

BAB II

TINJAUAN PUSTAKA

Prediksi harga saham tradisional umumnya mengandalkan model statistik seperti ARIMA, yang memecah deret waktu menjadi komponen autoregresif, integrasi (differencing), dan *moving average* untuk menangkap tren serta musiman dengan cukup baik [1]. Meski populer, ARIMA bergantung pada asumsi linearitas dan kestabilan pola yang sering dilanggar oleh data pasar saham yang sangat fluktuatif, sehingga performanya dapat menurun drastis ketika menghadapi lonjakan ekstrem atau pola non-linear [2].

Untuk mengatasi keterbatasan model tradisional, para peneliti telah mengembangkan berbagai pendekatan. Model seperti Regresi Linier [15] dan *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) [16] efektif pada data yang stabil namun masih kesulitan menangkap pola non-linear yang kompleks. Pendekatan modern seperti *Gated Recurrent Units* (GRU) [17] dan *Transformer* [18] menawarkan akurasi tinggi tetapi membutuhkan sumber daya komputasi besar dan kompleksitas implementasi yang signifikan.

Di antara berbagai alternatif, LSTM muncul sebagai solusi seimbang antara kemampuan menangkap pola kompleks dan efisiensi implementasi. Sebagai varian *Recurrent Neural Network* (RNN), LSTM memperkenalkan tiga mekanisme gate (forget, input, output) dan cell state yang mengatur aliran informasi sehingga model dapat mempertahankan memori jangka panjang tanpa terjebak pada masalah vanishing atau exploding gradient [2]. Studi dalam [3] melaporkan bahwa LSTM mampu menurunkan MSE hingga 15-20% dalam meramalkan indeks S&P 500 dibandingkan ARIMA. Penelitian dalam [4] juga menunjukkan bahwa kombinasi *stacked autoencoder* untuk ekstraksi fitur hierarkis sebelum pemrosesan oleh LSTM dapat meningkatkan akurasi hingga sekitar 12% pada berbagai saham global. Lebih lanjut, penelitian pada [5] membandingkan model LSTM dan ARIMA murni dalam konteks deret waktu ekonomi, dan hasilnya menunjukkan

bahwa LSTM secara konsisten mengungguli ARIMA, khususnya pada data non-linear dan pola yang tidak stabil.

Pada perkembangan terkini menunjukkan bahwa integrasi variabel sentimen pasar semakin dianggap penting dalam meningkatkan akurasi model prediksi harga saham. Studi dalam [6] membuktikan bahwa *mood* Twitter, yang diukur menggunakan *Google Profile of Mood States*, memiliki korelasi signifikan terhadap pergerakan indeks DJIA dengan tingkat akurasi hingga 87%. Penelitian lainnya, seperti [8], menemukan bahwa informasi dari *microblogging platform* seperti Twitter dapat meningkatkan performa model prediktif berbasis deret waktu ketika digabungkan dengan data sentimen sebagai variabel tambahan. Penelitian dalam [7] memanfaatkan skor sentimen Twitter sebagai masukan untuk model *Extreme Gradient Boosting* (XGBoost) dan melaporkan peningkatan nilai R^2 dari 0,72 menjadi 0,78. Studi lain pada [8] juga menggabungkan sentimen yang diukur dengan *VADER* bersama indikator teknikal ke dalam ARIMA, menghasilkan penurunan MAPE hingga 12%. Sementara itu, penelitian dalam [19] menunjukkan bahwa integrasi data sentimen berita (melalui *VADER*) dengan indikator teknikal dalam kerangka kerja *Apache Spark* dapat meningkatkan akurasi prediksi tren saham secara signifikan. Hasil serupa diperoleh dalam [10], di mana peneliti menambahkan skor sentimen langsung ke dalam model LSTM, baik secara tunggal maupun *ensemble*, dan menemukan penurunan rata-rata MAPE sekitar 5% dibanding model tanpa sentimen.

Selain penelitian-penelitian tersebut, terdapat beberapa studi terbaru yang memperkuat relevansi integrasi LSTM, ARIMA, dan sentimen pasar. Studi dalam [20] mengusulkan model berbasis LSTM yang menggabungkan data historis harga dan analisis sentimen untuk meningkatkan akurasi prediksi saham. Penelitian pada [21] membandingkan performa ARIMA dan LSTM pada saham perusahaan *FinTech*, dan hasilnya menunjukkan bahwa LSTM memberikan tingkat akurasi yang lebih tinggi dalam menangani data non-linear serta volatilitas tinggi. Sementara itu, penelitian dalam [22] menemukan bahwa model LSTM memberikan hasil prediksi yang lebih stabil dibandingkan ARIMA pada saham perusahaan

teknologi, terutama pada fase perubahan tren harga yang tajam.



Tabel 2.1. Tabel Perbandingan Penelitian

Peneliti (Tahun)	Judul	Model	Data Sentimen	Periode Studi	Temuan Kunci	Gap Terhadap Penelitian Ini
Bollen et al. (2011)	Twitter mood predicts the stock market	Analisis Korelasi	Twitter	2008–2009 (DJIA)	Korelasi mood–DJIA (87%)	Tidak membandingkan model prediksi
Siarni-Namini et al. (2019)	A Comparison of ARIMA and LSTM in Forecasting Time Series	ARIMA vs LSTM	Tidak	Berbagai dataset ekonomi (1913–2016)	LSTM unggul ARIMA pada data non-linear	Tidak sertakan sentimen
Mittal & Goel (2017)	Stock Prediction Using Twitter Sentiment Analysis	XGBoost	Twitter	2015–2016 (AAPL)	R ² meningkat 8%	Tidak uji periode fluktuasi tinggi
Fischer & Krauss (2018)	Deep learning with long short-term memory networks for financial market predictions	LSTM	Tidak	2006–2016 (S&P 500)	LSTM unggul 20% vs ARIMA	Tidak sertakan sentimen
Oliveira et al. (2017)	The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices	ARIMA + Sentimen	Twitter	2013–2015 (S&P 500)	MAPE turun 12%	Tidak uji LSTM
Patel et al. (2015)	Predicting stock and stock price index movement using	LSTM + Sentimen	Berita finansial	2004–2013 (Indeks Saham)	MAPE turun 5%	Tidak bandingkan dengan ARIMA

	Trend Deterministic Data Preparation and machine learning techniques					
Gupta et al. (2022)	HiSA-SMFM: Historical and Sentiment Analysis Based Stock Market Forecasting Model	LSTM + Sentimen	Twitter	2019–2021 (berbagai saham)	Integrasi data historis dan sentimen meningkatkan akurasi prediksi	Belum bandingkan dengan ARIMA
Zhang & Tong (2023)	Comparing the ARIMA and LSTM Models on the Stock Price of FinTech Companies	ARIMA vs LSTM	Tidak	2018–2023 (Perusahaan FinTech)	LSTM lebih efektif dalam pola non-linear	Tidak gunakan sentimen
Kirelli (2024)	Comparative Analysis of LSTM and ARIMA Models in Stock Price Prediction: A Technology Company Example	ARIMA vs LSTM	Tidak	2019–2024 (Perusahaan Teknologi)	LSTM lebih stabil pada volatilitas tinggi	Tidak gunakan data sentimen
Penelitian Ini	Prediksi Harga Saham Menggunakan Model Long Short-Term Memory	ARIMA vs LSTM	Twitter	2017–2024 (AAPL)	Pengaruh sentimen di 3 periode	Lengkap

Berdasarkan Tabel 2.1 di atas, terlihat bahwa meskipun berbagai kemajuan telah dicapai, masih terdapat gap penelitian yang perlu diisi: evaluasi dampak sentimen terhadap performa komparatif LSTM dengan *baseline model* ARIMA. Studi yang menguji kombinasi fitur teknikal dengan sentimen pada periode fluktuasi ekstrem seperti COVID-19 juga masih terbatas. Penelitian ini secara khusus bertujuan mengisi celah tersebut melalui komparasi langsung model LSTM dengan *baseline model* ARIMA berbasis sentimen dengan fitur teknikal pada tiga periode pasar berbeda.

