BAB 2
LANDASAN TEORI

Bab ini menjabarkan teori-teori yang mendasari penelitian mengenai prediksi pergerakan Indeks Harga Saham Gabungan (IHSG) menggunakan model *Hybrid* LSTM-IndoBERT berbasis sentimen berita ekonomi Indonesia. Teori yang dibahas dipilih berdasarkan keterkaitannya secara langsung dengan tahapan penelitian, yaitu data deret waktu IHSG, pembentukan target prediksi, model LSTM untuk data temporal, IndoBERT untuk representasi teks bahasa Indonesia, *zero-shot classification* berbasis *Natural Language Inference* (NLI), reduksi dimensi PCA, mekanisme *cross-attention*, teknik SMOTE, model ARIMA, serta metrik evaluasi yang digunakan. Pembahasan pada bab ini difokuskan pada teori yang secara langsung mendukung rancangan model, proses eksperimen, dan evaluasi penelitian.

2.1 Indeks Harga Saham Gabungan dan Representasi Deret Waktu Finansial

Indeks Harga Saham Gabungan (IHSG) merupakan indeks yang mengukur kinerja harga saham seluruh perusahaan tercatat pada papan utama dan papan pengembangan di Bursa Efek Indonesia [1]. Dalam penelitian prediksi pasar saham, nilai indeks seperti IHSG dapat diperlakukan sebagai data deret waktu karena setiap observasi harga tersusun berdasarkan urutan waktu dan memiliki keterkaitan dengan observasi sebelumnya [10]. Karakteristik deret waktu finansial cenderung tidak stasioner, mengandung *noise*, serta dipengaruhi oleh faktor eksternal seperti kondisi ekonomi, sentimen investor, dan informasi berita [2, 3].

Prediksi pada data finansial tidak selalu dilakukan secara langsung terhadap harga absolut. Harga absolut dapat memiliki skala yang besar dan dipengaruhi tren jangka panjang, sehingga perubahan relatif antartitik waktu lebih sering digunakan dalam pemodelan finansial [2, 4]. Salah satu bentuk representasi perubahan harga yang umum digunakan adalah *log return*. *Log return* digunakan karena mampu merepresentasikan perubahan relatif harga secara lebih stabil dibandingkan perubahan harga absolut [2, 3].

Secara matematis, *log return* dinyatakan sebagai berikut [2].

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2.1}$$

di mana P_t adalah harga penutupan pada waktu ke-t, sedangkan P_{t-1} adalah harga penutupan pada waktu sebelumnya. Dalam implementasi penelitian ini, target regresi dibentuk dari *log return* harga penutupan yang telah dihaluskan menggunakan *moving average* 3 hari (MA3). Penghalusan digunakan untuk mengurangi fluktuasi harian yang umum terjadi pada data pasar saham [2, 3].

$$P_t^{MA3} = \frac{P_t + P_{t-1} + P_{t-2}}{3} \tag{2.2}$$

Dengan demikian, target regresi yang digunakan dalam penelitian ini dapat ditulis sebagai berikut.

$$y_t = \ln\left(\frac{P_t^{MA3}}{P_{t-1}^{MA3}}\right) \tag{2.3}$$

Selain target regresi, penelitian ini juga mengevaluasi arah pergerakan IHSG, yaitu apakah IHSG mengalami kenaikan atau penurunan pada periode berikutnya. Evaluasi arah penting dalam konteks pasar modal karena keputusan perdagangan sering kali lebih bergantung pada ketepatan arah pergerakan dibandingkan besaran prediksi harga secara absolut [2, 3].

2.2 Indikator Teknikal, *Feature Engineering*, dan Pemilihan Fitur

Data pasar harian yang umum tersedia dari Yahoo Finance berbentuk atribut OHLCV, yaitu *open*, *high*, *low*, *close*, dan *volume*. Kelima atribut tersebut merepresentasikan informasi dasar mengenai harga pembukaan, harga tertinggi, harga terendah, harga penutupan, dan aktivitas perdagangan pada satu hari bursa. Dalam prediksi pasar saham, OHLCV sering digunakan sebagai titik awal karena dapat menggambarkan dinamika harga dan likuiditas harian. Namun, kelima atribut tersebut belum secara langsung mengekspresikan informasi turunan seperti momentum, volatilitas, dan perubahan relatif harga. Oleh karena itu, penelitian ini melakukan *feature engineering* dengan menurunkan beberapa indikator teknikal dari OHLCV.

Indikator teknikal digunakan untuk memperkaya representasi data pasar dengan menangkap karakteristik yang tidak selalu terlihat dari harga mentah. Penelitian terdahulu menunjukkan bahwa kombinasi data historis transaksi, indikator teknikal, dan sentimen dapat membantu model menangkap informasi

penting yang memengaruhi arah pergerakan saham [11]. Penelitian lain juga menggunakan indikator seperti RSI, MACD, Bollinger Bands, moving average, dan fitur volatilitas sebagai masukan model LSTM untuk prediksi pasar saham [12, 13, 14]. Dengan demikian, pemilihan fitur pada penelitian ini tidak dilakukan melalui seleksi otomatis seperti LASSO, melainkan melalui pendekatan *domain-driven feature engineering* berdasarkan indikator yang umum digunakan dalam analisis teknikal dan penelitian prediksi saham.

Lima fitur dasar OHLCV dikembangkan menjadi 12 fitur teknikal. Fitur tambahan yang dihitung adalah RSI, MACD, `macd_signal`, `macd_diff`, `bb_width`, `daily_return`, dan `volatility`. RSI digunakan untuk merepresentasikan momentum pasar, MACD dan turunannya digunakan untuk menangkap perubahan tren dan momentum, `bb_width` digunakan untuk menggambarkan volatilitas berbasis Bollinger Bands, `daily_return` digunakan untuk merepresentasikan perubahan relatif harian, sedangkan `volatility` dihitung sebagai standar deviasi bergerak dari *daily return*. Dengan demikian, proses pembentukan fitur dapat dipahami sebagai perubahan dari fitur mentah OHLCV menjadi fitur pasar yang mencakup kategori harga, volume, momentum, tren, return, dan volatilitas.

Perlu diperhatikan bahwa 12 fitur teknikal tersebut digunakan sebagai input model, sedangkan target prediksi penelitian adalah *log return* dari harga penutupan yang telah dihaluskan menggunakan MA3. Dengan kata lain, tidak semua 12 fitur diperlukan untuk menghitung *log return*. *Log return* secara matematis cukup dihitung dari harga penutupan atau harga penutupan yang telah dihaluskan. Namun, 12 fitur tetap digunakan sebagai variabel penjelas agar model memiliki informasi pasar yang lebih kaya dalam memprediksi target tersebut.

2.3 *Long Short-Term Memory* untuk Data Deret Waktu

Long Short-Term Memory (LSTM) merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk mempelajari pola sekuensial dan ketergantungan jangka panjang pada data deret waktu [6]. LSTM menggunakan mekanisme gerbang untuk mengatur informasi yang perlu disimpan, dilupakan, dan diteruskan ke langkah waktu berikutnya [6]. RNN konvensional memiliki keterbatasan dalam mempelajari dependensi jangka panjang karena rentan terhadap masalah *vanishing gradient*. LSTM mengatasi permasalahan tersebut melalui struktur *cell state* dan mekanisme gerbang yang membantu menjaga aliran informasi sepanjang sekuens [6].

Pada setiap waktu t, LSTM menerima input x_t, *hidden state* sebelumnya h_{t-1}, dan *cell state* sebelumnya C_{t-1}. Operasi utama LSTM dirumuskan sebagai berikut [6].

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2.4}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2.5}$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{2.6}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{2.7}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{2.8}$$

$$h_t = o_t \odot \tanh(C_t) \tag{2.9}$$

di mana f_t adalah *forget gate*, i_t adalah *input gate*, o_t adalah *output gate*, C_t adalah *cell state*, h_t adalah *hidden state*, σ adalah fungsi sigmoid, dan $\odot$ adalah operasi perkalian elemen-demi-elemen.

Dalam penelitian ini, LSTM digunakan untuk memproses data teknikal IHSG yang dibentuk dalam format *sliding window*. Penggunaan LSTM pada data saham relevan karena model ini mampu mempelajari pola temporal dari data historis dan indikator teknikal [7, 4]. Model yang digunakan adalah LSTM bidirectional, yaitu LSTM yang memproses sekuens dari dua arah. Representasi akhir diperoleh dari penggabungan arah maju dan arah mundur sebagai berikut [6].

$$h_t^{bi} = [\overrightarrow{h_t} \parallel \overleftarrow{h_t}] \tag{2.10}$$

Penggunaan LSTM bidirectional bertujuan untuk menghasilkan representasi temporal yang lebih kaya dari pola pergerakan harga dalam jendela input yang digunakan. Representasi ini menjadi dasar bagi model untuk menangkap pola historis IHSG sebelum digabungkan dengan fitur sentimen berita.

2.3.1 Pemilihan Panjang *Lookback Window*

Panjang *lookback window* menentukan banyaknya observasi historis yang diberikan kepada LSTM untuk membentuk satu prediksi. Window yang terlalu pendek berpotensi tidak menyediakan konteks temporal yang memadai, sedangkan

window yang terlalu panjang dapat memasukkan informasi lama yang kurang relevan, meningkatkan kompleksitas komputasi, dan mengurangi jumlah sekuens pelatihan. Penelitian mengenai pengaruh panjang sekuens pada prediksi indeks saham menunjukkan bahwa panjang window optimal dapat berbeda antarindeks dan karakteristik pasar. Oleh karena itu, panjang window perlu diperlakukan sebagai hyperparameter yang dievaluasi secara empiris, bukan sebagai nilai tetap yang berlaku universal [15].

Dalam penelitian ini, ukuran 3, 7, dan 14 hari dipilih sebagai kandidat analisis sensitivitas untuk mewakili tiga tingkat konteks historis jangka pendek. Window 3 hari mewakili reaksi pasar yang sangat baru, window 7 hari memberikan konteks pendek hingga menengah, sedangkan window 14 hari digunakan untuk menguji apakah tambahan memori historis meningkatkan kemampuan model. Ketiga nilai tersebut tidak dipilih karena dianggap optimal secara teoritis. Konfigurasi final tetap ditentukan berdasarkan pipeline eksperimen yang digunakan secara konsisten pada proses pelatihan, validasi, evaluasi, dan implementasi model.

2.4 BERT dan IndoBERT untuk Representasi Teks Bahasa Indonesia

Bidirectional Encoder Representations from Transformers (BERT) adalah model bahasa berbasis arsitektur *Transformer* yang menghasilkan representasi kata secara kontekstual [8, 16]. Berbeda dengan representasi statis, model berbasis BERT menghasilkan *embedding* yang bergantung pada konteks kalimat, sehingga makna kata dapat berubah sesuai kata-kata lain di sekitarnya [16, 8]. Model berbasis *Transformer* menggunakan mekanisme *attention* untuk mempelajari hubungan antar token secara adaptif [8].

Arsitektur BERT dibangun di atas mekanisme *self-attention*. Mekanisme ini menghitung hubungan antar token dalam teks menggunakan tiga komponen, yaitu *query* (Q), *key* (K), dan *value* (V). Rumus dasar *scaled dot-product attention* adalah sebagai berikut [8].

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.11}$$

di mana d_k adalah dimensi dari *key*. Mekanisme ini memungkinkan model menentukan bagian teks mana yang paling relevan terhadap token tertentu [8].

IndoBERT digunakan karena model berbasis BERT yang disesuaikan dengan bahasa Indonesia dapat mendukung tugas pemrosesan bahasa alami,

termasuk klasifikasi teks dan analisis sentimen bahasa Indonesia [9]. Penelitian terbaru mengenai IndoBERT menunjukkan bahwa representasi kontekstual berbasis BERT masih relevan untuk klasifikasi sentimen bahasa Indonesia [9]. Dalam penelitian ini, IndoBERT digunakan untuk mengubah teks berita ekonomi Indonesia menjadi vektor *embedding* berdimensi 768. Representasi tersebut kemudian digunakan sebagai fitur sentimen yang digabungkan dengan data teknikal IHSG.

2.5 *Zero-Shot Classification* Berbasis *Natural Language Inference*

Zero-shot classification adalah pendekatan klasifikasi teks yang memungkinkan model melakukan klasifikasi tanpa pelatihan langsung pada label target tertentu [17]. Salah satu pendekatan yang umum digunakan adalah memanfaatkan model *Natural Language Inference* (NLI). Pada NLI, model menentukan hubungan antara dua teks, yaitu *premise* dan *hypothesis*. Hubungan tersebut umumnya diklasifikasikan ke dalam tiga kategori, yaitu *entailment*, *contradiction*, dan *neutral* [17].

Dalam konteks *zero-shot classification*, teks berita diperlakukan sebagai *premise*, sedangkan label kandidat diubah menjadi bentuk hipotesis. Sebagai contoh, label "positif untuk pasar saham" dapat direpresentasikan sebagai hipotesis bahwa berita tersebut berdampak positif terhadap pasar saham. Model kemudian memilih label dengan probabilitas *entailment* tertinggi. Pendekatan ini sesuai untuk penelitian yang tidak memiliki label sentimen manual karena model dapat menggunakan relasi semantik antara teks dan label kandidat [17].

Secara umum, probabilitas label dapat dituliskan sebagai berikut [17].

$$\hat{y} = \arg\max_{c \in C} P(\textit{entailment} \mid x, h_c) \tag{2.12}$$

di mana x adalah teks berita, C adalah himpunan label kandidat, dan h_c adalah hipotesis yang dibentuk dari label c.

Pada penelitian ini, *zero-shot classification* digunakan untuk memberi label sentimen berita ekonomi Indonesia ke dalam tiga kategori, yaitu positif, negatif, dan netral terhadap pasar saham. Pendekatan ini dipilih karena dataset berita yang digunakan tidak memiliki label sentimen manual.

2.6 Normalisasi Fitur Numerik

Normalisasi diperlukan karena fitur teknikal memiliki rentang nilai yang berbeda. Harga IHSG berada pada skala ribuan, volume perdagangan dapat memiliki magnitudo yang jauh lebih besar, RSI memiliki rentang terbatas, sedangkan *daily return*, MACD, dan volatilitas memiliki nilai yang relatif kecil. Tanpa penyamaan skala, fitur dengan magnitudo numerik besar dapat memberi pengaruh yang tidak proporsional terhadap proses optimisasi jaringan saraf.

Penelitian ini menggunakan *MinMaxScaler* untuk memetakan setiap fitur ke rentang $[0,1]$. Metode ini mempertahankan urutan relatif nilai pada masing-masing fitur, tidak mensyaratkan data berdistribusi normal, dan memudahkan penerapan transformasi yang konsisten pada data latih, validasi, uji, serta data live. Penelitian perbandingan normalisasi pada LSTM multivariat menunjukkan bahwa normalisasi Min-Max dapat menghasilkan RMSE dan MAPE yang lebih rendah daripada normalisasi Z-score pada konteks data yang dievaluasi [18]. Meskipun demikian, hasil tersebut tidak berarti bahwa MinMaxScaler selalu menjadi metode terbaik untuk seluruh dataset. Dalam penelitian ini, MinMaxScaler dipilih sebagai konfigurasi pra-pemrosesan yang sesuai untuk menyamakan skala fitur heterogen; penelitian ini tidak mengklaim telah membuktikan keunggulannya terhadap StandardScaler atau RobustScaler tanpa eksperimen pembanding yang dikontrol.

Secara matematis, normalisasi Min-Max dirumuskan sebagai berikut.

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{2.13}$$

Nilai x_{min} dan x_{max} harus dihitung hanya dari data latih. Parameter tersebut kemudian digunakan untuk mentransformasi data validasi dan data uji agar informasi masa depan tidak masuk ke proses pelatihan.

2.7 *Principal Component Analysis*

Principal Component Analysis (PCA) adalah teknik reduksi dimensi linier yang digunakan untuk memproyeksikan data berdimensi tinggi ke ruang berdimensi lebih rendah dengan mempertahankan variasi data sebesar mungkin [19, 20]. PCA digunakan dalam penelitian ini untuk mereduksi dimensi *embedding* IndoBERT dari 768 dimensi menjadi jumlah komponen yang lebih kecil. Reduksi dimensi

diperlukan karena fitur teks berdimensi tinggi dapat meningkatkan kompleksitas komputasi dan risiko *overfitting* pada data terbatas [19].

Diberikan matriks data $X \in \mathbb{R}^{n \times d}$ yang telah dinormalisasi atau dikurangi rata-ratanya, PCA mencari vektor komponen utama yang memaksimalkan varians proyeksi data [19]. Proses ini dapat ditulis melalui dekomposisi berikut.

$$X = U\Sigma V^T \tag{2.14}$$

di mana U dan V adalah matriks ortogonal, sedangkan Σ berisi nilai singular. Jika hanya k komponen utama yang dipertahankan, maka proyeksi data menjadi sebagai berikut.

$$Z = XV_k \tag{2.15}$$

dengan V_k adalah matriks yang berisi k komponen utama pertama. Penelitian mengenai kompresi *embedding* BERT menunjukkan bahwa PCA dapat mengurangi dimensi representasi secara signifikan sambil tetap mempertahankan informasi yang berguna bagi tugas lanjutan dan meningkatkan efisiensi pelatihan [21].

Dalam penelitian ini, model regresi final menggunakan 128 komponen PCA. Konfigurasi tersebut mengurangi dimensi *embedding* IndoBERT dari 768 menjadi 128 dimensi atau sebesar sekitar 83,33%, tetapi tetap menyediakan representasi semantik yang relatif kaya sebelum proses fusi melalui *cross-attention*. Model klasifikasi berbasis BCE dan SMOTE menggunakan 32 komponen PCA, yang berarti pengurangan dimensi sekitar 95,83%. Kompresi yang lebih agresif digunakan pada model klasifikasi pembanding untuk membatasi kompleksitas input dan risiko *overfitting* pada jumlah sampel yang terbatas.

Nilai 32 dan 128 merupakan konfigurasi hyperparameter manual yang mempertimbangkan keseimbangan antara retensi informasi dan efisiensi komputasi. Kedua nilai tersebut tidak dinyatakan sebagai jumlah komponen yang optimal secara universal. Klaim optimalitas baru dapat dibuat apabila beberapa kandidat jumlah komponen dibandingkan menggunakan data validasi dan dilengkapi dengan nilai *cumulative explained variance*.

2.8 Mekanisme *Cross-Attention* untuk Fusi Multimodal

Penelitian ini menggunakan dua jenis data, yaitu data teknikal IHSG dan data sentimen berita ekonomi Indonesia. Kedua data tersebut memiliki karakteristik berbeda karena data teknikal berbentuk deret waktu numerik, sedangkan data berita berbentuk representasi semantik dari teks. Perbedaan karakteristik tersebut membuat proses fusi memerlukan mekanisme yang mampu menghubungkan representasi temporal dan representasi semantik secara adaptif [8, 5].

Cross-attention adalah mekanisme *attention* yang menggunakan *query* dari satu sumber data dan *key*/*value* dari sumber data lain [8]. Dalam penelitian ini, representasi LSTM digunakan sebagai *query*, sedangkan representasi sentimen IndoBERT digunakan sebagai *key* dan *value*. Mekanisme ini memungkinkan model mempelajari bagian sentimen yang relevan terhadap kondisi temporal tertentu pada data IHSG [4, 5]. Secara konseptual, mekanisme ini berlandaskan prinsip *scaled dot-product attention* pada arsitektur *Transformer*.

Rumus *cross-attention* yang digunakan adalah sebagai berikut [8].

$$\text{CrossAttention}(Q_{ts}, K_{sent}, V_{sent}) = \text{softmax}\left(\frac{Q_{ts}K_{sent}^{T}}{\sqrt{d_k}}\right)V_{sent} \tag{2.16}$$

di mana Q_{ts} adalah representasi data teknikal, sedangkan K_{sent} dan V_{sent} adalah representasi sentimen. Pada model final, mekanisme ini diterapkan dalam bentuk *multi-head attention* dengan empat kepala perhatian. Penggunaan beberapa kepala memungkinkan model mempelajari hubungan teknikal-sentimen dari beberapa subruang representasi secara paralel [8].

Setelah proses *attention*, penelitian ini juga menggunakan *residual connection* dengan menambahkan keluaran *attention* ke representasi LSTM. *Residual connection* membantu mempertahankan informasi temporal utama sekaligus menambahkan informasi sentimen yang dianggap relevan oleh mekanisme *attention*. Secara matematis, proses tersebut dapat ditulis sebagai berikut.

$$z_t = h_t^{LSTM} + \text{AttnOut}_t \tag{2.17}$$

2.9 *Explainable AI* dan *Integrated Gradients*

Model *deep learning* seperti LSTM, BERT, dan arsitektur *hybrid* berbasis *attention* sering memiliki tingkat kompleksitas yang tinggi sehingga sulit dijelaskan secara langsung. Oleh karena itu, *Explainable AI* (XAI) digunakan untuk membantu menganalisis kontribusi input terhadap keluaran model. Dalam konteks penelitian ini, XAI digunakan sebagai analisis tambahan untuk melihat kontribusi relatif antara fitur teknikal IHSG dan representasi sentimen berita ekonomi terhadap prediksi model.

Salah satu metode XAI yang digunakan adalah *Integrated Gradients*. Metode ini menghitung atribusi fitur berdasarkan integral gradien keluaran model terhadap input, dari suatu nilai acuan (*baseline*) menuju input aktual [22]. Secara umum, *Integrated Gradients* untuk fitur ke-i dapat dirumuskan sebagai berikut.

$$IG_i(x) = (x_i - x_i') \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha \tag{2.18}$$

di mana x adalah input aktual, x' adalah *baseline*, F adalah fungsi model, dan $IG_i(x)$ menunjukkan kontribusi fitur ke-i terhadap output model. Pada penelitian ini, *Integrated Gradients* diterapkan pada dua input model, yaitu input teknikal berbentuk sekuens LSTM dan input sentimen berbasis *embedding* IndoBERT yang telah direduksi menggunakan PCA.

Perlu diperhatikan bahwa hasil XAI pada penelitian ini digunakan sebagai analisis interpretabilitas tambahan, bukan sebagai bukti kausal bahwa suatu fitur secara langsung menyebabkan perubahan IHSG. Atribusi yang dihasilkan hanya menunjukkan kontribusi relatif input terhadap keputusan model pada sampel tertentu. Dengan demikian, hasil XAI membantu menjelaskan perilaku model, tetapi tidak menggantikan evaluasi utama seperti RMSE, MAPE, *Directional Accuracy*, *Balanced Accuracy*, MCC, dan simulasi strategi perdagangan.

2.10 Fungsi Kerugian `DirectionalMSELoss`

Model regresi pada penelitian ini tidak hanya diarahkan untuk meminimalkan kesalahan numerik, tetapi juga untuk memperhatikan ketepatan arah prediksi. Pendekatan evaluasi arah relevan dalam prediksi saham karena keputusan perdagangan sering kali bergantung pada ketepatan arah pergerakan harga [2, 3]. Oleh karena itu, penelitian ini menggunakan fungsi kerugian

`DirectionalMSELoss`, yaitu fungsi kerugian yang menggabungkan *Mean Squared Error* (MSE) dengan komponen arah berbasis *Binary Cross-Entropy with Logits*.

MSE digunakan untuk mengukur selisih kuadrat antara nilai aktual dan nilai prediksi. Rumus MSE dinyatakan sebagai berikut.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.19}$$

Sementara itu, komponen arah digunakan untuk memberi penalti ketika tanda prediksi tidak sesuai dengan arah aktual. Fungsi kerugian yang digunakan dalam penelitian ini dirumuskan sebagai berikut.

$$\mathscr{L} = \text{MSE}(\hat{y}, y) + \alpha \cdot \text{BCEWithLogits}(\hat{y}, \mathbb{I}(y > 0)) \tag{2.20}$$

dengan $\alpha = 0{,}3$. Komponen $\mathbb{I}(y > 0)$ menghasilkan label arah, yaitu 1 jika *log return* aktual bernilai positif dan 0 jika sebaliknya. Dengan demikian, fungsi kerugian ini mendorong model agar tetap akurat secara nilai regresi sekaligus lebih peka terhadap arah pergerakan IHSG.

2.11 SMOTE untuk Penanganan Ketidakseimbangan Kelas

Ketidakseimbangan kelas merupakan kondisi ketika jumlah data pada satu kelas lebih banyak dibandingkan kelas lainnya [23, 24]. Dalam prediksi arah pasar saham, ketidakseimbangan dapat menyebabkan model cenderung memprediksi kelas mayoritas dan gagal mempelajari kelas minoritas. SMOTE merupakan metode *oversampling* yang menghasilkan sampel sintetis pada kelas minoritas untuk membantu menyeimbangkan distribusi kelas [24, 23]. SMOTE membentuk sampel baru dengan melakukan interpolasi antara satu sampel minoritas dan tetangga terdekatnya [24].

$$x_{\text{sintetis}} = x_i + \lambda(\hat{x}_i - x_i), \quad \lambda \in [0, 1] \tag{2.21}$$

di mana x_i adalah sampel minoritas, $\hat{x}_i$ adalah salah satu tetangga terdekat dari kelas minoritas, dan λ adalah bilangan acak antara 0 dan 1. Dalam penelitian ini, SMOTE digunakan pada model klasifikasi pembanding, yaitu *Hybrid SMOTE Classifier*. Tujuannya adalah menguji apakah penyeimbangan kelas dapat

membantu model klasifikasi arah dalam memprediksi kelas naik dan turun secara lebih seimbang.

2.12 Evaluasi Deret Waktu dan Pencegahan *Data Leakage*

Evaluasi pada data deret waktu harus mempertahankan urutan waktu karena observasi masa depan tidak boleh digunakan untuk melatih model yang seharusnya hanya mengetahui informasi masa lalu. Praktik evaluasi yang tidak tepat pada data deret waktu dapat menyebabkan *look-ahead bias* atau *data leakage*, sehingga performa model terlihat lebih baik daripada kondisi sebenarnya [25]. Dalam konteks LSTM, kebocoran juga dapat terjadi apabila pembentukan sekuens, normalisasi, atau pemilihan parameter dilakukan dengan melibatkan data masa depan [26].

Oleh karena itu, penelitian ini menggunakan pembagian data secara kronologis dengan proporsi 70% data latih, 15% data validasi, dan 15% data uji. Pembagian ini berbeda dengan *k-fold cross-validation* acak yang umum digunakan pada data independen. Pada data deret waktu, *k-fold cross-validation* acak dapat merusak urutan temporal karena data masa depan dapat masuk ke proses pelatihan pada lipatan tertentu. Alternatif yang lebih sesuai untuk penelitian lanjutan adalah *walk-forward validation* atau *rolling-window validation*; namun, penelitian ini menggunakan satu pembagian kronologis karena fokus utama penelitian adalah membandingkan integrasi LSTM dan IndoBERT pada pipeline yang konsisten.

2.13 ARIMA sebagai Model Pembanding Statistik

Autoregressive Integrated Moving Average (ARIMA) adalah model statistik klasik untuk prediksi deret waktu [10]. ARIMA terdiri dari tiga komponen, yaitu *Autoregressive* (AR), *Integrated* (I), dan *Moving Average* (MA) [10]. Model ARIMA digunakan sebagai pembanding karena model statistik deret waktu masih sering dijadikan dasar evaluasi terhadap model prediksi modern [10, 2].

Model ARIMA ditulis dalam bentuk umum sebagai berikut [10].

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t \tag{2.22}$$

di mana B adalah operator *backshift*, d adalah orde diferensiasi, $\phi(B)$ adalah komponen autoregresif, $\theta(B)$ adalah komponen *moving average*, dan ε_t adalah galat

acak. Dalam penelitian ini, ARIMA digunakan sebagai model pembanding statistik terhadap model berbasis *deep learning*. Penggunaan ARIMA bertujuan untuk memberikan pembanding klasik pada data deret waktu IHSG, sehingga performa model *Hybrid* LSTM-IndoBERT dapat dievaluasi secara lebih objektif.

2.14 Metrik Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metrik regresi dan metrik klasifikasi arah. Metrik regresi digunakan untuk menilai kesalahan nilai prediksi, sedangkan metrik klasifikasi arah digunakan untuk menilai kemampuan model dalam membedakan arah naik dan turun. Penggunaan beberapa metrik diperlukan karena performa model prediksi finansial tidak hanya dinilai dari kedekatan nilai prediksi, tetapi juga dari kemampuan model menentukan arah pergerakan pasar [2, 3].

2.14.1 *Root Mean Squared Error*

Root Mean Squared Error (RMSE) mengukur akar dari rata-rata kuadrat kesalahan prediksi. RMSE digunakan untuk melihat seberapa jauh nilai prediksi menyimpang dari nilai aktual dalam satuan yang sama dengan target. RMSE dirumuskan sebagai berikut.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.23}$$

Nilai RMSE yang lebih kecil menunjukkan bahwa prediksi model lebih dekat terhadap nilai aktual.

2.14.2 *Mean Absolute Percentage Error*

Mean Absolute Percentage Error (MAPE) mengukur kesalahan prediksi dalam bentuk persentase relatif terhadap nilai aktual. MAPE digunakan karena hasilnya mudah diinterpretasikan sebagai persentase kesalahan prediksi. MAPE dirumuskan sebagai berikut.

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2.24}$$

2.14.3 *Directional Accuracy*

Directional Accuracy (DA) mengukur proporsi prediksi yang berhasil menebak arah pergerakan dengan benar [2]. Dalam konteks prediksi pasar saham, DA menjadi metrik penting karena keputusan perdagangan banyak bergantung pada arah pergerakan harga [2, 3]. DA dirumuskan sebagai berikut.

$$\text{DA} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[\text{sign}(\hat{y}_i) = \text{sign}(y_i)] \times 100\% \tag{2.25}$$

2.14.4 *Balanced Accuracy*

Balanced Accuracy digunakan untuk mengevaluasi performa klasifikasi pada data yang mungkin memiliki distribusi kelas tidak seimbang [27]. Metrik ini menghitung rata-rata *recall* dari setiap kelas, sehingga lebih sesuai dibandingkan *accuracy* biasa ketika jumlah kelas tidak seimbang [27]. *Balanced Accuracy* dirumuskan sebagai berikut.

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{2.26}$$

Nilai 0,5 menunjukkan performa setara tebakan acak pada klasifikasi biner yang seimbang.

2.14.5 *Matthews Correlation Coefficient*

Matthews Correlation Coefficient (MCC) adalah metrik evaluasi klasifikasi biner yang mempertimbangkan seluruh elemen *confusion matrix*, yaitu TP, TN, FP, dan FN [27]. MCC dinilai lebih informatif dibandingkan *accuracy* biasa ketika data memiliki ketidakseimbangan kelas [27]. MCC dirumuskan sebagai berikut.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{2.27}$$

Nilai MCC berada pada rentang $[-1, 1]$, di mana 1 menunjukkan prediksi sempurna, 0 menunjukkan prediksi setara acak, dan -1 menunjukkan prediksi berlawanan sepenuhnya dengan aktual [27].

2.14.6 F1-*Score*

F1-*Score* merupakan rata-rata harmonik antara *precision* dan *recall*. Metrik ini digunakan untuk melihat keseimbangan antara kemampuan model menghasilkan prediksi positif yang tepat dan kemampuan model menangkap kelas positif. F1-*Score* dirumuskan sebagai berikut.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.28}$$

F1-*Score* digunakan sebagai metrik tambahan pada evaluasi klasifikasi arah untuk memberikan gambaran performa model selain *Balanced Accuracy* dan MCC.

2.15 Uji McNemar

Uji McNemar adalah uji statistik yang digunakan untuk membandingkan dua model klasifikasi pada data uji yang sama. Uji ini berfokus pada sampel yang diprediksi berbeda oleh kedua model atau disebut *discordant pairs*. Dalam evaluasi model klasifikasi, uji McNemar dapat digunakan untuk melihat apakah perbedaan jumlah prediksi benar dan salah antar dua model terjadi secara signifikan [28].

Tabel kontingensi McNemar dapat dinyatakan sebagai berikut.

Tabel 2.1. Tabel Kontingensi Uji McNemar

	Model B Benar	**Model B Salah**
Model A Benar	n_{11}	n_{10}
Model A Salah	n_{01}	n_{00}

Statistik uji McNemar dengan koreksi kontinuitas dapat ditulis sebagai berikut [28].

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \tag{2.29}$$

Jika jumlah *discordant pairs* kecil, maka digunakan *exact McNemar test*. Dalam penelitian ini, uji McNemar digunakan untuk melihat apakah perbedaan prediksi antara model LSTM tanpa sentimen dan model *Hybrid* berbasis sentimen signifikan secara statistik.

2.16 Penelitian Terdahulu

Penelitian mengenai prediksi pasar saham telah berkembang dari pendekatan statistik klasik menuju pendekatan berbasis *machine learning* dan *deep learning* [2, 3]. Perkembangan tersebut dipengaruhi oleh karakteristik data finansial yang tidak stasioner, memiliki volatilitas tinggi, serta dipengaruhi oleh faktor eksternal seperti berita, sentimen investor, dan kondisi ekonomi [2, 4]. Pada penelitian berbasis *deep learning*, LSTM banyak digunakan karena kemampuannya dalam mempelajari pola sekuensial dan ketergantungan jangka panjang pada data deret waktu [6, 7]. Sementara itu, pada sisi pemrosesan teks, model berbasis *Transformer* dan BERT mampu menghasilkan representasi kontekstual yang relevan untuk analisis teks [8, 16].

Dalam konteks prediksi pasar saham, beberapa penelitian terdahulu menunjukkan bahwa penggabungan data harga historis dengan informasi tekstual dapat membantu model menangkap sinyal pasar yang tidak terlihat pada data teknikal saja [2, 4, 5]. Xiao dan Ihnaini [2] menggunakan analisis sentimen untuk membantu prediksi tren saham dan menunjukkan bahwa sentimen dapat menjadi fitur tambahan dalam prediksi pasar. Akyüz, Karadayı Ataş, dan Türkmen [3] mengkaji prediksi pasar saham berbasis analisis sentimen dan *deep learning*. Gu dkk. [4] mengintegrasikan sentimen berita finansial dengan LSTM untuk prediksi harga saham. Li dan Hu [5] mengembangkan kerangka *hybrid deep learning* yang mempertimbangkan sentimen investor dari forum daring.

Selain itu, penelitian terkait representasi bahasa Indonesia menunjukkan bahwa model berbasis IndoBERT relevan digunakan untuk tugas klasifikasi teks dan analisis sentimen bahasa Indonesia [9]. Saputra dkk. [9] mengembangkan pendekatan klasifikasi sentimen bahasa Indonesia berbasis IndoBERT yang memperhatikan konteks topik. Temuan tersebut mendukung penggunaan IndoBERT dalam penelitian ini untuk mengekstraksi representasi sentimen dari berita ekonomi Indonesia.

Dari sisi metode klasifikasi tanpa pelatihan langsung pada label target, Laurer dkk. [17] menunjukkan bahwa pendekatan NLI dapat digunakan untuk membangun pengklasifikasi universal. Pendekatan ini relevan dengan penelitian ini karena dataset berita ekonomi yang digunakan tidak memiliki label sentimen manual. Oleh karena itu, diperlukan metode *zero-shot classification* untuk mengklasifikasikan berita menjadi sentimen positif, negatif, dan netral terhadap pasar saham [17].

Ringkasan penelitian terdahulu yang relevan dengan penelitian ini disajikan pada Tabel 2.2.

Tabel 2.2. Ringkasan Penelitian Terdahulu

No.	**Peneliti**	**Fokus dan Metode Penelitian**	**Relevansi dan Celah Penelitian**
1	Qian Xiao dan Baraa Ihnaini [2]	Meneliti prediksi tren saham menggunakan analisis sentimen. Penelitian ini menunjukkan bahwa sentimen dapat digunakan sebagai fitur tambahan untuk membantu prediksi pasar saham.	Relevan karena menunjukkan hubungan antara sentimen dan prediksi pasar. Namun, penelitian tersebut belum berfokus pada IHSG dan belum menggunakan IndoBERT untuk memproses berita ekonomi Indonesia.
2	Sevi Özögür Akyüz, Pınar Karadayı Ataş, dan Selen Türkmen [3]	Mengkaji prediksi pasar saham berbasis analisis sentimen dan *deep learning*. Penelitian ini menggabungkan informasi tekstual dengan model prediksi pasar.	Relevan sebagai dasar integrasi data tekstual dan data pasar. Namun, penelitian tersebut belum secara spesifik menggunakan IndoBERT pada konteks pasar modal Indonesia.
3	Wenjun Gu, Yihao Zhong, Shizun Li, Changsong Wei, Liting Dong, Zhuoyue Wang, dan Chao Yan [4]	Mengembangkan pendekatan prediksi harga saham dengan mengintegrasikan analisis sentimen berita finansial dan LSTM.	Relevan karena menggunakan kombinasi sentimen berita dan LSTM. Namun, penelitian tersebut menggunakan konteks berita finansial umum dan belum berfokus pada IHSG maupun berita ekonomi Indonesia.
4	Huiyu Li dan Junhua Hu [5]	Mengembangkan kerangka *hybrid deep learning* untuk prediksi saham dengan mempertimbangkan sentimen investor dari forum daring.	Relevan karena menunjukkan peran sentimen investor dalam prediksi saham. Namun, sumber sentimen yang digunakan bukan berita ekonomi Indonesia dan belum menggunakan IndoBERT.

No.	Peneliti	Fokus dan Metode Penelitian	Relevansi dan Celah Penelitian
5	Benyamin Ghojogh dan Ali Ghodsi [6]	Membahas konsep RNN dan LSTM, termasuk mekanisme gerbang, pemrosesan data sekuensial, dan varian LSTM.	Menjadi dasar teori penggunaan LSTM untuk menangkap pola temporal IHSG. Namun, penelitian tersebut tidak membahas integrasi LSTM dengan sentimen berita.
6	Yi Tay, Mostafa Dehghani, Dara Bahri, dan Donald Metzler [8]	Menyajikan survei mengenai perkembangan model *Transformer* dan mekanisme *attention*.	Relevan sebagai dasar teori mekanisme *attention*. Namun, penelitian tersebut tidak secara spesifik membahas prediksi IHSG berbasis sentimen berita ekonomi Indonesia.
7	Muhammad Apriandito Arya Saputra, Andry Alamsyah, Dian Puteri Ramadhani, Thomhert Suprapto Siadari, dan Hanif Fakhrurroja [9]	Mengembangkan pendekatan klasifikasi sentimen bahasa Indonesia berbasis IndoBERT dengan mempertimbangkan konteks topik.	Relevan dengan penggunaan IndoBERT untuk teks bahasa Indonesia. Namun, penelitian tersebut tidak diterapkan pada prediksi pasar saham atau IHSG.
8	Moritz Laurer, Wouter van Atteveldt, Andreu Casas, dan Kasper Welbers [17]	Membahas klasifikasi teks berbasis NLI untuk membangun pengklasifikasi universal.	Relevan untuk pelabelan sentimen secara *zero-shot*. Namun, pendekatan tersebut belum digunakan untuk prediksi IHSG berbasis berita ekonomi Indonesia.
9	Davide Chicco, Matthijs J. Warrens, dan Giuseppe Jurman [27]	Membahas penggunaan MCC sebagai metrik evaluasi klasifikasi biner yang mempertimbangkan seluruh komponen *confusion matrix*.	Relevan karena prediksi arah pasar dapat mengalami ketidakseimbangan kelas. Namun, penelitian tersebut berfokus pada metrik evaluasi, bukan prediksi pasar saham.

No.	Peneliti	Fokus dan Metode Penelitian	Relevansi dan Celah Penelitian
10	Damien Dablain, Bartosz Krawczyk, dan Nitesh V. Chawla [24]	Membahas penggunaan SMOTE dalam konteks data tidak seimbang dan pembelajaran mendalam.	Relevan sebagai dasar penggunaan SMOTE pada model klasifikasi pembanding. Namun, penelitian tersebut tidak berfokus pada prediksi IHSG.

Selain penelitian umum mengenai integrasi sentimen dan data teknikal, terdapat beberapa penelitian yang lebih dekat dengan konteks penelitian ini, baik karena menggunakan IHSG maupun karena menggabungkan data teknikal dengan sentimen berbasis model bahasa. Perbandingan tersebut dirangkum pada Tabel 2.3.

Tabel 2.3. Perbandingan Penelitian Terkait Prediksi Saham dan IHSG

No.	Penelitian	Data dan Metode	Hasil Utama	Posisi Penelitian Ini
1	Rasyid dkk. [29]	Prediksi IHSG menggunakan data Yahoo Finance dan model LSTM/GRU.	LSTM memperoleh RMSE 71,2896 dan GRU memperoleh RMSE 70,6187 pada skala harga.	Penelitian ini juga menggunakan IHSG dari Yahoo Finance, tetapi target utama berupa *log return* dan evaluasi arah dengan DA, *Balanced Accuracy*, MCC, serta simulasi strategi.
2	Majid dkk. [30]	Prediksi IHSG berbasis Bi-LSTM dengan faktor indeks global.	Bi-LSTM memperoleh MAPE 0,572314%, lebih baik daripada LSTM 0,74326%.	Penelitian ini memperoleh MAPE 0,3626% pada evaluasi final dan menambahkan berita ekonomi Indonesia berbasis IndoBERT.

No.	Penelitian	Data dan Metode	Hasil Utama	Posisi Penelitian Ini
3	Saputra dkk. [31]	Prediksi pergerakan saham dengan LSTM menggunakan indikator teknikal, fundamental, IHSG, kurs, dan sentimen Twitter berbasis IndoBERT.	IndoBERT untuk sentimen memperoleh akurasi 0,69; model terbaik setelah optimisasi menghasilkan RMSE 332,66 pada data uji.	Penelitian ini menggunakan IndoBERT untuk berita ekonomi, bukan Twitter, dan menggabungkannya dengan LSTM melalui *cross-attention*.
4	Ko dan Chang [32]	Prediksi harga saham menggunakan data transaksi historis dan sentimen teks berbasis BERT-LSTM.	Integrasi sentimen dan data teknikal meningkatkan performa RMSE dibandingkan model pembanding.	Mendukung gagasan bahwa informasi teks dapat memperkaya model prediksi saham; penelitian ini mengadaptasi gagasan tersebut pada konteks IHSG dan bahasa Indonesia.
5	Yang dkk. [11]	Prediksi arah saham dengan gabungan indikator teknikal dan sentimen finansial, disertai seleksi variabel LASSO-LSTM.	Penggunaan indikator teknikal dan sentimen meningkatkan akurasi rata-rata sebesar 8,53% dibandingkan LSTM standar.	Penelitian ini tidak menggunakan LASSO, tetapi melakukan *domain-driven feature engineering* dan menambahkan *cross-attention* untuk fusi teknikal-sentimen.
6	Gona dkk. [33]	Model LSTM-Transformer dengan indikator teknikal dan sentimen FinBERT.	Menghasilkan *directional accuracy* 76,4% pada konteks data yang digunakan.	Penelitian ini memperoleh DA 62,72% pada IHSG; hasil tidak dibandingkan langsung karena dataset, pasar, fitur, dan metrik implementasi berbeda.

Berdasarkan perbandingan tersebut, model yang dikembangkan dalam penelitian ini dapat dikatakan kompetitif dalam konteks IHSG karena MAPE

final sebesar 0,3626% lebih rendah daripada beberapa penelitian IHSG berbasis LSTM/Bi-LSTM yang dilaporkan pada skala MAPE. Namun, perbandingan antarpenelitian tetap harus dilakukan secara hati-hati karena perbedaan periode data, target prediksi, horizon prediksi, sumber fitur, dan cara pembagian data dapat memengaruhi hasil evaluasi.

Berdasarkan Tabel 2.2, terdapat beberapa celah penelitian yang diidentifikasi. Pertama, penelitian terdahulu telah menunjukkan bahwa sentimen dapat membantu prediksi pasar saham, tetapi belum banyak studi yang secara spesifik mengintegrasikan sentimen berita ekonomi Indonesia dengan data teknikal IHSG menggunakan IndoBERT dan LSTM [2, 4, 5]. Kedua, penelitian terdahulu yang menggunakan model berbasis sentimen umumnya belum memanfaatkan mekanisme *cross-attention* untuk menghubungkan representasi temporal dari data harga dengan representasi semantik dari berita [8, 5]. Ketiga, penelitian terkait IndoBERT umumnya berfokus pada tugas klasifikasi teks bahasa Indonesia, tetapi belum secara langsung mengevaluasi kontribusi representasi IndoBERT terhadap prediksi arah IHSG [9].

Penelitian ini hadir untuk mengisi celah tersebut dengan mengembangkan model *Cross-Attention Hybrid* LSTM-IndoBERT. Model ini mengintegrasikan 12 fitur teknikal IHSG dengan representasi sentimen berita ekonomi Indonesia berbasis IndoBERT. Dataset berita yang digunakan terdiri atas 3.592 artikel dan 1.274 tanggal unik yang memiliki berita. Selain itu, penelitian ini tidak hanya mengevaluasi performa menggunakan metrik regresi seperti RMSE dan MAPE, tetapi juga menggunakan metrik arah seperti *Directional Accuracy*, *Balanced Accuracy*, MCC, Uji McNemar, dan simulasi strategi perdagangan. Dengan demikian, penelitian ini diharapkan dapat memberikan evaluasi yang lebih komprehensif terhadap kontribusi sentimen berita ekonomi Indonesia dalam prediksi pergerakan IHSG.