

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saham adalah sebuah tanda penyertaan modal seseorang atau instansi dalam sebuah perusahaan atau Perseroan Terbatas [1]. Melalui perdagangan saham, perusahaan yang menjual sahamnya bisa mendapatkan modal untuk melakukan operasi bisnis sehingga perusahaan bisa tetap berjalan dan terus berkembang untuk terus mendorong roda perekonomian [2]. Di sisi lain, investor juga bisa mendapatkan keuntungan jika harga saham yang dibeli mengalami kenaikan harga. Sebaliknya, jika harga saham mengalami penurunan, maka investor juga akan mengalami kerugian.

Ada banyak faktor yang mempengaruhi perubahan harga saham seperti kondisi fundamental perusahaan, proyeksi kinerja perusahaan di masa yang akan datang, kebijakan pemerintah, dan sentimen pasar [3]. Setiap faktor ini dapat berubah setiap saat sehingga sulit bagi investor untuk terus mengikuti masing-masing perubahan yang terjadi. Akibatnya, melakukan analisis dan membuat keputusan yang cepat dan tepat untuk berinvestasi menjadi tantangan tersendiri bagi investor. Salah satu solusi yang ditawarkan dengan adanya perkembangan teknologi adalah investor dapat melakukan analisis saham dengan bantuan *machine learning*. Dengan menggunakan *machine learning*, investor dapat menganalisis data saham yang besar secara lebih efisien dan akurat. Salah satu bentuk pemanfaatan *machine learning* adalah dengan memprediksi harga saham menggunakan harga-harga saham sebelumnya.

Berdasarkan sejumlah penelitian, *deep learning* menjadi metode *machine learning* dengan performa yang paling baik dibandingkan metode lainnya untuk memprediksi harga saham [2], [4]. *Deep learning* sendiri merupakan cabang dari *machine learning* yang menggunakan beberapa lapisan *neural network*. Di antara berbagai arsitektur *deep learning* yang tersedia, *Long-Short Term Memory* (LSTM)

merupakan arsitektur dengan performa paling superior jika dibandingkan dengan alternatif lainnya seperti *Recurrent Neural Network* (RNN) atau *Convolutional Neural Network* (CNN) [5], [6]. Tidak hanya itu, LSTM juga dapat dimodifikasi sehingga melahirkan varian-varian LSTM lainnya seperti *special* LSTM dan *Bidirectional*-LSTM. Selain itu, LSTM juga dapat digabungkan dengan arsitektur lainnya seperti *Convolutional Neural Network* atau CNN dan *Effective Channel Attention* atau ECA. Studi yang dilakukan sebelumnya menunjukkan bahwa modifikasi dari LSTM dapat memberikan performa yang lebih baik lagi [5], [6]. Modifikasi LSTM yang dimaksud adalah CNN-BiSLSTM dan CNN-BiLSTM-ECA.

Motivasi pemilihan model CNN-BiSLSTM dan CNN-BiLSTM-ECA dalam penelitian ini didasarkan pada keunggulan teknis yang relevan dengan karakteristik kompleks data harga saham. CNN-BiSLSTM menggabungkan kekuatan *Convolutional Neural Network* (CNN) dalam mengekstraksi fitur spasial dengan kemampuan *Bidirectional Special LSTM* (BiSLSTM) yang mampu mempelajari pola temporal dari dua arah sekaligus. Struktur khusus pada BiSLSTM ini meningkatkan kemampuan model dalam menangani data time series yang sangat bising dan *volatile*, seperti harga saham, sehingga diharapkan mampu memberikan prediksi yang lebih akurat. Di sisi lain, CNN-BiLSTM-ECA menambahkan modul *Effective Channel Attention* (ECA) yang berfungsi menyoroti fitur-fitur paling relevan di antara banyak channel input tanpa meningkatkan kompleksitas model secara signifikan. Hal ini penting karena data saham biasanya memiliki banyak fitur dengan kontribusi yang bervariasi terhadap prediksi harga. Tabel 1.1 membandingkan kelebihan dan kekurangan masing-masing model.

Tabel 1. 1 Perbandingan Kelebihan dan Kekurangan Model LSTM, CNN-BiSLSTM, dan CNN-BiLSTM-ECA

Model	Kelebihan	Kekurangan
-------	-----------	------------

LSTM	Dapat menangani data sekuensial jangka panjang dan mengatasi <i>vanishing gradient</i>	Tidak menangkap konteks dua arah secara optimal.[6]
CNN-BiSLSTM	Memiliki kemampuan ekstraksi fitur spasial dari CNN dan menangkap konteks dua arah serta pola jangka panjang dari BiSLSTM.	Rentan terhadap <i>overfitting</i> tanpa regularisasi yang tepat. [6]
CNN-BiLSTM-ECA	Modul ECA mampu menyoroti <i>channel</i> penting tanpa menambah kompleksitas signifikan, sehingga memperkuat performa prediksi.	Performa sangat bergantung pada pemilihan hyperparameter dan bisa tidak stabil jika tidak optimal. [5]

Selain itu, kedua model ini sebelumnya diuji pada dataset saham di pasar yang berbeda, sehingga belum ada perbandingan langsung yang *valid* antara keduanya. Studi yang mengusulkan model CNN-BiSLSTM menggunakan objek penelitian berupa Shenzhen Component Index serta studi yang mengusulkan model CNN-BiLSTM-ECA menggunakan objek penelitian berupa Shanghai Composite Index, China Unicom, dan CSI 300 [5], [6]. Penelitian ini bertujuan untuk membandingkan kinerja tiga model *deep learning*, yaitu LSTM sebagai tolak ukur serta CNN-BiSLSTM dan CNN-BiLSTM-ECA, yang sebelumnya telah dikembangkan dalam dua studi terpisah. Masing-masing studi mengklaim bahwa model mereka memberikan hasil terbaik dalam konteks prediksi harga saham. CNN-BiSLSTM menunjukkan performa yang sangat baik pada dataset saham tertentu, sementara CNN-BiLSTM-ECA unggul pada dataset yang berbeda. Belum ada penelitian yang membandingkan kedua model ini secara langsung

menggunakan dataset yang identik. Oleh karena itu, penelitian ini dirancang untuk melatih dan menguji ketiga model tersebut menggunakan dataset yang sama. Penelitian ini akan menggunakan 2 kelompok dataset, yaitu indeks saham dan saham individual. Data indeks saham yang akan digunakan adalah data harga S&P 500, Dow Jones Industrial, dan NASDAQ Composite. Ketiga indeks ini dipilih karena mereka adalah indeks saham yang paling populer saat penelitian ini ditulis [8]. Kemudian untuk kelompok saham individual, digunakan data harga saham AAPL (Apple Inc.), MSFT (Microsoft Corporation), dan GOOG (Alphabet Inc.). Ketiga saham ini dipilih karena per tanggal 7 November 2023 pukul 11.50 WIB mereka memiliki kapitalisasi pasar terbesar di Amerika Serikat [7].

Largest companies by market cap - US Stock Market												
Market capitalization shows the value of a corporation by multiplying the stock price by the number of stocks outstanding. Here you can see the companies with the largest market cap. Large-cap stocks are usually industry and sector leaders and represent well-known, established companies.												
Overview Performance Valuation Dividends Profitability Income Statement Balance Sheet Cash Flow Technicals												
Symbol			Market cap	Price	Change %	Volume	Relative Volume	P/E	EPS diluted TTM	EPS diluted growth % TTM YoY	Dividend yield % TTM	Sector
AAPL	Apple Inc.		2.844T usd	182.89 usd	0.59%	49.339M	0.78	29.84	6.13 usd	0.45%	0.52%	Electronic Technology
MSFT	Microsoft Corporation		2.699T usd	363.20 usd	0.74%	26.767M	0.92	35.17	10.33 usd	11.28%	0.75%	Technology Services
GOOG	Alphabet Inc.		1.658T usd	133.26 usd	0.65%	15.093M	0.54	25.55	5.22 usd	3.56%	0.00%	Technology Services
AMZN	Amazon.com, Inc.		1.468T usd	142.08 usd	-0.44%	44.522M	0.65	74.19	1.92 usd	75.87%	0.00%	Retail Trade
NVDA	NVIDIA Corporation		1.15T usd	465.74 usd	1.35%	34.671M	0.81	112.51	4.14 usd	35.76%	0.03%	Electronic Technology
META	Meta Platforms, Inc.		821.791B usd	319.78 usd	0.30%	13.61M	0.50	28.23	11.33 usd	7.92%	0.00%	Technology Services
BRK.A	Berkshire Hathaway Inc.		757.002B usd	526660.00 usd	0.26%	7.252K	0.92	8.82	59732.00 usd	694.31%	0.00%	Finance
TSLA	Tesla, Inc.		706.07B usd	222.11 usd	-0.03%	106.585M	0.91	71.52	3.11 usd	-4.05%	0.00%	Consumer Durables
LLY	Eli Lilly and Company		587.745B usd	619.13 usd	3.20%	7.776M	2.42	114.28	5.42 usd	-18.63%	0.73%	Health Technology
V	Visa Inc.		496.695B usd	243.91 usd	-0.35%	5.118M	0.88	29.78	8.19 usd	19.59%	0.74%	Commercial Services

Gambar 1. 1 Perusahaan - Perusahaan dengan Kapitalisasi Pasar Terbesar di Amerika Serikat [7]

Gambar 1.1 menjelaskan statistik dari saham-saham perusahaan dengan kapitalisasi pasar terbesar di Amerika Serikat. Berdasarkan gambar tersebut dapat dilihat pada kolom *Market cap* bahwa AAPL bernilai 2.844 triliun dolar AS, MSFT bernilai 2.699 triliun dolar AS, dan GOOG bernilai 1.658 triliun dolar AS. Data indeks saham juga akan digunakan pada penelitian kali ini untuk melihat bagaimana performa algoritma ini dalam memprediksi data yang relatif lebih stabil.

Selain dataset saham yang digunakan sama, model LSTM konvensional juga akan diterapkan sebagai model pembandingan untuk memberikan gambaran

yang lebih menyeluruh mengenai keefektifan masing-masing model. Pendekatan ini diharapkan dapat mengevaluasi kembali apakah modifikasi yang diterapkan pada model CNN-BiSLSTM dan CNN-BiLSTM-ECA benar membawa peningkatan performa terutama tugas dalam memprediksi harga saham AS.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah ditulis, maka dirumuskan permasalahan berikut.

1. Bagaimana hasil evaluasi dari algoritma LSTM dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite?
2. Bagaimana hasil evaluasi dari algoritma CNN-BiSLSTM dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite?
3. Bagaimana hasil evaluasi dari algoritma CNN-BiLSTM-ECA dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite?
4. Manakah model terbaik antara LSTM, CNN-BiSLSTM, dan CNN-BiLSTM-ECA untuk memprediksi saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite berdasarkan nilai RMSE, MSE, MAPE, dan MAE masing-masing model?

1.3 Batasan Masalah

Berikut adalah batasan masalah yang ditentukan oleh penulis untuk mencapai tujuan analisis.

1. Penelitian ini berfokus pada prediksi saham perusahaan AAPL, MSFT, dan GOOG serta indeks saham S&P 500, Dow Jones Industrial, dan NASDAQ Composite.
2. Data yang digunakan dimulai sejak awal saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite terdaftar di bursa sampai 24 Januari 2025 dengan interval waktu harian.

3. Sumber data diambil dari laman Yahoo Finance.
4. Algoritma yang akan digunakan pada penelitian ini adalah CNN-BiSLSTM, CNN-BiLSTM-ECA, dan LSTM sebagai tolak ukur.
5. Tidak sampai tahap *deployment*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang tertulis di atas, berikut adalah tujuan dari penelitian ini.

1. Mengetahui hasil evaluasi dari algoritma LSTM dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.
2. Mengetahui hasil evaluasi dari algoritma CNN-BiSLSTM dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.
3. Mengetahui hasil evaluasi dari algoritma CNN-BiLSTM-ECA dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.
4. Menentukan algoritma terbaik untuk memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.

1.4.2 Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini.

1. Mengetahui penerapan dari algoritma CNN-BiSLSTM dan CNN-BiLSTM-ECA dalam memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.
2. Memberikan rekomendasi algoritma terbaik antara LSTM, CNN-BiSLSTM dan CNN-BiLSTM-ECA untuk memprediksi harga saham AAPL, MSFT, GOOG, S&P 500, Dow Jones Industrial, dan NASDAQ Composite.

1.5 Sistematika Penulisan

Berikut ini merupakan sistematika penulisan yang digunakan dalam penulisan Skripsi ini.

BAB 1 PENDAHULUAN

Bagian ini berisi latar belakang yang membahas tentang saham, *deep learning*, algoritma yang akan digunakan, dan juga saham-saham yang akan digunakan. Dalam bab ini juga dijelaskan mengenai rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika dari penelitian.

BAB II LANDASAN TEORI

Bagian ini berisi tentang pembahasan dasar teori yang digunakan penelitian ini. Dasar teori yang dibahas pada bagian ini adalah saham, CNN, LSTM, ECA, RMSE, MSE, MAE, MAPE, Python, Microsot Visual Studio Code, dan penelitian terdahulu.

BAB III METODOLOGI PENELITIAN

Bagian ini berisi penjelasan mengenai alur penelitian yang digunakan untuk mengumpulkan data dan mengolah data menggunakan metode yang diperlukan.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bagian ini berisi tentang proses implementasi pemodelan menggunakan algoritma LSTM, CNN-BiSLSTM, CNN-BiLSTM-ECA yang kemudian dilanjutkan dengan evaluasi dari masing-masing algoritma.

BAB V SIMPULAN DAN SARAN

Bagian ini berisikan mengenai simpulan atau hasil dari penelitian yang merujuk kepada tujuan penelitian. Selain itu bagian ini juga berisi saran untuk penelitian selanjutnya.