

Prediksi Return Saham Berdasarkan Data Histori dan Data Fundamental Menggunakan LSTM

Naufal Abdurrahman Burhani
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
zeknous@telkomuniversity.ac.id

Deni Saepudin
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
denisaepudin@telkomuniversity.ac.id

Didit Adytia
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
adytia@telkomuniversity.ac.id

Abstrak — Prediksi harga saham merupakan tantangan signifikan dalam dunia keuangan karena sifat pasar yang dinamis dan kompleks. Penelitian ini bertujuan untuk memprediksi harga saham dengan menggunakan model Long Short-Term Memory (LSTM), yang efektif dalam menangkap pola sekuensial pada data historis dan fundamental. LSTM dipilih karena kemampuannya untuk memodelkan ketergantungan jangka panjang, yang penting dalam prediksi harga saham. Evaluasi dilakukan dengan membandingkan kinerja model LSTM pada data historis dan kombinasi data historis dengan data fundamental. Hasil penelitian menunjukkan bahwa meskipun model LSTM berhasil menangkap pola pada data training, model tersebut mengalami overfitting yang signifikan pada data testing. Penambahan data fundamental tidak secara signifikan meningkatkan akurasi prediksi pada fase testing. Kesimpulannya, meskipun LSTM membantu dalam menangkap pola kompleks dari data historis dan fundamental, penelitian ini menekankan pentingnya eksplorasi lebih lanjut terhadap metode lain untuk mengatasi overfitting dan meningkatkan kinerja prediksi saham pada data testing.

Kata kunci— LSTM, Data Fundamental, Analisis Saham, Prediksi Pasar

I. PENDAHULUAN

Penelitian ini akan membahas perbedaan nilai akurasi ketika model prediksi LSTM digunakan dan data fundamental ditambahkan. Karena pemilihan saham yang tepat memiliki potensi untuk menghasilkan keuntungan besar, tetapi juga melibatkan risiko yang perlu dikelola dengan hati-hati dan bijaksana, masalah ini harus diperhatikan. Terdapat beberapa aspek yang perlu dipertimbangkan dalam tugas akhir ini, seperti analisis fundamental yang diisi dengan laporan dan rasio keuangan yang menggunakan parameter seperti Return on Equity, Return on Asset, Debt-to-Equity Ratio, Financial Leverage, Current Ratio, Quick Ratio, Receivable Turnover Ratio, Inventory Turnover Ratio, Net Margin, Income Margin, Operating Revenue Growth Rate dan Net Income Growth Rate. Untuk membedakan dari penelitian sebelumnya [1] yang menggunakan data fundamental yang sama, meskipun menggunakan metode dan tujuan yang berbeda, penelitian

saat ini menggunakan data yang sama dengan metode LSTM. Banyak penelitian yang telah membahas masalah ini, di antaranya dengan menggunakan metode Single-Layer Feedforward Network (SLFN), Extreme Learning Machine (ELM), Differential Evolution (DE) dan Genetic Algorithm (GA). Penggunaan SLFN digunakan sebagai sistem untuk meramalkan harga saham di masa depan, Penggunaan ELM digunakan untuk mengeksplorasi faktor utama yang mempengaruhi harga saham, Penggunaan DE digunakan untuk menentukan saham mana yang harus dipertimbangkan untuk investasi dan digunakan untuk memilih saham yang paling mungkin pada penelitian tahap kedua, Penggunaan GA digunakan untuk membangun portofolio yang paling diinginkan pada penelitian tahap ketiga. Hasilnya menunjukkan bahwa sistem yang diusulkan mempunyai potensi menjadi alternatif yang baik terhadap metode yang ada [1] dan pada penelitian sebelumnya menggunakan metode Artificial Bee Colony (ABC) Algorithm, Directed Artificial Bee Colony (DABC), Back-Propagation (BP) dan Artificial Neural Network (ANN) yang dimana ABC digunakan untuk mengidentifikasi solusi optimal berkualitas tinggi dan menawarkan keseimbangan antara kompleksitas dan kinerja, serta mengoptimalkan prediksi secara efektif. DABC digunakan untuk mengoptimalkan bobot dan bias ANN sebelumnya melatih jaringan dengan algoritma BP. Hasilnya menunjukkan bahwa metode tersebut belum bisa diandalkan untuk mengintegrasikan proses pelatihan yang tepat ke dalam Neural Network. Hasilnya memiliki kinerja prediktif yang memuaskan, namun masih memiliki beberapa kekurangan. Penelitian di masa depan cenderung untuk menentukan dampak kritis dari variabel-variabel analisis fundamental tertentu terhadap kualitas prediksi harga saham [2]. Dalam penelitian ini, masalah yang dibahas diselesaikan dengan menggunakan metode Long Short Term Memory (LSTM) dan Data Fundamental. Dengan menganalisis data historis saham, menemukan pola, dan membuat prediksi, alat-alat ini dapat digunakan untuk membuat Stock Prediction untuk investasi saham. LSTM digunakan karena menggunakan fungsi aktivasi non-linear, mereka memungkinkan untuk memodelkan hubungan yang kompleks dalam data. Untuk memahami pasar keuangan yang kompleks dan berubah-ubah, fleksibilitas ini sangat

penting. Selain itu, LSTM dirancang untuk mengatasi masalah hilangnya gradien, yang sering terjadi dalam RNN tradisional [4]. LSTM dapat menangkap dan mengingat ketergantungan jangka panjang dalam data karena arsitektur LSTM yang sangat penting ketika berurusan dengan harga saham yang mungkin dipengaruhi oleh tren historis. Selain itu, LSTM memiliki kemampuan untuk secara otomatis mempelajari fitur yang relevan dari data input, sehingga mengurangi kebutuhan untuk melakukan rekayasa fitur secara manual. Ini sangat bermanfaat dalam analisis data keuangan, yang mengidentifikasi fitur yang relevan mungkin sulit. Fokus utama dari penelitian ini adalah penerapan model LSTM dengan data fundamental untuk menawarkan pendekatan prediksi harga saham yang lebih luas. Penelitian ini membandingkan model LSTM yang menggunakan fitur PCA dan model tanpa fitur PCA, menunjukkan bagaimana rekayasa fitur mempengaruhi kinerja prediksi harga saham. Selain itu, penelitian ini menekankan pentingnya data fundamental untuk memperbaiki prediksi harga saham.

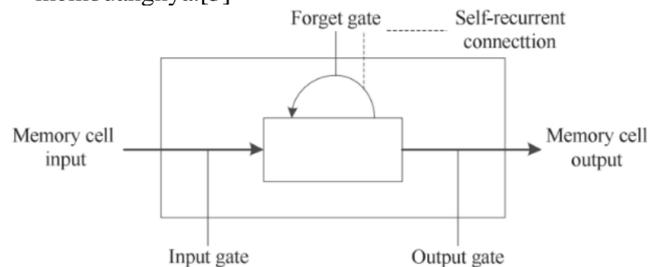
Penelitian ini mengambil topik pasar saham dengan cara memprediksi data dari tahun 2009-2023 yang didapatkan dari Yahoo Finance. Data yang digunakan adalah "Close" karena harga penutupan dapat diakses dan dicatat secara konsisten sepanjang hari perdagangan; itu juga mencerminkan harga perdagangan akhir pada akhir hari perdagangan, memberikan titik referensi yang konsisten. dan untuk Data Fundamental digunakan 12 fitur yaitu Return on Equity, Return on Asset, Debt-to-Equity Ratio, Financial Leverage, Current Ratio, Quick Ratio, Receivable Turnover Ratio, Inventory Turnover Ratio, Net Margin, Income Margin, Operating Revenue Growth Rate dan Net Income Growth Rate. Penggunaan dataset pada penelitian ini yaitu TLKM, BBCA dan ASII karena pilihan yang baik untuk digunakan dalam model LSTM karena harga penutupan menunjukkan informasi penting tentang pergerakan pasar, memiliki pola volatilitas yang dapat dipelajari, dan sering mencerminkan kondisi teknis dan fundamental perusahaan. Batasan dalam penelitian ini adalah data yang digunakan butuh dalam jumlah banyak untuk menghasilkan predikat yang akurat.

Penelitian ini bertujuan untuk memprediksi harga saham 1 bulan ke depan dengan cara mengimplementasikan LSTM pada data saham dengan menggunakan data fundamental. Penelitian ini juga menginvestigasi dampak penggunaan data fundamental dalam prediksi data. Diharapkan bahwa penelitian ini akan memberikan informasi bermanfaat tentang cara meningkatkan akurasi analisis prediksi dengan menambahkan data fundamental tersebut.

II. KAJIAN TEORI

A. LSTM

LSTM adalah evolusi dari Recurrent Neural Network (RNN) yang bertujuan untuk mengidentifikasi ketergantungan jangka panjang dalam data urutan. Berbeda dengan RNN, LSTM melakukan perhitungan status tersembunyi dengan cara yang lebih kompleks dengan menggantikan neuron lapisan tersembunyi tradisional dengan seperangkat sel memori. LSTM mampu mengatur informasi melalui tiga struktur gerbang, dan gerbang-gerbang ini memperbarui informasi secara selektif dengan belajar jenis informasi apa dalam urutan yang penting untuk disimpan atau mengidentifikasi informasi yang tidak berguna dan kemudian membuangnya.[3]



Gambar 1. LSTM Model Memory Cell

B. Data Fundamental

Konsep dasar yang mencakup pengumpulan, transformasi, pemrosesan, deskripsi, dan pemodelan data dikenal sebagai "data fundamental". Ini adalah komponen penting dalam ilmu data, yang mencakup berbagai teknik dan pendekatan untuk mendapatkan informasi bermanfaat dari data yang tidak terstruktur. Ilmu dan memainkan peran penting dalam penelitian ilmu komputer dan teknik, melibatkan pengumpulan, transformasi, pemrosesan, deskripsi, dan pemodelan data [9][10]. Tujuan menggunakan Data Fundamental adalah memberikan gambaran menyeluruh tentang kinerja dan nilai intrinsik perusahaan, membantu investor dan analis untuk membuat keputusan investasi yang lebih informasi.

C. Data Historis

Data historis adalah bagian dari analisis teknikal yang melibatkan grafik dan data pergerakan harga saham masa lampau, untuk mencari pola dan tren mengenai kondisi perusahaan dan pasar saham saat ini. Studi menunjukkan bahwa menggunakan data historis saham membantu investor membuat pilihan yang lebih baik. Prediksi harga saham menjadi lebih akurat berkat algoritma pembelajaran mesin dan pembelajaran mendalam[11], analisis data historis [12], dan fusion data [13]. Analisis pasar saham juga sangat dipengaruhi oleh sumber data baru dan teknologi canggih. Investor dapat menemukan peluang baru dan mengurangi risiko dengan strategi investasi berbasis data ini.

D. PCA

Tujuan penggunaan Principal Component Analysis (PCA) adalah teknik multivariat yang menganalisis tabel data di mana pengamatan dijelaskan oleh beberapa variabel dependen kuantitatif yang saling berkorelasi. Tujuannya adalah untuk mengekstrak informasi penting dari data, menampilkannya sebagai sekumpulan variabel ortogonal baru yang disebut komponen utama, dan menunjukkan pola kesamaan antara pengamatan

dan variabel di dalam peta. Kualitas model PCA dapat dievaluasi menggunakan teknik validasi silang seperti bootstrap dan jackknife. PCA dapat digeneralisasikan sebagai analisis korespondensi (Correspondence Analysis/CA) untuk menangani variabel kualitatif dan sebagai analisis faktor ganda (Multiple Factor Analysis/MFA) untuk menangani kumpulan variabel yang heterogen. Secara matematis, PCA bergantung pada dekomposisi eigen dari matriks semidefinit positif dan pada dekomposisi nilai singular (Singular Value Decomposition/SVD) dari matriks persegi panjang [8]. Cara kerja PCA yaitu membuat komponen baru yang menyimpan informasi paling berharga dari fitur-fitur dengan menangkap variansi yang tinggi [5].

E. Time Series Forecasting

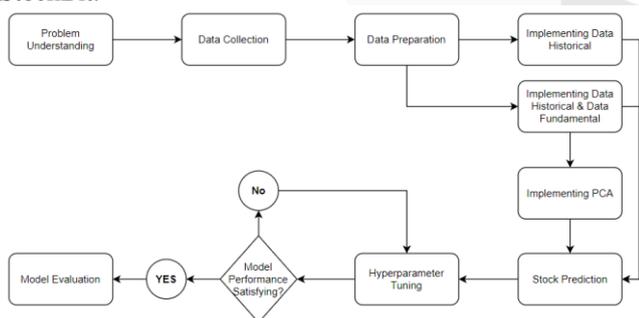
Time Series Forecasting adalah tugas memprediksi nilai masa depan dari suatu urutan tertentu menggunakan data historis. Para peneliti di bidang pembelajaran mesin baru-baru ini tertarik dengan tugas ini untuk mengatasi keterbatasan metode forecasting tradisional, yang memakan waktu dan penuh dengan kompleksitas. Teknik forecasting yang kuat yang dapat menentukan ketergantungan stokastik antara nilai masa lalu dan masa depan sangat dibutuhkan karena ketersediaan data historis yang semakin meningkat. [7].

F. Random Search

Random Search adalah salah satu metode yang paling banyak digunakan untuk Optimasi Hyperparameter, dan sangat penting bagi keberhasilan model deep learning. Meskipun kinerjanya yang mengagumkan, sedikit teori non-heuristik yang telah dikembangkan untuk menggambarkan mekanisme kerja dasarnya. Makalah ini memberikan penjelasan teoritis tentang Random Search. [6].

III. METODE

Dataset yang digunakan untuk penelitian ini adalah dari perusahaan ASII, BBCA dan TLKM, dimana column yang digunakan adalah "Close" dan jangka waktu 2009-2023. bersama dengan data fundamental yang diambilkan dari web StockBit.



Gambar 2. Diagram sistem kerja

Terdapat ada beberapa tahapan yang dilakukan dalam membangun sistem, seperti Problem Understanding, Data Collection, Data Preparation, Implementation model LSTM, Stock Prediction, Hyperparameter Tuning, Implementing Data Fundamental, Model Evaluation. Berikut diagram alir pembuatan sistem tersebut:

1) Problem Understanding difokuskan pada proses mendefinisikan masalah utama yang menjadi fokus penelitian dan menerapkan model. Dalam proses ini, definisi

masalah utama akan diikuti dengan fokus pada rumusan masalah penelitian.

2) Data Understanding adalah tahap kedua proses analisis data, analis atau ilmuwan data memperoleh pemahaman tentang karakteristik, struktur, dan isi dataset yang sedang mereka pelajari. Proses ini melibatkan eksplorasi data untuk memahami karakteristiknya, menemukan pola, menilai kualitas dan kelengkapan data, dan memahami hubungan antara variabel. Fase ini sangat penting untuk membentuk fondasi untuk pengolahan, pembersihan, dan analisis data selanjutnya, yang memungkinkan pengambilan keputusan yang akurat dan interpretasi data yang relevan. Pada tahap ini, sangat penting untuk menggunakan data saham "close" karena harga penutupan dapat diakses dan dicatat secara konsisten sepanjang hari perdagangan. Selain itu, data ini mencerminkan harga perdagangan akhir pada akhir hari perdagangan, memberikan titik referensi yang konsisten. Karena lebih stabil dan tidak terpengaruh oleh fluktuasi besar intraday, data "close" ini biasanya digunakan. Ini memberikan gambaran yang lebih akurat dan dapat diandalkan tentang bagaimana saham berjalan pada hari tertentu. Dengan memahami data "close", analis dapat menemukan pola pergerakan harga dan tren jangka panjang yang penting untuk prediksi dan analisis.

3) Data Preparation merupakan proses pengolahan data, Pada langkah ini akan dilakukan normalisasi data dengan menetapkan tanggal dan kolom yang akan digunakan misalnya menggunakan kolom "Close" pada data yang akan digunakan.

4) Implementing Data Historical digunakan dengan tujuan memperkirakan atau memprediksi pergerakan harga saham di pasar keuangan di masa depan. Cara implementasi data historis pada model prediksi yaitu:

1) Data Collection: Mengumpulkan data dari sumber terpercaya seperti Yahoo Finance

2) Data Preparation:

Normalization/Scaling: normalisasikan atau standarisasi data untuk memastikan bahwa berbagai data "close" berkontribusi secara proporsional pada model.

3) Data Integration:

Consistency Check: Memastikan semua kumpulan data disinkronkan dan disejajarkan dengan benar untuk menghindari ketidakcocokan waktu.

5) Implementing Data Fundamental digunakan untuk menambah data pada model prediksi dengan tujuan memperkirakan pergerakan harga saham di pasar keuangan. Identifikasi data fundamental yang Relevan: Mengumpulkan informasi finansial seperti Return on Equity, Return on Asset, Debt-to-Equity Ratio, Financial Leverage, Current Ratio, Quick Ratio, Receivable Turnover Ratio, Inventory Turnover Ratio, Net Margin, Income Margin, Operating Revenue Growth Rate dan Net Income Growth Rate. Cara implementasi data fundamental untuk model prediksi yaitu:

1) Data Collection: Mengumpulkan data dari sumber terpercaya seperti StockBit.

2) Data Cleaning dan Preparation:

Menangani Nilai yang Hilang: Memasukkan nilai yang hilang menggunakan metode yang tepat

Normalization/Scaling: Normalisasikan atau standarisasi data untuk memastikan bahwa berbagai fitur berkontribusi secara proporsional pada model.

3) Feature Engineering:

Lagged Data: Data fundamental seringkali mempunyai efek tertunda pada harga saham. Terapkan fitur-fitur yang tertinggal untuk mencerminkan penundaan ini.

4) Data Integration:

Combine Data Sources: menggabungkan data harga saham dengan data fundamental berdasarkan periode waktu (misalnya menyelaraskan data keuangan triwulanan dengan harga saham terkait).

Consistency Check: Memastikan semua kumpulan data disinkronkan dan disejajarkan dengan benar untuk menghindari ketidakcocokan waktu.

6) Implementing PCA Pada tahap ini, PCA akan memberikan jalan alternatif untuk penerapan data historis dan data fundamental. Ini akan memungkinkan pengumpulan data bernilai tinggi dari atribut yang memiliki kemampuan untuk menangkap varians tertinggi.

Table 1. Penangkapan komponen variansi

Cumulative Explained Variance	
Component 1	28.35%
Component 2	50.25%
Component 3	62.98%
Component 4	72.97%
Component 5	81.14%
Component 6	88.21%
Component 7	93.94%
Component 8	97.70%
Component 9	99.10%
Component 10	99.63%
Component 11	99.98%
Component 12	100.00%

Alasan menggunakan empat dan delapan fitur karena PCA dapat mempertahankan sekitar 73% variansi dengan empat fitur, yang berarti sebagian besar informasi penting dari data asli dapat diwakili oleh komponen-komponen ini. Sebaliknya, dengan delapan fitur, PCA dapat mempertahankan sekitar 98% variansi, yang berarti hampir semua informasi dari data asli terjaga. Pemilihan fitur 4 dan 8 dilakukan untuk mengukur sejauh mana jumlah komponen yang berbeda dapat mempengaruhi hasil prediksi dan untuk menentukan apakah peningkatan jumlah fitur yang mencakup lebih banyak variansi data dapat meningkatkan akurasi model prediksi atau justru meningkatkan risiko overfitting. Cara implementasi PCA pada model prediksi yaitu:

1) Data Preparation:

Normalization/Scaling: Normalisasikan atau standarisasi data untuk memastikan bahwa berbagai data "close" berkontribusi secara proporsional pada model.

2) Feature Engineering:

Lagged Data: Data fundamental seringkali mempunyai efek tertunda pada harga saham. Terapkan fitur-fitur yang tertinggal untuk mencerminkan penundaan ini.

3) Data Integration:

Combine Data Sources: menggabungkan data harga saham dengan data fundamental berdasarkan periode waktu (misalnya menyelaraskan data keuangan triwulanan dengan harga saham terkait).

Consistency Check: Memastikan semua kumpulan data disinkronkan dan disejajarkan dengan benar untuk menghindari ketidakcocokan waktu.

7) Stock Prediction adalah untuk mengetahui bagaimana harga saham dapat berubah di masa depan, sehingga pedagang dan investor dapat membuat keputusan yang lebih baik tentang pembelian, penjualan, atau kepemilikan saham. Salah satu metode yang sering digunakan untuk prediksi saham adalah LSTM, yang merupakan jenis jaringan saraf tiruan (neural network) yang khusus dirancang untuk mengenali pola dalam data urutan waktu. Berikut adalah langkah-langkah umum dalam implementasi LSTM untuk prediksi:

1) Data preparation:

Normalization/Scaling: Normalisasikan atau standarisasi data untuk memastikan bahwa berbagai data "close" berkontribusi secara proporsional pada model.

Data Splitting: Melakukan pembagian data untuk set pelatihan dan set testing. pada tahap ini telah membagikan dengan ratio 65:35 untuk pengujian model tersebut.

2) Pembangunan Model LSTM: Model LSTM dibangun menggunakan kerangka kerja machine learning seperti TensorFlow atau Keras. Arsitektur model ditentukan dengan mengatur jumlah lapisan LSTM dan neuron di setiap lapisan.

3) Pelatihan Model: Model LSTM dilatih menggunakan data pelatihan. Selama pelatihan, model belajar dari pola dalam data historis untuk membuat prediksi yang lebih akurat.

4) Data Integration:

Combine Data Sources: menggabungkan data harga saham dengan data fundamental berdasarkan periode waktu (misalnya menyelaraskan data keuangan triwulanan dengan harga saham terkait).

Consistency Check: Memastikan semua kumpulan data disinkronkan dan disejajarkan dengan benar untuk menghindari ketidakcocokan waktu.

8) Hyperparameter Tuning proses yang dilakukan bersama untuk melakukan uji coba pada keseluruhan parameter yang digunakan pada model prediksi. Tuning menggunakan Random Search untuk menemukan kombinasi hyperparameter terbaik untuk model prediksi.

9) Model Evaluation adalah hasil langkah yang berasal dari langkah Hyperparameter Tuning. Jika nilai yang diinginkan tercapai, jika tidak, akan diulang dari langkah prediksi saham hingga mencapai nilai yang diinginkan. Tujuan dari langkah ini adalah untuk mendapatkan nilai optimum dari setiap saham dan akan ditransfer ke langkah berikutnya. Model evaluasi yang digunakan pada tahap ini adalah Root Mean Square Error (RMSE) dan Mean Absolute Percentage Error (MAPE) untuk menentukan nilai prediksi dengan nilai aktual. RMSE digunakan untuk melakukan analisis terhadap kesalahan besar, misalnya dalam kasus fluktuasi harga saham yang luar biasa yang dapat menyebabkan kerugian finansial yang signifikan sementara MAPE digunakan untuk mengukur kinerja model secara konsisten di berbagai jenis

saham, terutama dalam kasus di mana perbedaan harga antar saham cukup signifikan. berikut adalah rumus RMSE dan MAPE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

Keterangan:

\hat{y}_i = Nilai hasil prediksi

y_i = Nilai aktual / sebenarnya

n = Banyaknya data

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (2)$$

Keterangan:

A_t = Nilai aktual / sebenarnya

F_t = Nilai hasil prediksi

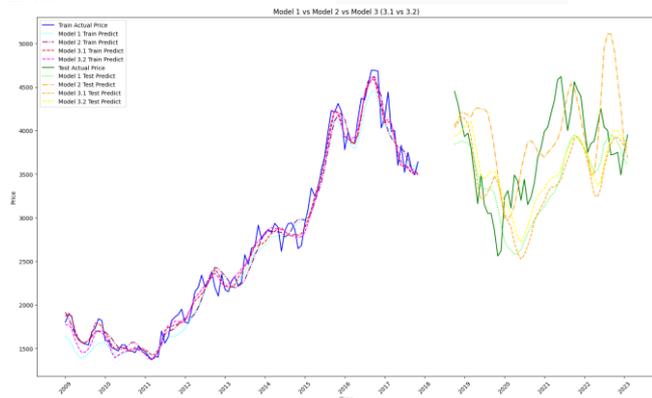
n = Banyaknya data

IV. HASIL DAN PEMBAHASAN

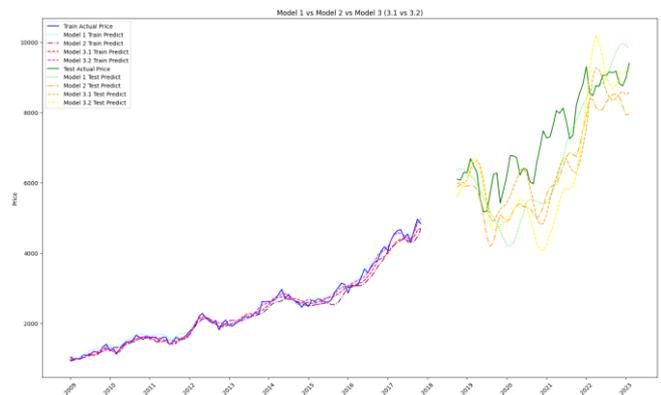
Pada penelitian ini, dilakukan tiga skenario pengujian untuk mengevaluasi beberapa faktor. Pertama, terdapat analisis model prediksi menggunakan LSTM dengan data historis. Kedua, pengaruh data fundamental jika ditambah pada model prediksi. Ketiga, menambahkan PCA dengan 4 fitur dan PCA dengan 8 fitur pada model prediksi.

A. Hasil Pengujian

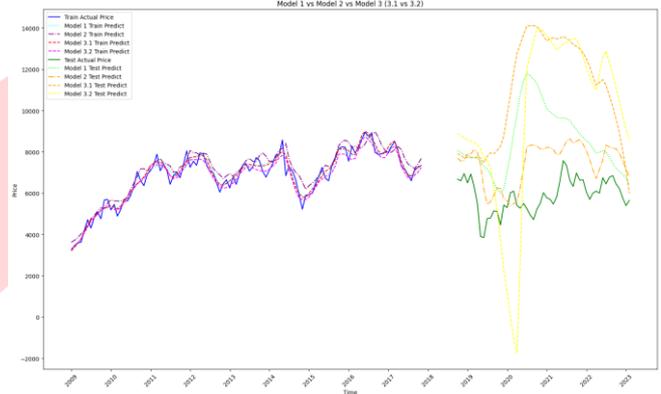
Pertama, pada eksperimen ini dilakukan pengujian menggunakan rasio data 65:35 untuk data pengujian. Terdapat empat model yang diuji untuk setiap dataset. Untuk skenario pertama menggunakan metode LSTM dengan data historis. Kedua, Dalam pengujian selanjutnya dilakukan pengujian pengaruh data fundamental pada model prediksi tersebut. Ketiga, Untuk pengujian selanjutnya, kali ini menambah PCA dengan empat fitur dan PCA dengan delapan fitur.



Gambar 4. Prediksi Saham TLKM



Gambar 5. Prediksi Saham BBCA



Gambar 6. Prediksi Saham ASII

Table 2. Hasil Pengujian Keseluruhan Model Prediksi

Ratio	Saham	Model	Trainin g Phase (RMSE)	Testing Phase (RMSE)	Trainin g Phase (MAP E)	Testing Phase (MAP E)
65:35	TLKM	Model Data Historis	0.14	0.19	5.46	5.96
	BBCA		0.25	0.81	4.76	5.34
	ASII		0.33	0.5	5.51	10.00
	TLKM	Model Data Historis dengan Data Funda mental	0.11	0.33	4.39	8.79
	BBCA		0.27	1.31	5.12	5.87
	ASII		0.29	0.79	5.09	14.92
65:35	TLKM	Model PCA dengan empat fitur	0.08	0.56	3.17	16.22
	BBCA		0.21	1.98	3.97	11.64
	ASII		0.21	4.3	3.22	87.30
	TLKM	Model PCA dengan delapa n fitur	0.09	0.39	3.37	11.49
	BBCA		0.15	2.71	2.86	16.61

	ASII		0.22	4.19	3.42	85.26
--	------	--	------	------	------	-------

B. Analisis Hasil Pengujian

Pada tahap pengujian pertama, model LSTM dengan data historis menunjukkan performa yang cukup baik pada fase training, dengan nilai RMSE yang lebih rendah dibandingkan fase testing. Namun, saat diuji menggunakan data testing, terjadi peningkatan signifikan pada nilai RMSE, terutama untuk saham BBCA dan ASII. Metrik MAPE juga menunjukkan hasil yang stabil pada fase training, tetapi mengalami peningkatan pada fase testing, yang mengindikasikan bahwa model mengalami kesulitan dalam melakukan generalisasi terhadap data baru.

Pada tahap pengujian kedua, penambahan data fundamental memberikan dampak pada peningkatan nilai RMSE testing dibandingkan dengan model yang hanya menggunakan data historis. Namun, untuk saham TLKM dan BBCA, peningkatan RMSE testing tidak terlalu signifikan, meskipun tetap lebih tinggi daripada fase training. MAPE testing juga mengalami peningkatan signifikan, terutama untuk saham ASII, di mana MAPE melonjak dari 10.00 menjadi 14.92. Hal ini menunjukkan bahwa meskipun data fundamental telah ditambahkan, model masih mengalami kesulitan dalam memberikan prediksi yang akurat.

Pada tahap pengujian ketiga, penggunaan model PCA, baik dengan empat maupun delapan fitur, menghasilkan peningkatan nilai RMSE dan MAPE yang sangat tinggi pada fase testing dibandingkan dengan fase training. Hasil ini menunjukkan bahwa model PCA tidak mampu menangkap pola yang cukup untuk menghasilkan prediksi yang akurat pada data testing, yang mengindikasikan adanya overfitting yang parah. MAPE yang sangat tinggi pada fase testing memperlihatkan bahwa prediksi model PCA sangat jauh dari nilai aktual, sehingga model ini tidak dapat diandalkan untuk tujuan prediksi.

V. KESIMPULAN

Peningkatan performa prediksi harga saham satu bulan ke depan tidak dapat dicapai dengan penggunaan LSTM dengan data historis dan data fundamental, serta teknik pengurangan dimensi seperti PCA. Namun, temuan penelitian menunjukkan bahwa setiap model LSTM yang diuji mengalami masalah overfitting yang signifikan. Model tidak menghasilkan hasil yang akurat pada data pengujian, meskipun mereka berhasil dengan data pelatihan. Selain tidak mampu mengatasi masalah overfitting, penambahan data fundamental dalam model LSTM tidak meningkatkan akurasi prediksi secara signifikan. Selain itu, penggunaan PCA dengan empat dan delapan fitur menunjukkan gejala overfitting yang signifikan, menunjukkan bahwa model LSTM tidak dapat menangkap pola yang relevan dalam data pengujian. Penelitian lanjutan harus mempertimbangkan pengayaan data fundamental, penggunaan hyperparameter tambahan, dan pengujian algoritma atau metode lain untuk meningkatkan kinerja model prediksi pada dataset yang sama.

REFERENSI

[1] E. Solares, V. De-León-Gómez, F. G. Salas, and R. Díaz, "A comprehensive decision support system

for stock investment decisions," **Expert Systems with Applications**, vol. 210, 2022, Art. no. 118485. DOI: 10.1016/j.eswa.2022.118485.

- [2] T. Khuat, T. M. H. Le, Q. Le, and B. Nguyen, "Forecasting Stock Price using Wavelet Neural Network optimized by Directed Artificial Bee Colony Algorithm," **Journal of Telecommunications and Information Technology**, no. 2, pp. 43-52, 2016.
- [3] Y. Guo, "Stock Price Prediction Based on LSTM Neural Network: the Effectiveness of News Sentiment Analysis," in **Proc. 2020 2nd Int. Conf. Economic Management and Model Engineering (ICEMME)**, Chongqing, China, 2020, pp. 1018-1024. DOI: 10.1109/ICEMME51517.2020.00206.
- [4] J. Cahyani, S. Mujahidin, and T. Fiqar, "Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional," **Jurnal Sistem dan Teknologi Informasi (JustIN)**, vol. 11, no. 2, pp. 346-354, 2023. DOI: 10.26418/justin.v11i2.57395.
- [5] A. K. Gárate-Escamila, A. H. El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," **Informatics in Medicine Unlocked**, vol. 19, 2020, Art. no. 100330. DOI: 10.1016/j.imu.2020.100330.
- [6] C. Han, Y. Feng, and T. Wang, "From Random Search to Bandit Learning in Metric Measure Spaces," **ArXiv**, May 2023. DOI: 10.48550/arXiv.2305.11509.
- [7] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," **Neurocomputing**, vol. 323, pp. 203-213, 2019. DOI: 10.1016/j.neucom.2018.09.082.
- [8] H. Abdi and L. Williams, "Principal component analysis," **Wiley Interdisciplinary Reviews: Computational Statistics**, vol. 2, pp. 433-459, 2010. DOI: 10.1002/wics.101.
- [9] T. Jeyaprakash and K. Padmaveni, "Introduction to Data Science - An Overview," **International Journal of Science and Management Studies (IJSMS)**, 2021. DOI: 10.51386/25815946/ijms-v4i4p137.
- [10] S. Paul et al., "Components of Data Science and its Applications," **International Journal of Engineering Applied Sciences and Technology**, vol. 7, no. 5, 2022. DOI: 10.33564/ijeast.2022.v07i05.039.
- [11] J. Halawa, A. Hermawan, and J. ., "Implementation of Linear Regression Algorithm to Predict Stock Prices Based on Historical Data," **bit-Tech**, vol. 5, no. 2, pp. 123-132, 2022. DOI: 10.32877/bt.v5i2.616.
- [12] K. M., "Analysis of Stock Market Data for Guidance of Stock Investors," **International Journal of Scientific Research in Engineering and Management**, vol. 5, no. 2, 2023. DOI: 10.55041/ijrsrem24672.
- [13] A. Li, Q. Wei, Y. Shi, and Z. Liu, "Research on stock price prediction from a data fusion perspective," **Data Science in Finance and Economics**, vol. 3, no. 1, pp. 56-70, 2023. DOI: 10.3934/dsfe.2023014.

