BAB I
PENDAHULUAN

1.1 Latar Belakang

Harga saham berubah setiap hari dan sulit ditebak arahnya, sehingga prediksinya menjadi kajian yang banyak diminati dalam bidang keuangan maupun kecerdasan buatan [1]. Kemampuan memperkirakan pergerakan harga berguna untuk memahami perilaku pasar modal sekaligus menjadi bahan pengujian metode peramalan [2]. Penelitian ini memilih saham Alphabet Inc. (GOOGL) karena rekam jejak perdagangannya panjang dan datanya tersedia lengkap selama lebih dari dua dekade, sehingga menyediakan contoh yang cukup banyak bagi model untuk belajar [3]. Saham ini juga sangat likuid dan diperdagangkan di bursa Amerika Serikat dengan data terbuka untuk umum, sehingga hasilnya mudah ditelusuri ulang oleh peneliti lain. Saham dalam negeri tidak dipilih karena penelitian ini merupakan perbandingan arsitektur, bukan kajian atas satu pasar, sehingga objeknya perlu berupa saham yang lazim dipakai penelitian sejenis agar hasilnya dapat disandingkan dengan temuan terdahulu. Kerangka perbandingan yang diuji tetap dapat diterapkan pada saham mana pun, termasuk saham di Bursa Efek Indonesia.

Berbagai pendekatan telah dipakai untuk memprediksi harga saham. Pendekatan statistik klasik seperti ARIMA dan *exponential smoothing* sudah lama digunakan karena perhitungannya sederhana dan hasilnya mudah dijelaskan [4], tetapi pendekatan ini menganggap hubungan antardata bersifat lurus dan polanya cenderung tetap, padahal harga saham kerap berubah arah secara tiba-tiba [5]. Pendekatan pembelajaran mendalam (*deep learning*) kemudian banyak dipilih karena mampu mengenali pola yang rumit tanpa perlu ditentukan bentuk hubungannya terlebih dahulu [6]. Kelemahannya, model jenis ini menuntut data yang banyak, waktu pelatihan yang lebih lama, dan hasilnya lebih sulit ditelusuri [7].

Ketiga arsitektur CNN, LSTM, dan GRU telah banyak dibandingkan maupun digabungkan pada penelitian terdahulu, namun kesimpulannya belum seragam. Rezaei dan rekan pada 2024 memadukan pembelajaran mendalam dengan dekomposisi frekuensi untuk memprediksi harga saham [8], sedangkan Selvin dan rekan menguji LSTM, RNN, dan CNN memakai skema jendela geser (*sliding window*) [9]. Setiawan dan Tjahjadi pada 2024 membandingkan beberapa arsitektur pembelajaran mendalam untuk peramalan harga saham pada sektor teknologi di Asia Tenggara [10]. Perhatian tersebut berlanjut pada karya terbaru:

Bhanujyothi dan Jacob pada 2025 mengusulkan model hibrida CNN-LSTM berbasis mekanisme perhatian (*attention*) [11], dan Chohan dan rekan pada 2026 mengembangkan hibrida CNN-LSTM-GRU untuk peringatan dini risiko keuangan [12]. Karya tersebut menunjukkan ketiga arsitektur masih aktif diteliti, tetapi umumnya berupa model gabungan pada pasar dan konteks yang berbeda-beda.

Meskipun demikian, tinjauan sistematis atas peramalan deret waktu keuangan menunjukkan bahwa penelitian di bidang ini memakai pengaturan yang sangat beragam, mulai dari jumlah lapisan hingga cara pembagian data [7], dan keberagaman serupa juga tampak pada perbandingan yang lebih baru [10]. Akibatnya, selisih hasil yang dilaporkan dapat berasal dari perbedaan pengaturan, bukan dari perbedaan kemampuan arsitekturnya, padahal perbandingan model menuntut kondisi pengujian yang setara agar kesimpulannya sahih [13]. Sebagian penelitian juga masih membagi data secara acak tanpa menjaga urutan waktu, sehingga berisiko terjadi kebocoran data (*data leakage*) [14]. Perbandingan arsitektur tunggal yang setara pada satu saham likuid dengan validasi yang menjaga urutan waktu belum banyak dikaji, sehingga pertanyaan mengenai arsitektur mana yang terbaik masih terbuka [1]. Menjawab pertanyaan itulah yang menjadi urgensi penelitian ini. Oleh karena itu, penelitian ini diarahkan untuk menjawab empat pertanyaan pokok, yaitu arsitektur mana yang terbaik pada pengaturan yang sama, apakah skema validasinya stabil dan bebas kebocoran data, arsitektur mana yang terpilih beserta pola kesalahannya, dan seperti apa proyeksi harganya. Keempat pertanyaan inilah yang dirumuskan pada Subbab 1.2.

Hasil penelitian ini ditujukan bagi pembaca akademik, yaitu peneliti dan mahasiswa yang menekuni peramalan deret waktu keuangan, bukan bagi investor di dalam maupun di luar negeri. Yang ditawarkan adalah kerangka pengujian yang setara beserta gambaran apa adanya mengenai batas kemampuan ketiga arsitektur, sehingga temuannya dipakai untuk menilai metode dan bukan untuk mengambil keputusan jual-beli.

Penelitian ini menawarkan perbandingan yang setara antara CNN, LSTM, dan GRU pada saham GOOGL. Ketiga model dilatih dengan pengaturan yang disamakan dan dinilai memakai 10-Fold *Time Series Split*, yaitu skema pengujian bertahap yang selalu menempatkan data uji sesudah data latih sehingga urutan waktu tetap terjaga dan informasi masa depan tidak bocor ke dalam pelatihan [14]. Penilaian memakai beberapa ukuran kesalahan agar kesimpulan tidak bergantung pada satu ukuran saja [15]. Berdasarkan uraian tersebut, penelitian ini mengambil judul "Prediksi Harga Saham Alphabet Inc. (GOOGL) Menggunakan Algoritma CNN, LSTM, dan GRU Berbasis K-Fold Cross Validation".

1.2 Rumusan Masalah

Permasalahan yang dirumuskan dalam penelitian ini adalah sebagai berikut:

1. Di antara CNN, LSTM, dan GRU, arsitektur manakah yang memberikan kinerja prediksi terbaik pada saham GOOGL apabila dinilai menggunakan empat ukuran evaluasi, yaitu MAE, R^2, *Directional Accuracy* (DA), dan MASE?
2. Apakah skema 10-Fold *TimeSeriesSplit* menghasilkan estimasi kinerja yang stabil dan bebas dari kebocoran data bagi ketiga arsitektur tersebut?
3. Arsitektur manakah yang terpilih sebagai model terbaik, dan bagaimana pola kesalahan prediksi yang dihasilkan oleh model tersebut?
4. Bagaimana proyeksi harga saham GOOGL pada periode Mei hingga Juni 2026 yang dihasilkan oleh model terbaik tersebut?

1.3 Tujuan

Penelitian ini memiliki empat tujuan utama, yaitu:

1. Mengidentifikasi algoritma yang memberikan tingkat akurasi prediksi tertinggi di antara CNN, LSTM, dan GRU pada data pelatihan saham GOOGL (5.472 baris perdagangan harian dari 19 Agustus 2004 hingga 19 Mei 2026), berdasarkan empat metrik evaluasi (MAE, R^2, DA, MASE) yang diperoleh dari 10-Fold *Time Series Cross Validation* dalam satuan USD.
2. Mengukur konsistensi dan kestabilan performa ketiga algoritma menggunakan skema 10-Fold *TimeSeriesSplit* yang bebas *data leakage* pada 5.472 baris data pelatihan GOOGL melalui 30 siklus pelatihan (3 model × 10 fold).
3. Menentukan model terbaik dengan cara memberi peringkat pada keempat ukuran evaluasi, lalu memilih model yang peringkat gabungannya paling baik. Apabila terdapat model yang peringkatnya sama, pemilihan dilanjutkan dengan membandingkan ukuran berikutnya secara berurutan. Penelitian ini juga menelaah pola kesalahan yang muncul ketika model diterapkan pada periode dengan kenaikan harga yang tajam pada April hingga Mei 2026.
4. Membangun proyeksi harga GOOGL untuk 28 hari perdagangan (20 Mei hingga 30 Juni 2026) menggunakan model terbaik dari hasil 10-Fold *Cross Validation* sebagai luaran utama penelitian

1.4 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut:

1. Menyediakan kerangka perbandingan yang setara antararsitektur CNN, LSTM, dan GRU, yaitu ketiga model diuji pada objek yang sama dengan pengaturan pelatihan yang disamakan dan dinilai memakai kerangka validasi yang menjaga urutan waktu, sehingga perbedaan hasil dapat ditelusuri pada arsitekturnya dan bukan pada perbedaan pengaturan. Penelitian ini juga menunjukkan keterbatasan ketiga model apabila dibandingkan dengan peramalan dasar sebagai pembanding dasar, sehingga diperoleh gambaran yang jujur mengenai sejauh mana model pembelajaran mendalam benar-benar memberikan peningkatan.
2. Dapat mengetahui algoritma mana di antara CNN, LSTM, dan GRU yang menghasilkan akurasi prediksi terbaik pada data saham GOOGL berdasarkan hasil 10-Fold *Cross Validation*.
3. Dapat melengkapi literatur mengenai *machine learning* dalam bidang prediksi harga saham.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini meliputi hal-hal sebagai berikut:

1. Objek penelitian ini adalah saham Alphabet Inc. Kelas A (kode GOOGL), yaitu perusahaan induk Google yang sahamnya diperdagangkan di bursa Amerika Serikat dan menjadi salah satu komponen utama indeks S&P 500 serta NASDAQ-100. Saham ini dipilih karena memiliki rekam jejak perdagangan yang panjang dan pergerakan harga yang aktif, sehingga cocok untuk menguji kemampuan model prediksi.
2. Data yang digunakan adalah data historis harga saham GOOGL yang diunduh dari Yahoo Finance dalam bentuk berkas CSV sebanyak 5.472 baris perdagangan, mencakup periode 19 Agustus 2004 hingga 19 Mei 2026, dengan tujuh kolom (Tanggal, *Open*, *High*, *Low*, *Close*, *Volume*, dan *Adj Close*). Harga penutupan yang telah disesuaikan (*Adj Close*) dipilih sebagai nilai yang diprediksi karena sudah dikoreksi terhadap aksi korporasi seperti pemecahan saham dan pembagian dividen.
3. Tahap pra-pemrosesan data mencakup transformasi seluruh fitur harga menjadi *log-return*, penyetaraan skala antarfitur menggunakan normalisasi *Min-Max*, dan pembentukan jendela geser (*sliding window*) sepanjang 15 hari. Transformasi *log-return* dipilih karena harga GOOGL tidak stabil secara statistik dan tumbuh ratusan kali lipat sepanjang periode data, sehingga deret perlu distabilkan agar model tidak salah belajar dari tren tingkat harga. Normalisasi *Min-Max* dipilih karena keenam fitur memiliki satuan dan besaran yang sangat berbeda, yaitu volume bernilai jutaan sedangkan harga

hanya puluhan hingga ratusan dolar, sehingga penyetaraan skala mencegah fitur berskala besar mendominasi pembelajaran. Parameter normalisasi di-fit hanya pada data latih setiap *fold* untuk mencegah kebocoran data (*data leakage*). Konsekuensinya, nilai pada data uji dapat sedikit berada di luar rentang [0, 1] apabila melampaui rentang nilai data latihnya. Panjang jendela geser ditetapkan melalui percobaan yang diuraikan pada BAB III.

4. Model yang dibandingkan dibatasi pada tiga algoritma *deep learning*, yaitu CNN, LSTM, dan GRU. Ketiga model menggunakan enam fitur masukan (Open, High, Low, Close, Volume, dan Adj Close) serta diuji dengan pengaturan pelatihan yang sama agar perbandingannya adil, sedangkan rincian setiap pengaturannya diuraikan pada BAB III.
5. Pembagian dan pengujian data menggunakan metode 10-Fold *Time Series Split* sehingga model selalu diuji pada data yang lebih baru daripada data latihnya dan tidak terjadi kebocoran informasi masa depan.
6. Kinerja model diukur dengan empat metrik yang saling melengkapi, yaitu MAE, R^2, *Directional Accuracy*, dan MASE, untuk menilai ketepatan prediksi dari berbagai sisi.
7. Proyeksi harga dilakukan untuk 28 hari perdagangan ke depan (20 Mei hingga 30 Juni 2026) menggunakan model terbaik dengan skema proyeksi rekursif (*recursive forecasting*) tanpa pemasukan data harga sebenarnya, yaitu hasil prediksi satu hari dijadikan masukan untuk memprediksi hari berikutnya secara berantai. Pada setiap langkah proyeksi, keenam fitur diprediksi dan diperbarui sekaligus oleh model, sehingga tidak ada fitur yang dibekukan pada nilai hari terakhir data latih. Skema rekursif murni dipilih karena mencerminkan kondisi peramalan nyata ketika data masa depan benar-benar belum tersedia, sehingga model hanya boleh mengandalkan hasil prediksinya sendiri. Memasukkan data harga sebenarnya sebagai masukan akan memberikan gambaran kinerja yang terlalu optimistis dan tidak realistis untuk masa proyeksi ke depan. Jumlah hari mengikuti kalender resmi bursa New York sehingga hari libur bursa tidak dihitung sebagai hari perdagangan. Lingkup analisis bersifat kuantitatif dan tidak mencakup pemberian rekomendasi jual-beli atau keputusan investasi.
8. Penelitian ini dijalankan pada lingkungan Google Colab menggunakan bahasa pemrograman Python dengan pustaka TensorFlow versi 2.20.0, NumPy versi 2.0.2, dan pandas versi 2.2.2.
9. Seluruh percobaan dirancang agar dapat diulang dengan hasil yang konsisten melalui penetapan nilai acak (*seed*) sebesar 42 pada pustaka yang digunakan.

10. Pengujian *overfitting* dilakukan secara terstruktur mengikuti kerangka uji hipotesis dan *effect size* melalui empat uji yang saling melengkapi (Shapiro-Wilk, *paired t-test*, Wilcoxon signed-rank, dan Cohen's d) atas selisih *error* MAE validasi dan latih pada skala log-return di seluruh *fold*. Keputusan akhir mengenai keparahan *overfitting* bertumpu pada ukuran besar pengaruh (*effect size*) Cohen's d, bukan semata pada nilai p (*p-value*). Nilai p adalah peluang memperoleh selisih *error* seekstrem hasil pengamatan apabila sebenarnya tidak ada perbedaan antara *error* data latih dan validasi. Nilai p yang kecil (lazimnya di bawah 0,05) menandakan perbedaan yang signifikan secara statistik. Cohen's d dipilih sebagai indikator utama karena nilai p sangat dipengaruhi oleh ukuran sampel sehingga pada jumlah *fold* yang relatif kecil (sepuluh) daya ujinya terbatas dan rawan gagal mendeteksi perbedaan yang nyata, sedangkan Cohen's d mengukur besarnya perbedaan secara relatif dan tidak bergantung pada ukuran sampel sehingga lebih sahih untuk menilai keparahan *overfitting*.