

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Indeks Harga Saham Gabungan (IHSG) merupakan indeks yang mengukur kinerja harga saham dari seluruh perusahaan tercatat pada papan utama dan papan pengembangan di Bursa Efek Indonesia [1]. IHSG dapat digunakan sebagai representasi pergerakan pasar modal Indonesia karena indeks pasar saham mencerminkan perubahan harga secara agregat dari sekumpulan saham yang tercatat di bursa [1, 2]. Pergerakan IHSG menjadi penting untuk dianalisis karena perubahan arah indeks dapat memengaruhi keputusan investasi, manajemen risiko, dan strategi perdagangan di pasar modal [2, 3].

Prediksi pasar saham merupakan permasalahan yang menantang karena data finansial bersifat tidak stasioner, fluktuatif, dan dipengaruhi oleh berbagai faktor eksternal [2, 3]. Pergerakan harga tidak hanya dipengaruhi oleh data historis, tetapi juga oleh informasi tekstual seperti berita ekonomi dan sentimen investor [4, 5]. Penelitian terdahulu menunjukkan bahwa integrasi data pasar dan sentimen teks dapat membantu meningkatkan kemampuan model dalam memprediksi tren saham [2, 4]. Oleh karena itu, berita ekonomi Indonesia dapat dimanfaatkan sebagai sumber informasi tambahan dalam membangun model prediksi pergerakan IHSG.

Salah satu model yang banyak digunakan dalam prediksi deret waktu finansial adalah *Long Short-Term Memory* (LSTM) karena mampu mempelajari pola temporal dan ketergantungan jangka panjang pada data sekuensial [6, 7]. Namun, LSTM yang hanya menggunakan data teknikal belum secara langsung memanfaatkan informasi sentimen dari berita ekonomi. Di sisi lain, model bahasa berbasis *Transformer* seperti BERT mampu menghasilkan representasi teks kontekstual, sedangkan IndoBERT dapat digunakan untuk memproses teks berbahasa Indonesia [8, 9]. Dengan demikian, integrasi LSTM dan IndoBERT dapat menjadi pendekatan yang relevan untuk menggabungkan pola historis IHSG dan sentimen berita ekonomi Indonesia.

Integrasi data teknikal dan data sentimen tidak cukup hanya dilakukan dengan penggabungan fitur secara langsung karena kedua jenis data memiliki karakteristik yang berbeda [4, 5]. Data teknikal berbentuk deret waktu numerik, sedangkan berita ekonomi direpresentasikan dalam bentuk fitur semantik dari teks.

Oleh karena itu, penelitian ini menggunakan pendekatan *Cross-Attention Hybrid* LSTM-IndoBERT untuk menghubungkan representasi temporal dari data IHSG dengan representasi sentimen dari berita ekonomi Indonesia. Mekanisme *cross-attention* digunakan agar model dapat mempelajari hubungan antara sinyal teknikal dan informasi sentimen secara lebih adaptif [8, 4].

Berdasarkan permasalahan tersebut, penelitian ini mengangkat judul **“Prediksi Indeks Harga Saham Gabungan (IHSG) Berbasis Sentimen Berita Ekonomi Indonesia Menggunakan Integrasi IndoBERT dan LSTM”**. Penelitian ini berfokus pada pengembangan dan evaluasi model *Hybrid* LSTM-IndoBERT untuk memprediksi pergerakan IHSG, serta mengukur kontribusi fitur sentimen berita ekonomi Indonesia terhadap akurasi prediksi dibandingkan model LSTM tanpa sentimen.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana kinerja model *Hybrid* LSTM-IndoBERT dalam memprediksi pergerakan IHSG?
2. Sejauh mana integrasi fitur sentimen berita ekonomi Indonesia menggunakan IndoBERT berkontribusi terhadap akurasi prediksi dibandingkan model LSTM tanpa sentimen?

1.3 Batasan Masalah

Agar penelitian ini lebih terfokus, batasan masalah yang digunakan adalah sebagai berikut.

1. Objek penelitian terbatas pada Indeks Harga Saham Gabungan (IHSG).
2. Periode data yang digunakan adalah 1 Januari 2020 hingga 31 Desember 2024.
3. Data historis IHSG yang digunakan berupa data harian OHLCV, yaitu *Open*, *High*, *Low*, *Close*, dan *Volume*.
4. Fitur teknikal yang digunakan terdiri dari 12 fitur, yaitu *open*, *high*, *low*, *close*, *volume*, *RSI*, *MACD*, *MACD signal*, *MACD diff*, *Bollinger Band Width*, *daily return*, dan volatilitas.

5. Data tekstual yang digunakan berupa berita ekonomi Indonesia yang dikumpulkan dari beberapa portal berita daring.
6. Representasi teks berita diekstraksi menggunakan IndoBERT dan direduksi dimensinya menggunakan PCA.
7. Model utama yang dikembangkan adalah *Cross-Attention Hybrid LSTM-IndoBERT*.
8. Evaluasi model dilakukan menggunakan metrik RMSE, MAPE, *Directional Accuracy*, *Balanced Accuracy*, MCC, *F1-score*, uji McNemar, dan simulasi strategi perdagangan.
9. Penelitian ini tidak bertujuan untuk memberikan rekomendasi investasi, melainkan berfokus pada evaluasi performa model prediksi.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Menganalisis kinerja model *Hybrid LSTM-IndoBERT* dalam memprediksi pergerakan IHSG.
2. Mengukur kontribusi integrasi fitur sentimen berita ekonomi Indonesia menggunakan IndoBERT terhadap akurasi prediksi dibandingkan model LSTM tanpa sentimen.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. **Bagi peneliti:** Menjadikan penelitian ini sebagai referensi untuk pengembangan model prediksi pasar saham yang menggabungkan data teknikal dan sentimen berita ekonomi Indonesia.
2. **Bagi akademisi:** Menambah referensi penelitian mengenai penerapan LSTM, IndoBERT, dan mekanisme *cross-attention* dalam prediksi pergerakan IHSG.
3. **Bagi pengembang sistem prediksi finansial:** Memberikan gambaran rancangan model berbasis data numerik dan teks yang dapat digunakan sebagai dasar pengembangan sistem prediksi pasar modal.

1.6 Sistematika Penulisan

Sistematika penulisan Skripsi ini adalah sebagai berikut.

1. Bab 1 Pendahuluan

Bab ini menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. Bab 2 Landasan Teori

Bab ini membahas teori yang mendasari penelitian, meliputi IHSG, deret waktu finansial, LSTM, IndoBERT, *zero-shot classification*, PCA, *cross-attention*, fungsi kerugian, model pembandingan, dan metrik evaluasi.

3. Bab 3 Metodologi Penelitian

Bab ini menjelaskan tahapan penelitian, mulai dari pengumpulan data, pra-pemrosesan data, ekstraksi fitur sentimen, pembentukan sekuens, perancangan model, pelatihan model, hingga evaluasi.

4. Bab 4 Hasil dan Diskusi

Bab ini menyajikan hasil implementasi, hasil evaluasi model, perbandingan dengan model pembandingan, uji statistik, simulasi strategi perdagangan, dan pembahasan hasil penelitian.

5. Bab 5 Simpulan dan Saran

Bab ini berisi simpulan yang menjawab rumusan masalah serta saran untuk pengembangan penelitian selanjutnya.

