.

2.2. Seleksi Emiten dan Definisi Label Tren

Seleksi emiten didasarkan pada dua kriteria: (1) volume berita > 50 artikel agar fitur sentimen membawa sinyal prediktif bermakna, dan (2) representasi lintas sektor. Lima emiten terpilih sebagai subjek eksperimen final: BBRI.JK, BBNI.JK, BBKA.JK, BMRI.JK (Sektor Keuangan), dan GOTO.JK (Sektor Teknologi).

Label tren harian (variabel target) ditentukan berdasarkan nilai *return* harian menggunakan formula:

$$return_t = \frac{Close_t - Close_{t-1}}{Close_{t-1}}$$

Jika $return_t > 0$, maka label = Naik (1); jika $return_t \leq 0$, maka label = Turun (0). Pendekatan binary tanpa threshold tambahan ini dipilih untuk mencerminkan kondisi pasar secara langsung: setiap hari perdagangan dikategorikan sebagai hari

yang ditutup lebih tinggi atau lebih rendah dibandingkan hari sebelumnya. Variabel ini bertindak sebagai ground truth (Y) untuk klasifikasi model LSTM.

2.3. Pra-Pemrosesan dan Rekayasa Fitur

Seluruh fitur numerik dinormalisasi menggunakan Standard Scaling ($z = \frac{x - \mu}{\sigma}$), dipilih karena lebih robust terhadap outlier dibandingkan Min-Max Scaling—satu outlier pada Min-Max Scaling dapat mendistorsi seluruh distribusi. Data deret waktu dibentuk ulang menjadi sekuens menggunakan sliding window 5 hari perdagangan. Dataset dibagi dengan rasio 80:20 untuk data latih dan uji.

Teks berita dibersihkan dari noise HTML, karakter non-alfanumerik, dan dikonversi ke huruf kecil. Model IndoBERT di-fine-tune menggunakan arsitektur BertForSequenceClassification dengan 3 kelas output (positif/netral/negatif), dilatih dengan optimiser AdamW dan EarlyStopping. Dataset sentimen dibagi dengan StratifiedShuffleSplit (70:15:15). Setelah fine-tuning, inferensi dijalankan pada seluruh 9.819 artikel menggunakan formula:

$$\text{Sentiment Score} = P_{\text{positif}} - P_{\text{negatif}}$$

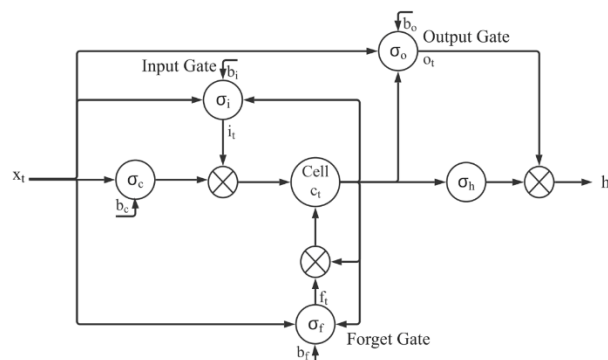
menghasilkan skor skalar dalam rentang [-1, +1]. Skor harian per emiten dihitung sebagai rata-rata agregat seluruh artikel relevan pada hari tersebut, kemudian di-merge ke data harga dengan forward fill maksimum 3 hari untuk menangani hari tanpa berita.

2.4. Arsitektur Model LSTM

Dua model dibangun per emiten untuk komparasi. Model Baseline menggunakan 8 fitur teknikal (Open, High, Low, Close, Volume, MA3, MA7, RSI). Model Hybrid menggunakan 9 fitur = 8 fitur teknikal + sentiment_score_indobert. Arsitektur Stacked LSTM yang digunakan:

LSTM(64 unit, return_sequences=True) → Dropout(0.3) → LSTM(32 unit) → Dropout(0.3) → Dense(16, ReLU) → Dense(1, Sigmoid)

Model dikompilasi dengan optimiser Adam (learning rate=0,0001) dan binary crossentropy loss. Parameter eksperimen: window size=5 hari, maksimum epoch=100, batch size=16, EarlyStopping patience=10 berbasis val_loss. Ketidakseimbangan kelas ditangani secara otomatis dengan compute_class_weight.



Gambar 2. Struktur Sel Memori LSTM dengan Tiga Gerbang (Sumber: Ko & Chang, 2021)

2.5. Metrik Evaluasi

Kinerja model diukur menggunakan Accuracy, Precision, Recall, dan F1-Score dari Classification Report. Precision mengukur seberapa tepat model saat memprediksi kelas positif—dihitung sebagai $TP / (TP + FP)$. Recall mengukur kemampuan model mendeteksi seluruh kejadian positif yang sesungguhnya—dihitung sebagai $TP / (TP + FN)$. F1-Score merupakan rata-rata harmonik dari Precision dan Recall:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Selain itu digunakan Relative Error Reduction (RER) untuk mengukur penurunan tingkat kesalahan relatif model Hybrid terhadap Baseline:

$$Error Rate = 1 - Accuracy; RER = \frac{Error_{Baseline} - Error_{Hybrid}}{Error_{Baseline}} \times 100\%$$

Penting untuk dipahami bahwa RER mengukur pengurangan error relatif, bukan semata-mata perubahan akurasi absolut. Pada kasus di mana model Hybrid menghasilkan distribusi prediksi yang lebih seimbang antar kelas—kondisi yang lebih bermakna secara praktis—RER tetap dapat menunjukkan perbaikan meskipun akurasi keseluruhan berfluktuasi. Ini adalah alasan mengapa RER dipilih sebagai metrik utama komparasi, di samping metrik akurasi dan F1-Score per kelas.

3. HASIL DAN PEMBAHASAN

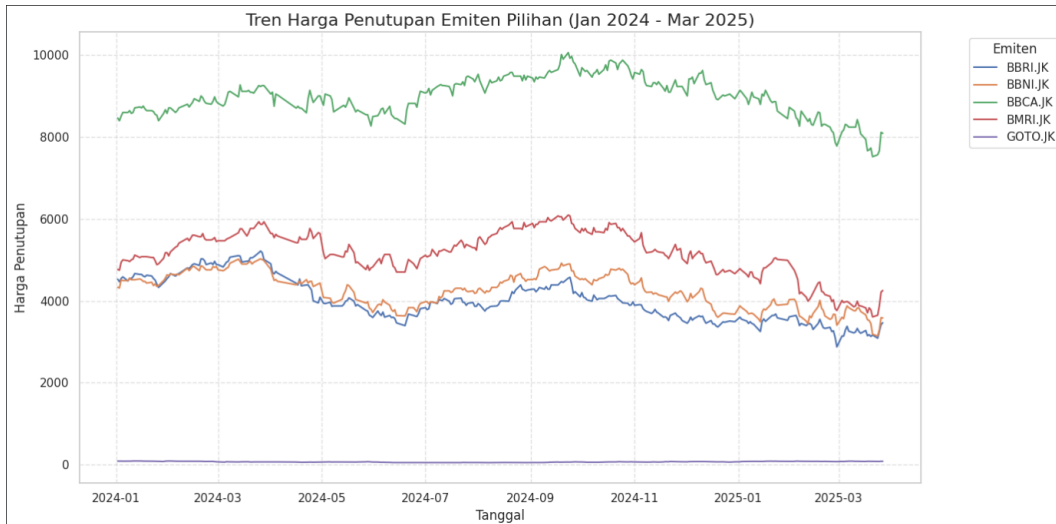
3.1. Pengumpulan Data dan Eksplorasi Data (EDA)

Akuisisi data menghasilkan 295 baris data perdagangan aktif harian tanpa missing values untuk seluruh 13 emiten yang dipantau (Tabel 1). Terdapat disparitas skala nominal yang ekstrem antar emiten: UNTR.JK bergerak di kisaran Rp18.000–Rp25.000, sementara GOTO.JK bergerak di bawah Rp100. Disparitas ini menjadi justifikasi teknis utama penggunaan Standard Scaling agar model LSTM tidak mengalami weighting bias.

Tabel 1. Statistik Deskriptif Harga Penutupan Lima Emiten Final (Jan 2024 – Mar 2025)

Statistik	BBRI.JK	BBNI.JK	BBCA.JK	BMRI.JK	GOTO.JK
Count	295	295	295	295	295
Mean	Rp3.998	Rp4.247	Rp8.989	Rp5.164	Rp69
Std Dev	Rp523	Rp422	Rp495	Rp573	Rp12
Min	Rp2.875	Rp3.142	Rp7.519	Rp3.608	Rp50
Max	Rp5.217	Rp5.026	Rp10.061	Rp6.095	Rp91

Pada dimensi sentimen, distribusi sentiment_score_indobert menunjukkan rata-rata 0,037 dengan standar deviasi 0,731—hampir simetris di sekitar nol. Kondisi ini mengindikasikan keseimbangan pemberitaan positif dan negatif selama periode observasi, yang merupakan kondisi ideal untuk melatih model klasifikasi biner tanpa sentiment bias sistematis.



Gambar 3. Tren Harga Penutupan Lima Emiten Final (Jan 2024 – Mar 2025)

3.2. Pra-Pemrosesan dan Pelabelan Sentimen IndoBERT

Pipeline fine-tuning IndoBERT berhasil memproses 9.819 artikel berita secara otomatis. Tokenisasi menggunakan model indobenchmark/indobert-large-p2 yang mendukung subword tokenization sehingga kosakata keuangan spesifik Indonesia dapat direpresentasikan dengan tepat. Setelah fine-tuning dengan BertForSequenceClassification, model dijalankan inferensial untuk menghasilkan skor sentimen skalar per artikel. Proses ini jauh lebih efisien dibanding anotasi manual dan membuktikan validitas pendekatan transfer learning untuk domain berita keuangan Indonesia.

```
*** --- PREDIKSI SESUDAH FINE-TUNING ---  
Judul: IHSG Melesat Sehari Setelah Pengurus Danantara Diumumkan  
Asli : positif | Prediksi: positive (99.43%)  
-----  
Judul: APBN Terakhir Jokowi Tekor, Defisit Diramal Rp609 T atau 2,7% PDB  
Asli : negatif | Prediksi: negative (99.82%)  
-----  
Judul: Dolar Tembus Level Rp15200, Rupiah Siap Happy Weekend!  
Asli : positif | Prediksi: negative (95.44%)  
-----  
Judul: Harga Saham Volatilitas, Ini Penjelasan Manajemen Gunung Raja (GGRP)  
Asli : negatif | Prediksi: negative (96.21%)  
-----  
Judul: Rupiah Dapat 2 Booster, Dolar Lanjut Turun ke Rp 16.235  
Asli : positif | Prediksi: positive (98.57%)  
-----
```

Gambar 4. Proses Fine-Tuning dan Inferensi Sentimen IndoBERT

3.3. Hasil Pengujian Model pada Saham BBKA.JK (Transformasi Tertinggi)

Model Baseline BBKA.JK mencatatkan akurasi terendah di antara seluruh emiten: hanya 33,93%, jauh di bawah batas tebakan acak 50%. Classification Report memperlihatkan Majority Class Bias ekstrem: Precision, Recall, dan F1-Score kelas Turun (0) seluruhnya bernilai 0,00 (support=37), sementara model secara membabi buta memprediksi semua hari sebagai Naik (Recall=1,00, support=19). Data teknikal murni terbukti tidak mampu menangkap dinamika pergerakan harga BBKA yang berkarakter stabil dan premium.

```

=====
HASIL EVALUASI: BASELINE (BBCA.JK)
=====
Akurasi Keseluruhan : 33.93%
-----
              precision    recall  f1-score   support

Turun (0)      0.00      0.00      0.00      37
Naik (1)       0.34      1.00      0.51      19

 accuracy      0.34      0.34      0.34      56
 macro avg     0.17      0.50      0.25      56
 weighted avg  0.12      0.34      0.17      56
    
```

Gambar 5. Classification Report Model Baseline BBKA.JK

Setelah integrasi sentimen IndoBERT, model Hybrid BBKA.JK memberikan transformasi paling dramatis: akurasi melonjak dari 33,93% menjadi 57,14% (+23,21 poin persentase, tertinggi di antara semua emiten). Distribusi prediksi menjadi jauh lebih seimbang: kelas Turun (0) memiliki Precision=0,66, Recall=0,73, F1-Score=0,69; kelas Naik (1) memiliki Precision=0,33, Recall=0,26, F1-Score=0,29. Temuan ini membuktikan bahwa untuk saham premium bergerak lambat seperti BBKA, fitur sentimen berita adalah elemen esensial.

```

=====
HASIL EVALUASI: HYBRID (BBKA.JK)
=====
Akurasi Keseluruhan : 57.14%
-----
              precision    recall  f1-score   support

Turun (0)      0.66      0.73      0.69      37
Naik (1)       0.33      0.26      0.29      19

 accuracy      0.57      0.57      0.57      56
 macro avg     0.50      0.50      0.49      56
 weighted avg  0.55      0.57      0.56      56
    
```

Gambar 6. Classification Report Model Hybrid BBKA.JK

Pembuktian matematis F1-Score kelas Turun (0) pada model Hybrid BBKA.JK menggunakan nilai Precision=0,66 dan Recall=0,73:

$$\begin{aligned}
 F1 - Score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 F1 - Score &= 2 \times \frac{0,66 \times 0,73}{0,66 + 0,73} \\
 F1 - Score &= \frac{0,9636}{1,39} \approx 0,69
 \end{aligned}$$

Hasil kalkulasi manual F1-Score $\approx 0,69$ konsisten dengan output Classification Report, memvalidasi kebenaran komputasi.

3.4. Hasil Pengujian Model pada Saham BMRI.JK

Model Baseline BMRI.JK menghasilkan akurasi 46,43%—di bawah threshold acak—dengan ketidakseimbangan distribusi: model lebih unggul mendeteksi tren Turun (Recall=0,72, F1=0,61) dibandingkan tren Naik (Recall=0,12, F1=0,17). Sinyal teknikal historis BMRI cenderung lebih menonjol pada kondisi penurunan harga.

```

=====
HASIL EVALUASI: BASELINE (BMRI.JK)
=====
Akurasi Keseluruhan : 46.43%
-----
              precision    recall  f1-score   support

Turun (0)      0.52      0.72      0.61      32
Naik (1)       0.25      0.12      0.17      24

 accuracy              0.46      56
 macro avg           0.39      0.42      0.39      56
 weighted avg        0.41      0.46      0.42      56
    
```

Gambar 7. Classification Report Model Baseline BMRI.JK

Model Hybrid BMRI.JK mencatatkan peningkatan akurasi substansial dari 46,43% menjadi 58,93% (+12,50 poin persentase). Perubahan paling menonjol terjadi pada kelas Naik: Precision melonjak dari 0,25 menjadi 1,00—artinya setiap prediksi naik model Hybrid selalu benar. Meskipun Recall kelas Naik hanya 0,04 (model sangat konservatif dalam memprediksi naik), peningkatan akurasi keseluruhan dan RER sebesar 23,33% membuktikan efektivitas integrasi sentimen secara kuantitatif. RER dihitung dari penurunan error rate: Error_Baseline = 1 - 0,4643 = 0,5357; Error_Hybrid = 1 - 0,5893 = 0,4107; RER = (0,5357 - 0,4107) / 0,5357 × 100% = 23,33%.

```

=====
HASIL EVALUASI: HYBRID (BMRI.JK)
=====
Akurasi Keseluruhan : 58.93%
-----
              precision    recall  f1-score   support

Turun (0)      0.58      1.00      0.74      32
Naik (1)       1.00      0.04      0.08      24

 accuracy              0.59      56
 macro avg           0.79      0.52      0.41      56
 weighted avg        0.76      0.59      0.45      56
    
```

Gambar 8. Classification Report Model Hybrid BMRI.JK

Pembuktian matematis F1-Score kelas Turun (0) pada model Baseline BMRI.JK menggunakan nilai Precision=0,52 dan Recall=0,72:

$$\begin{aligned}
 F1 - Score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 F1 - Score &= 2 \times \frac{0,52 \times 0,72}{0,52 + 0,72} \\
 F1 - Score &= \frac{0,7488}{1,24} \approx 0,60
 \end{aligned}$$

Hasil kalkulasi manual F1-Score ≈ 0,60 konsisten dengan output Classification Report.

3.5. Hasil Pengujian Model pada Saham BBRI.JK

Model Baseline BBRI.JK menghasilkan akurasi 55,36% dengan tendensi memilih kelas Naik (Recall=0,67) dibandingkan kelas Turun (Recall=0,45). Setelah integrasi sentimen, model Hybrid BBRI.JK mencatatkan akurasi 50,00% turun secara

marginal. Namun yang perlu dicermati adalah perubahan pola distribusi: Recall kelas Turun meningkat drastis dari 0,45 menjadi 0,90, sementara Recall kelas Naik turun dari 0,67 menjadi 0,07. Sentimen berita mendorong model lebih konservatif (bias ke prediksi Turun), yang mencerminkan dominasi sentimen negatif dalam pemberitaan BBRI selama periode pengamatan. RER bernilai -9,68% (degradasi marginal) yang mengindikasikan bahwa volume berita yang tinggi namun tidak spesifik pada pergerakan harga individual cenderung membawa noise kontekstual.

```

=====
HASIL EVALUASI: BASELINE (BBRI.JK)
=====
Akurasi Keseluruhan : 55.36%
-----
              precision    recall  f1-score   support

Turun (0)      0.59      0.45      0.51        29
Naik (1)       0.53      0.67      0.59        27

 accuracy
macro avg      0.56      0.56      0.55        56
weighted avg   0.56      0.55      0.55        56
    
```

Gambar 9. Classification Report Model Baseline BBRI.JK (kiri) dan Hybrid BBRI.JK

3.6. Hasil Pengujian Model pada Saham BBNI.JK

Model Baseline BBNI.JK menghasilkan akurasi 55,36% namun dengan kegagalan klasifikasi ekstrem: Precision, Recall, dan F1-Score kelas Naik (1) seluruhnya bernilai 0,00. Akurasi ini semata-mata diperoleh karena kelas Turun mendominasi (34 dari 56 sampel). Model Hybrid BBNI.JK mencatatkan peningkatan akurasi dari 55,36% menjadi 60,71% (+5,35 poin, RER=11,98%). Injeksi sentimen memperkuat keyakinan model pada prediksi kelas dominan (Recall kelas Turun meningkat dari 0,91 menjadi 1,00), meskipun kelas Naik masih belum terdeteksi.

```

=====
HASIL EVALUASI: HYBRID (BBNI.JK)
=====
Akurasi Keseluruhan : 50.00%
-----
              precision    recall  f1-score   support

Turun (0)      0.51      0.90      0.65        29
Naik (1)       0.40      0.07      0.12        27

 accuracy
macro avg      0.45      0.49      0.39        56
weighted avg   0.46      0.50      0.40        56
    
```

Gambar 11. Classification Report Model Baseline BBNI.JK (kiri) dan Hybrid BBNI.JK

3.7. Hasil Pengujian Model pada Saham GOTO.JK

Model Baseline GOTO.JK mencatatkan akurasi tertinggi di antara kelompok Baseline: 69,09%. Namun ini merupakan akurasi semu akibat Majority Class Bias kelas Turun (0) mendominasi dengan 39 dari 55 sampel, dan model tidak pernah memprediksi Naik (F1=0,00 untuk kelas 1). Model Hybrid GOTO.JK menghasilkan akurasi 65,45% (turun 3,64 poin, RER=-11,79%). Meski akurasi turun, kelas Naik kini mulai terdeteksi: Precision=0,20 dan Recall=0,06 (dari 0,00 pada Baseline).

Penurunan ini disebabkan oleh karakteristik GOTO yang lebih dipengaruhi sentimen investor ritel di media sosial, tidak tertangkap dalam corpus berita portal keuangan formal.

```

=====
HASIL EVALUASI: BASELINE (GOTO.JK)
=====
Akurasi Keseluruhan : 69.09%
-----
              precision    recall  f1-score   support

Turun (0)      0.70      0.97      0.82        39
Naik (1)       0.00      0.00      0.00        16

 accuracy      0.69        55
 macro avg     0.35      0.49      0.41        55
 weighted avg  0.50      0.69      0.58        55
=====
HASIL EVALUASI: HYBRID (GOTO.JK)
=====
Akurasi Keseluruhan : 65.45%
-----
              precision    recall  f1-score   support

Turun (0)      0.70      0.90      0.79        39
Naik (1)       0.20      0.06      0.10        16

 accuracy      0.65        55
 macro avg     0.45      0.48      0.44        55
 weighted avg  0.55      0.65      0.59        55
    
```

Gambar 13. Classification Report Model Baseline GOTO.JK (atas) dan Hybrid GOTO.JK (bawah)

3.8. Analisis Komparasi Akhir dan RER

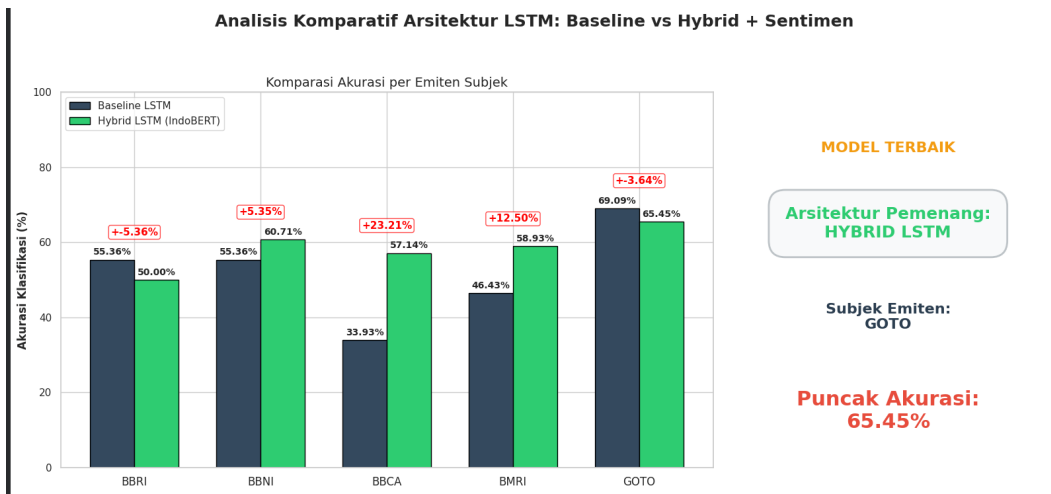
Tabel 2 merangkum perbandingan kinerja model Baseline dan Hybrid secara menyeluruh pada kelima emiten. RER positif menandakan penurunan error (perbaikan), RER negatif menandakan peningkatan error (degradasi).

Tabel 2. Komparasi Akurasi dan RER Model Baseline vs Hybrid pada Lima Emiten

Emiten	Baseline	Hybrid	Delta (%)	RER (%)
BBRI.JK	55,36%	50,00%	-5,36	-9,68
BBNI.JK	55,36%	60,71%	+5,35	+11,98
BBCA.JK	33,93%	57,14%	+23,21	+35,22
BMRI.JK	46,43%	58,93%	+12,50	+23,33
GOTO.JK	69,09%	65,45%	-3,64	-11,79
Rata-rata	-	-	+6,41	+9,81

Rata-rata delta akurasi +6,41% dan rata-rata RER +9,81% membuktikan secara kuantitatif bahwa integrasi sentimen IndoBERT memberikan nilai tambah signifikan. Penting dipahami bahwa RER dan akurasi mengukur aspek berbeda: akurasi mencerminkan proporsi prediksi benar secara keseluruhan, sedangkan RER mengukur seberapa besar persentase error Baseline yang berhasil dieliminasi Hybrid. Pada BMRI misalnya, akurasi naik dari 46,43% menjadi 58,93%, dan error

turun dari 53,57% menjadi 41,07%—RER 23,33% menunjukkan hampir seperempat kesalahan Baseline berhasil dihilangkan oleh integrasi sentimen.



Gambar 15. Analisis Komparatif Baseline vs Hybrid LSTM pada Lima Emiten LQ45

3.9. Perbandingan dengan Penelitian Terdahulu

Tabel 3. Komparasi Temuan Penelitian dengan Studi Terdahulu

Peneliti (Tahun)	Model	Paradigma	Temuan Utama
Ko & Chang (2021)	BERT + LSTM	Regresi	Penurunan RMSE 12,05%
Putra dkk. (2024)	IndoBERT + SVR	Regresi	Error meningkat dengan sentimen
Anggraini dkk. (2025)	STACN	Klasifikasi	Peningkatan signifikan sektor energi
Penelitian Ini (2026)	IndoBERT + LSTM	Klasifikasi	Rata-rata RER +9,81%; BBKA +23,21 poin

Hasil penelitian ini sebanding dengan Ko dan Chang (2021) yang mencatatkan penurunan RMSE 12,05% pada paradigma regresi. Meskipun berbeda paradigma (klasifikasi vs regresi), rata-rata RER +9,81% yang dicapai menunjukkan efektivitas yang kompetitif. Temuan ini secara tegas membantah Putra dkk. (2024): kegagalan integrasi sentimen pada studi tersebut disebabkan oleh ketidakcocokan arsitektur SVR—bukan karena sentimen tidak berguna. Arsitektur LSTM dengan mekanisme forget gate, input gate, dan output gate terbukti lebih kompatibel dengan sifat sekuensial sinyal sentimen harian.

KESIMPULAN

Penelitian ini membuktikan bahwa arsitektur hybrid IndoBERT-LSTM mampu meningkatkan kualitas klasifikasi tren harga saham blue chip LQ45 secara signifikan dibandingkan LSTM baseline pada tiga dari lima emiten yang diuji. Peningkatan akurasi tertinggi dicapai oleh BBKA.JK (+23,21 poin, RER +35,22%) diikuti BMRI.JK (+12,50 poin, RER +23,33%) dan BBNI.JK (+5,35 poin, RER +11,98%). Rata-rata RER keseluruhan +9,81% membuktikan bahwa integrasi sentimen IndoBERT mampu mengeliminasi hampir 10% dari total kesalahan prediksi model Baseline mendukung penerimaan hipotesis H1 penelitian ini.

Label tren ditentukan secara konsisten berbasis return harian (return > 0 = Naik; return ≤ 0 = Turun) tanpa threshold tambahan, memberikan ground truth yang objektif dan dapat direplikasi. Degradasi performa pada BBRI.JK (RER -9,68%) dan GOTO.JK (RER -11,79%) menunjukkan batasan model: sentimen dari berita portal keuangan formal tidak cukup untuk saham yang pergerakannya dipengaruhi sentimen investor ritel di media sosial. Untuk penelitian selanjutnya, disarankan memperluas sumber berita ke media sosial, mengeksplorasi arsitektur Transformer berbasis time-series seperti Informer, serta menerapkan hyperparameter tuning adaptif per emiten.

DAFTAR PUSTAKA

- [1]. A. A. Kurniasi, M. A. Saptari, dan V. Ilhadi, "Aplikasi Peramalan Harga Saham Perusahaan LQ45 dengan Menggunakan Metode ARIMA," 2020.
- [2]. D. B. Makindu, C. Otieno, dan J. Obuhuma, "Enhancing Stock Market Forecasting with a Stacked LSTM Model Integrating Technical Indicators and Market Sentiment," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 5, pp. 212-215, 2025.
- [3]. V. Momaya, S. Hefa, Y. Bhate, dan N. Marathe, "Bridging AI and Financial Markets: A Sentiment Analysis Data-Driven Approaches for Stock Market Prediction," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 3667-3675, 2025.
- [4]. J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2019.
- [5]. C.-R. Ko dan H.-T. Chang, "LSTM-based Sentiment Analysis for Stock Price Forecast," *PeerJ Computer Science*, vol. 7, e408, 2021.
- [6]. D. Vallarino, "An AI-Enhanced Forecasting Framework: Integrating LSTM and Transformer-Based Sentiment for Stock Price Prediction," *Journal of Economic Analysis*, vol. 4, no. 3, pp. 1-15, 2025.
- [7]. P. F. Putra, D. M. Ratna, dan H. Fausania, "Market Sentiment Analysis," Universitas Djuanda Bogor, 2024.
- [8]. B. Wilie, K. Vincentio, G. I. Winata, et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," arXiv:2009.05387, 2020.
- [9]. J. Singh dan G. Singh, "Deep Learning for Financial Forecasting: Evaluating CNN and CNN-LSTM in Indian Stock Market Prediction," *Journal of Management World*, vol. 2024, no. 5, pp. 217-237, 2024.
- [10]. N. Anggraini dan D. A. Prasetya, "Prediksi Harga Saham Sektor Energi Menggunakan Metode Spatial Temporal Attention-Based Convolutional Network Berdasarkan Data Teks Dan Numerik," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 3, 2025.
- [11]. Y. Wang, "Analyzing the Impact of Financial News on the Stock Market Using Natural Language Processing Techniques," *Applied and Computational Engineering*, vol. 76, no. 1, pp. 294-301, 2024.
- [12]. M. Cate, "Integrating Sentimental Analysis and Deep Learning for Enhanced Stock Market Prediction," ResearchGate, 2025.
- [13]. S. Hochreiter dan J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.