

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi ini.

1.1 Latar Belakang

Pasar modal Indonesia menyediakan berbagai instrumen investasi melalui Bursa Efek Indonesia (BEI). Saham menjadi salah satu instrumen yang paling aktif diperdagangkan di pasar modal Indonesia. Investor umumnya memperhatikan saham berkapitalisasi besar atau *blue-chip* karena saham tersebut memiliki likuiditas tinggi, kapitalisasi pasar besar, dan fundamental perusahaan yang relatif kuat. Indeks LQ45 mencerminkan karakteristik tersebut karena indeks ini berisi saham-saham perusahaan dengan likuiditas tinggi dan kapitalisasi pasar besar. Oleh karena itu, indeks LQ45 sering digunakan sebagai acuan dalam analisis pasar saham Indonesia (Syukur & Istiawan, 2021).

Pergerakan harga saham memiliki karakteristik *non-linear* dan *non-stationary*. Dinamika pasar yang kompleks juga memengaruhi perubahan harga saham. Karakteristik tersebut membuat harga saham sulit dimodelkan menggunakan pendekatan konvensional. Kondisi ini menunjukkan bahwa prediksi harga saham memerlukan metode yang mampu menangkap pola data secara lebih efektif (Syukur & Istiawan, 2021). Prediksi harga saham menjadi penting karena hasil prediksi dapat digunakan sebagai dasar dalam pengambilan keputusan investasi (Ahmed, 2020). Prediksi yang lebih akurat juga dapat membantu meningkatkan kualitas alokasi portofolio melalui estimasi *return* yang lebih tepat (Bessler & Wolff, 2024).

Kebutuhan prediksi harga saham menjadi semakin menantang ketika investor tidak hanya membutuhkan prediksi untuk satu hari ke depan. Investor dalam praktik jangka pendek memerlukan proyeksi harga untuk beberapa hari ke depan agar keputusan dapat dilakukan dengan lebih terarah. Proyeksi tersebut juga membantu investor dalam mengelola risiko ketika kondisi pasar berubah. Namun, pasar saham memiliki karakteristik yang kompleks karena pergerakan harga bersifat dinamis, *non-linear*, dan mengandung *noise*

jangka pendek. Kondisi tersebut dapat menutupi pola utama pergerakan harga, sehingga proses prediksi menjadi tidak sederhana (Abdullah dkk., 2024). Oleh karena itu, prediksi harga saham secara *multi-horizon* membutuhkan model yang mampu menangkap hubungan pergerakan harga dalam jangka pendek dan jangka panjang.

Model prediksi saham perlu mengolah berbagai jenis data secara bersamaan. Data tersebut dapat berupa riwayat harga saham, informasi waktu perdagangan, dan informasi dari luar data harga yang dapat memengaruhi pergerakan pasar (Lim dkk., 2021). Penelitian berbasis data numerik umumnya menggunakan data harga dan volume sebagai dasar pembentukan fitur. Data tersebut kemudian diolah menjadi indikator teknikal untuk menggambarkan tren, momentum, dan volatilitas pasar. Nti dkk. (2020) menunjukkan bahwa indikator teknikal merupakan salah satu jenis fitur yang paling sering digunakan dalam penelitian prediksi harga saham dan telah diterapkan pada berbagai pendekatan pemodelan.

Informasi dari luar pasar dapat berupa berita keuangan dan berita korporasi. Berita tersebut dapat memengaruhi ekspektasi investor dalam menilai kondisi perusahaan. Dampak dari informasi tersebut juga dapat tercermin pada respons pasar. Oleh sebab itu, berita dapat digunakan sebagai pelengkap indikator teknikal agar kondisi pasar dapat digambarkan secara lebih menyeluruh (Chen dkk., 2024). Pemanfaatan berita dalam prediksi saham umumnya dilakukan melalui ekstraksi sentimen, yaitu proses mengubah teks berita menjadi nilai numerik yang dapat digunakan sebagai fitur model. Namun, ekstraksi sentimen finansial bukan proses yang sederhana karena teks keuangan bersifat kompleks, tidak terstruktur, dan sangat bergantung pada konteks (Fatouros dkk., 2023).

Perkembangan *Large Language Model* (LLM) memberikan peluang untuk mengolah sentimen berita secara lebih baik. Model ini mampu memahami konteks dan hubungan makna dalam teks secara mendalam, sehingga penilaian sentimen dapat dilakukan dengan lebih sesuai terhadap isi berita (Yang dkk., 2020). Beberapa penelitian juga menunjukkan bahwa model bahasa seperti GPT mampu menilai implikasi ekonomi dari berita dan menghasilkan sinyal yang relevan untuk memprediksi reaksi pasar (Kırtac & Germano, 2024; Lopez-Lira & Tang, 2023).

Temporal Fusion Transformer (TFT) merupakan salah satu model berbasis *Transformer* yang dirancang untuk prediksi deret waktu secara *multi-horizon*. Model ini

mampu mengolah deret waktu multivariat, memanfaatkan berbagai jenis data masukan, serta menyediakan kemampuan penjelasan model melalui mekanisme *variable selection*, *gating*, dan *attention* (Lim dkk., 2021). Karakteristik tersebut membuat TFT sesuai untuk menangani permasalahan prediksi saham yang kompleks, melibatkan banyak variabel, dan sensitif terhadap perubahan kondisi pasar.

Penelitian yang mengintegrasikan sentimen berita berbasis LLM dengan TFT masih perlu diteliti lebih lanjut, khususnya pada prediksi harga saham blue-chip di Bursa Efek Indonesia. Penelitian yang membandingkan TFT berbasis indikator teknikal dengan TFT yang diperkaya sentimen berita berbasis LLM dalam rancangan evaluasi yang sama juga masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengembangkan model LLM-TFT dan membandingkannya dengan TFT dalam prediksi harga saham secara multi-horizon. Perbandingan tersebut dilakukan untuk mengetahui apakah penambahan fitur sentimen berita berbasis LLM dapat meningkatkan kinerja prediksi harga saham dibandingkan penggunaan indikator teknikal saja.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan masalah yang diangkat pada penelitian ini adalah bagaimana mengembangkan model *Large Language Model-Temporal Fusion Transformer* (LLM-TFT) untuk prediksi harga penutupan saham masing-masing emiten di Bursa Efek Indonesia secara *multi-horizon*, serta menguji pengaruh penambahan fitur sentimen berita berbasis LLM terhadap kinerja prediksi dibandingkan dengan model *Temporal Fusion Transformer* (TFT) berbasis indikator teknikal.

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah mengembangkan model *Large Language Model-Temporal Fusion Transformer* (LLM-TFT) untuk memprediksi harga penutupan saham emiten di Bursa Efek Indonesia secara *multi-horizon*. Selain itu, penelitian ini bertujuan membandingkan kinerja model *Temporal Fusion Transformer* (TFT) berbasis indikator teknikal dengan model LLM-TFT yang diperkaya fitur sentimen berita berbasis LLM.

Manfaat yang diharapkan dari penelitian ini adalah memberikan gambaran mengenai pengaruh sentimen berita berbasis LLM terhadap kinerja prediksi harga saham serta

mendukung pengembangan sistem prediksi yang lebih adaptif terhadap perubahan kondisi pasar modal.

1.4 Ruang Lingkup

Ruang lingkup dirumuskan dengan tujuan agar dapat memberikan target pada tugas akhir ini agar memiliki batasan dan tidak menyimpang sehingga sesuai dengan target yang diinginkan. Berikut merupakan ruang lingkup pada tugas akhir ini:

1. Penelitian menggunakan data harga saham harian empat emiten *blue-chip* di Bursa Efek Indonesia (BEI), yaitu ASII, BBRI, BMRI, dan TLKM.
2. Prediksi harga saham dilakukan menggunakan data harian dengan horizon prediksi H+1, H+2, dan H+3.
3. Model yang digunakan dalam penelitian adalah *Temporal Fusion Transformer* (TFT) dan *Large Language Model-Temporal Fusion Transformer* (LLM-TFT).
4. Evaluasi model dilakukan menggunakan metrik MAE, RMSE, MAPE, R^2 , dan *Directional Accuracy* (DirAcc).

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian tugas akhir ini terbagi dalam beberapa bab, yaitu:

BAB I PENDAHULUAN

Bab ini memaparkan latar belakang penelitian, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan sebagai gambaran awal mengenai arah dan batasan penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan landasan teori dan studi terdahulu yang relevan, meliputi konsep pasar modal dan karakteristik saham *blue-chip*, dasar-dasar prediksi deret waktu (*time series forecasting*), indikator teknikal yang digunakan dalam penelitian, konsep analisis sentimen pada berita finansial, pemanfaatan *Large Language Model* (LLM) untuk ekstraksi/klasifikasi sentimen, serta teori dan komponen utama *Temporal Fusion Transformer*

(TFT) untuk prediksi multi-horizon beserta aspek interpretabilitasnya (*attention* dan *variable selection*).

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan rancangan penelitian dan tahapan pelaksanaan secara sistematis, mencakup sumber data harga saham dan data berita, metode pengumpulan data (misalnya dari Yahoo Finance dan RSS), proses prapemrosesan data (pembersihan, penyelarasan waktu, penanganan data hilang), pembentukan fitur indikator teknikal, proses penentuan label/skor sentimen berita menggunakan LLM dan proses agregasi harian, pembentukan dataset TFT dan dataset LLM-TFT, pembagian data (train/validation/test), rancangan skenario prediksi multi-horizon, konfigurasi dan prosedur pelatihan model TFT, serta metode evaluasi kinerja menggunakan metrik MAE, RMSE, MAPE, R^2 , dan DirAcc beserta prosedur perbandingan hasil.

BAB IV HASIL DAN ANALISIS

Bab ini menyajikan hasil implementasi dan eksperimen penelitian. Pembahasan mencakup lingkungan dan perangkat keras yang digunakan, skenario pelatihan model TFT dan LLM-TFT, hasil pelatihan dan pemilihan skenario terbaik, *setting* data *testing* dan definisi sampel evaluasi, hasil evaluasi global, analisis kinerja berdasarkan horizon dan emiten, interpretabilitas model melalui pola *attention*, studi kasus visual prediksi, serta pembahasan kontribusi fitur sentimen pada model LLM-TFT.

BAB V KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan yang disusun berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, serta saran untuk pengembangan penelitian selanjutnya, termasuk pengayaan sumber berita, penyempurnaan pembentukan fitur sentimen, dan perluasan pengujian model maupun horizon prediksi.