

# Perbandingan Metode *Long Short-Term Memory* dan *Double Random Forest* dalam Prediksi Harga Penutupan Saham

Muhamad Terbit Zikri, Freza Riana, Gibtha Fitri Laxmi

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Ibn Khaldun Bogor, Indonesia

\*E-mail koresponden: [freza@uika-bogor.ac.id](mailto:freza@uika-bogor.ac.id)

Diserahkan 3 Agustus 2024; Direview 8 Oktober 2024; Dipublikasikan 18 Oktober 2024

## Abstrak

*Long Short-Term Memory (LSTM)* merupakan bagian dari *Recurrent Neural Network (RNN)* yang memiliki keunggulan dalam memproses data sekuensial dan mengenali pola serta ketergantungan dalam data berurutan, sementara *Double Random Forest (DRF)* adalah metode ensemble yang mampu menangkap pola yang kompleks dengan memanfaatkan pohon keputusan dari data pelatihan secara keseluruhan. Kedua metode ini dapat digunakan untuk melakukan forecasting, termasuk dalam kasus prediksi harga saham. Penelitian ini bertujuan untuk membandingkan metode *LSTM* dan *DRF* dalam konteks prediksi harga penutupan saham pada PT. Indofood Sukses Makmur, Tbk, menggunakan dataset berukuran 1.253 data yang akan dibagi menjadi 80% data latih dan 20% data uji. Eksperimen ini mencakup tiga skenario, yaitu penggunaan semua fitur yang tersedia, penggunaan fitur dengan korelasi linier positif tinggi, dan penggunaan fitur dengan korelasi linier rendah. Error metrics yang digunakan pada penelitian ini adalah *RMSE*, *MAE*, dan *MAPE*. Dari hasil penelitian prediksi harga penutupan PT. Indofood Sukses Makmur, Tbk diperoleh bahwa *LSTM* memiliki hasil yang lebih baik dibandingkan *DRF* dengan nilai *RMSE* (80.99), *MAE* (61.08), dan *MAPE* (0.94). Hasil dari riset ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem prediksi harga saham berbasis machine learning.

**Kata kunci:** *DRF*, *LSTM*, *Machine Learning*, *Peramalan*, *Saham*.

## Abstract

*Long Short-Term Memory (LSTM)* is a type of *Recurrent Neural Network (RNN)* that excels in processing sequential data and identifying patterns and dependencies within ordered datasets. In contrast, *Double Random Forest (DRF)* is an ensemble method capable of capturing complex patterns by utilizing decision trees constructed from the entire training dataset. Both methods can be applied for forecasting purposes, including stock price prediction. This study aims to compare the performance of *LSTM* and *DRF* in predicting the closing stock price of PT. Indofood Sukses Makmur, Tbk, using a dataset consisting of 1,253 records, which will be divided into 80% training data and 20% testing data. The experiment involves three scenarios: using all available features, using features with high positive linear correlation, and using features with low linear correlation. The error metrics employed in this study are *RMSE*, *MAE*, and *MAPE*. The findings indicate that *LSTM* outperforms *DRF* in predicting the closing stock price of PT. Indofood Sukses Makmur, Tbk, with lower error values — *RMSE* (80.99), *MAE*

(61.08), and MAPE (0.94). The results of this research are expected to provide a valuable contribution to the development of machine learning-based stock price prediction systems.

**Keywords:** DRF, Forecasting, LSTM, Machine Learning, Stock Prices.

## PENDAHULUAN

Saham menjadi pilihan investasi menarik berkat perkembangan teknologi yang pesat. Hal ini didukung jurnal ilmiah yang menunjukkan bahwa kemajuan teknologi memiliki pengaruh positif dan signifikan terhadap minat investasi. Perkembangan teknologi yang semakin pesat telah mempermudah akses ke informasi pasar modal, dan hal ini telah berhasil meningkatkan minat investor atau calon investor untuk terlibat dalam investasi [1]. Melakukan investasi di pasar saham bisa memberikan potensi keuntungan yang besar, namun sekaligus juga membawa risiko yang tinggi. Hal ini sesuai dengan prinsip umum investasi “*high risk, high return*”. Artinya, semakin tinggi potensi risiko investasi, semakin tinggi pengembalinya. Sebaliknya, setiap investasi dengan potensi risiko rendah juga akan memiliki tingkat pengembalian yang rendah [2].

Para investor perlu melakukan strategi investasi yang tepat dalam investasi saham, dan analisis yang mendalam sebelum mengambil keputusan [3]. Oleh karena itu, diperlukan analisis pergerakan harga saham agar investor dapat memanfaatkan hasil analisis tersebut dalam pengambilan keputusan investasi. Melalui data historis tren nilai saham investor dapat melakukan peramalan pada waktu yang akan datang, salah satu metode umum yang digunakan dalam menentukan peramalan harga saham adalah metode *forecast* [4]. Metode yang dapat digunakan untuk peramalan harga saham yaitu Double Random Forest (DRF) dan Long Short-Term Memory (LSTM) [5].

Salah satu pendekatan dalam jaringan saraf berulang adalah LSTM. Model jaringan saraf berulang ini memiliki keunggulan dalam menangani data yang memiliki urutan atau sifat sekuensial. Dalam penggunaannya, model ini mampu mengidentifikasi pola dan keterkaitan antar elemen dalam data yang tersusun secara berurutan, seperti pada teks, sinyal audio, maupun deretan peristiwa berdasarkan waktu [6]. Penerapan metode LSTM pada data sekuensial digunakan untuk prediksi terhadap data deret waktu. LSTM bisa mengidentifikasi data yang akan disimpan dan data yang tidak terpakai akan dihapus. LSTM memiliki empat lapisan *neuron* yang disebut gerbang yang mengontrol memori setiap *neuron* [7].

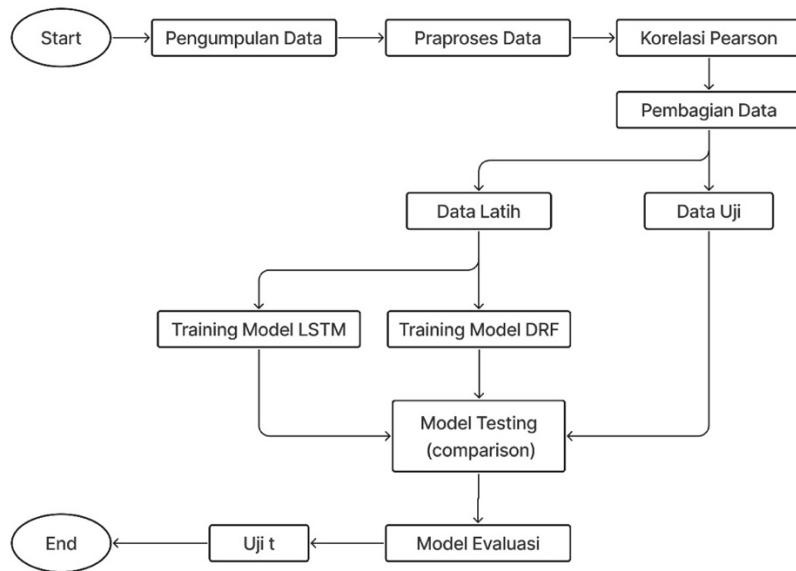
DRF adalah metode *ensemble* dalam analisis data dan pembuatan model prediksi. Berbeda dari metode *Random Forest* (RF), DRF menggunakan seluruh data pelatihan untuk menumbuhkan pohon keputusan, sehingga pohon-pohon yang dihasilkan cenderung lebih besar dan dapat menangkap pola yang lebih kompleks. Keunggulan DRF terletak pada kemampuannya mengatasi masalah bias prediksi dan menghadirkan *ensemble* pohon yang lebih beragam dan efisien, serta *ensemble* dapat meningkatkan akurasi model-modelnya [20]. DRF dalam skala besar dapat dijalankan pada sistem yang terdistribusi untuk mengatasi data besar dan kompleks dengan baik [8].

Pada penelitian ini harga saham yang diprediksi adalah saham milik PT. Indofood Sukses Makmur, Tbk (INDF). Perusahaan ini termasuk dalam kategori perusahaan *Fast Moving Consumer Goods* (FMCG) terkemuka di Indonesia [9]. INDF dalam beberapa tahun terakhir masuk pada daftar saham LQ45. Indeks LQ45 mencerminkan performa saham dari 45 perusahaan dengan likuiditas tinggi dan kapitalisasi pasar besar, serta didukung oleh fundamental perusahaan yang kuat. [10]. Dalam data harga saham ini, terdapat tujuh fitur yaitu, tanggal (*date*), harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), harga penutupan yang disesuaikan (*adj close*), dan volume perdagangan (*volume*).

Penelitian ini bertujuan untuk mengembangkan model prediksi harga penutupan saham INDF dan membandingkan kinerja dua metode, yaitu LSTM dan DRF, untuk menentukan metode mana yang lebih baik dalam melakukan prediksi harga saham tersebut.

## METODE PENELITIAN

Percobaan yang akan dilakukan adalah membandingkan hasil Hasil dari model LSTM dengan DRF yang akan dibandingkan menggunakan *error metrics* seperti RMSE, MAE, dan MAPE. Terdapat tiga macam percobaan, yaitu menggunakan semua fitur yang ada, menggunakan fitur yang memiliki hubungan linier kuat dan diatasnya, yang terakhir adalah fitur dengan hubungan linier sangat rendah. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1 Alur penelitian

## Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data historis saham PT. Indofood Sukses Makmur, Tbk yang diperoleh dari website Yahoo Finance [11]. Data yang dikumpulkan terdapat tujuh fitur, yaitu tanggal (*date*), harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), harga penutupan yang disesuaikan (*adj close*), dan volume perdagangan (*volume*). Data ini meliputi periode waktu mulai dari 1 Januari 2018 hingga 30 Desember 2022. Data saham bisa dilihat pada Tabel 1.

Tabel 1 Data Saham

Date	Open	High	Low	Close	Adj Close	Volume
1/1/2018	7625	7625	7625	7625	6258.996	0
1/2/2018	7650	7700	7550	7550	6197.431	6875700
1/3/2018	7600	7700	7600	7650	6279.517	11532300
1/4/2018	7650	7725	7600	7725	6341.081	6861700
1/5/2018	7700	7700	7625	7675	6300.038	8060200
...	...	...	...	...	...	...
12/26/2022	6775	6875	6775	6875	6875	1596500
12/27/2022	6875	6900	6800	6800	6800	1878100
12/28/2022	6800	6850	6675	6725	6725	5291600
12/29/2022	6700	6800	6700	6800	6800	4037800
12/30/2022	6800	6800	6700	6725	6725	3503600

## Praproses Data

Praproses data ini penting untuk memastikan data yang digunakan dalam pelatihan model berkualitas dan siap untuk analisis lebih lanjut. Berikut adalah langkah-langkah yang akan dilakukan dalam tahap pra-pemrosesan data:

### **Mengisi Nilai yang Kosong**

Nilai kosong dalam *dataset*, dapat diatasi menggunakan nilai sebelum dan sesudahnya. Pendekatan ini membantu dalam mengisi kekosongan data berdasarkan konteks sekitarnya.

### **Normalisasi Data**

Normalisasi dilakukan hanya pada data untuk model LSTM agar bekerja lebih baik. Normalisasi data ini menggunakan teknik *min-max scaling* pada data set yang telah disiapkan, sebelum dinormalisasi kolom *date* dihilangkan terlebih dahulu. Hasil normalisasi dapat dilihat pada Tabel 2.

**Tabel 2** Normalisasi Data

Open	High	Low	Close	Adj Close	Volume
0.784483	0.75	0.813953	0.804688	0.643176	0
0.793103	0.777778	0.790698	0.78125	0.621327	0.08475
0.775862	0.777778	0.806202	0.8125	0.650459	0.142148
0.793103	0.787037	0.806202	0.835938	0.672308	0.084578
...	...	...	...	...	...
0.491379	0.472222	0.550388	0.570313	0.861795	0.019679
0.525862	0.481481	0.55814	0.546875	0.835177	0.02315
0.5	0.462963	0.51938	0.523438	0.80856	0.065225
0.465517	0.444444	0.527132	0.546875	0.835177	0.04977
0.5	0.444444	0.527132	0.523438	0.80856	0.043186

### **Segmentasi Data**

Segmentasi merupakan proses pemisahan dan pengelompokan data pada rentang tertentu. Segmentasi pada penelitian ini dilakukan hanya pada model LSTM dengan *time step* tujuh data agar menghasilkan prediksi yang lebih baik. Pada *input* pertama akan menggunakan data kesatu sampai ketujuh, *input* kedua menggunakan data kedua sampai kedelapan dan seterusnya. Ilustrasi segmentasi dapat dilihat pada Gambar 2.



**Gambar 2** Segmentasi data

### **Korelasi Pearson**

Korelasi adalah analisis *Bivariat* yang mengukur kekuatan hubungan antara dua variabel. Dalam statistik, nilai koefisien korelasi berkisar antara +1 hingga -1. Ketika nilai koefisien korelasi berada di sekitar 1 atau -1, maka dikatakan memiliki hubungan yang sempurna antara dua variabel tersebut. Seiring dengan nilai koefisien korelasi mendekati 0, hubungan antara dua

variabel akan menjadi lebih lemah[12]. Kriteria dari koefisien korelasi Pearson dapat dilihat pada Tabel 3 [13]. Selain itu, proses identifikasi fitur dilakukan menggunakan metode korelasi Pearson. Hasil korelasi Pearson dapat dilihat pada Tabel 4. Fitur *open*, *high*, *low*, *close* memiliki korelasi positif yang sangat kuat dengan variabel target, sementara fitur *adj close* memiliki korelasi positif yang kuat. Fitur *volume* di sisi lain memiliki korelasi yang sangat rendah dengan Variabel target.

**Tabel 3** Kriteria Korelasi Pearson

Nilai r	Interpretasi
0.00 – 0.199	Sangat Rendah
0.20 – 0.399	Rendah
0.40 – 0.599	Sedang
0.60 – 0.799	Kuat
0.80 – 1.000	Sangat Kuat

**Tabel 4** Korelasi Pearson

Fitur	Hasil
Open	0.960219
High	0.972519
Low	0.969952
Close	0.978697
Adj Close	0.770870
Volume	0.087776

**Tabel 5** Pembagian Data

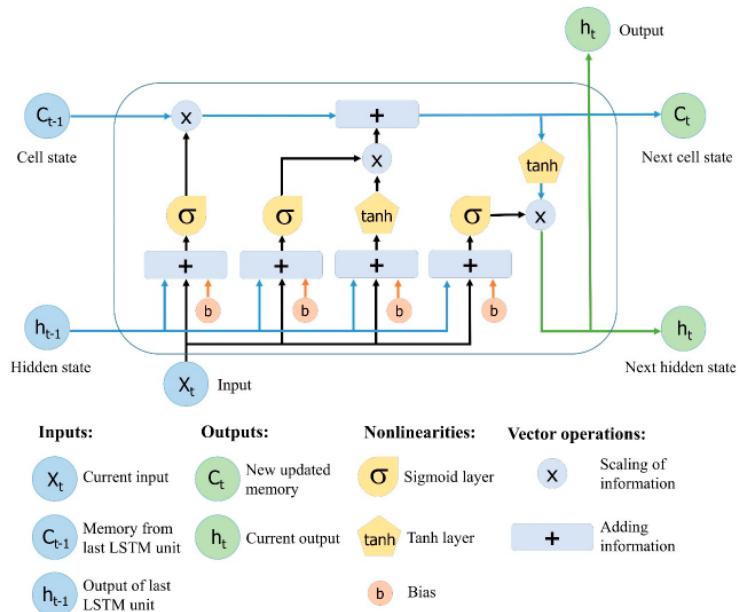
Metode	Data Latih	Data Uji	Total
LSTM	997	250	1247
DRF	1002	251	1253

## Pembagian Data

Penelitian ini mengalokasikan data menjadi dua jenis, yaitu data training atau data latih, dan data testing atau data uji. Data tersebut dibagi menggunakan perbandingan 80% : 20%, dimana 80% dari total data digunakan sebagai data latih, sedangkan 20% sisanya digunakan sebagai data uji. Data untuk model LSTM akan lebih sedikit dibandingkan dengan DRF karena data input untuk LSTM akan dilakukan segmentasi setiap inputnya berjumlah tujuh data. Pembagian data untuk dilakukan proses pemodelan dapat dilihat pada Tabel 5.

## Long Short-Term Memory

Salah satu jenis RNN yang populer yaitu *Long Short-Term Memory* (LSTM). Kemampuan LSTM yang dimiliki adalah memperpanjang memori jaringan dengan mengontrol aliran informasi ke dalam dan ke luar dari memori. Hal ini memungkinkan LSTM untuk mengatasi masalah *vanishing gradient* yang dapat terjadi pada RNN biasa saat menghadapi urutan waktu yang panjang [14]. Arsitektur LSTM dapat dilihat pada Gambar 3[15].

**Gambar 3** LSTM[15]

Pada percobaan kali ini pembuatan model LSTM menggunakan satu layer LSTM, *batch size* 32, *loss function mean square error*, *optimizer Adam*, *timestep* tujuh data dan jumlah *output* satu. Karena ada tiga jenis percobaan yang akan dilakukan, maka akan dibuat tiga jenis model akan menyesuaikan dengan jumlah fitur tiap percobaan.

### Double Random Forest

*Double Random Forest* (DRF) adalah metode *ensemble* yang bertujuan untuk meningkatkan kinerja RF. Metode ini menggunakan seluruh data pelatihan untuk membuat *decision tree* pada setiap tahap *ensemble*, berbeda dengan RF yang menggunakan sampel *bootstrap* acak. Hal ini menyebabkan DRF cenderung menghasilkan pohon-pohon yang lebih besar dibandingkan dengan RF, hal ini meningkatkan ukuran pohon, sehingga mengurangi bias. Selain itu, metode ini menambahkan ketidakpastian pada aturan partisi melalui *bootstrap* dan subkumpulan fitur, yang memungkinkan himpunan pohon menjadi lebih beragam [8]. Berikut merupakan algoritma DRF [5]:

$D$  : data pelatihan dengan  $n$  instansi,  $p$  fitur, dan variabel target.

$D_t$  : data pelatihan dengan  $n_t$  instansi,  $p$  fitur, dan variabel target pada node  $t$ .

$K$  : jumlah kelas dalam variabel target.

$B$  : jumlah pengklasifikasi dalam DRF.

1. Gunakan data pelatihan  $D$ .
2. Bangun sebuah pohon menggunakan subkumpulan fitur acak dan sampel *bootstrap* acak dari instansi di data pelatihan  $D$ .

Untuk sebuah node  $t$

- i. Jika  $n_t > n \times 0.1$ , buat sampel *bootstrap*  $D_t$  dari sampel  $D$ . jika tidak  $D_t = D$
  - ii. Pilih secara acak  $p \approx m3$  atau  $\sqrt{p}$  features.
  - iii. Cari fitur dan titik pemotongan terbaik menggunakan subkumpulan fitur acak.
  - iv. Kirim data ke bawah menggunakan fitur dan titik pemotongan terbaik. Ulangi langkah (i)-(iv) hingga aturan penghentian terpenuhi.
3. Konstruksi pengklasifikasi yang terlatih  $C_b$ .

### Model Testing

Setelah model berhasil dibuat, Model LSTM dan DRF akan diuji menggunakan data uji yang telah dipersiapkan yaitu sebanyak 20% secara acak dari keseluruhan data pada tahap pembagian data yang tidak termasuk dalam data latih.

### Evaluasi Model

Pada tahap evaluasi model, hasil dari kedua model yang telah diuji menggunakan data uji sebelumnya, yaitu model LSTM dan DRF akan dibandingkan. Pada tahap evaluasi menggunakan *error metrics* seperti RMSE, MAE, dan MAPE.

### RMSE

*Root Mean Square Error* (RMSE) adalah akar kuadrat dari *Mean Square Error* (MSE) adalah metrik yang umum digunakan untuk mengukur rata-rata perbedaan kuadrat antara nilai prediksi dan nilai sebenarnya. Persamaan 1 digunakan untuk menghitung RMSE [16].

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

Keterangan :

- $n$  adalah jumlah data poin yang dievaluasi.
- $y_i$  adalah nilai sebenarnya dari data poin ke- $i$ .
- $\hat{y}_i$  adalah nilai prediksi dari data poin ke- $i$ .

### MAE

*Mean Absolute Error* (MAE) adalah nilai mutlak dari selisih antara nilai sebenarnya dengan nilai prediksi dan kemudian menghitung rata-rata dari nilai-nilai mutlak tersebut. Persamaan 2 digunakan untuk menghitung MAE [16]:

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

Keterangan:

- $n$  adalah jumlah data poin yang dievaluasi.
- $y_i$  adalah nilai sebenarnya dari data poin ke- $i$ .

$\hat{y}_i$  adalah nilai prediksi dari data poin ke- $i$ .

### MAPE

*Mean Absolute Percentage Error* (MAPE) adalah persentase kesalahan rata-rata antara nilai sebenarnya dan nilai prediksi, untuk mengevaluasi kesalahan prediksi dalam perspektif persentase. Persamaan 3 digunakan untuk menghitung MAE [17]:

$$\text{MAPE} = \frac{\sum_{t=1}^n \left| \frac{a - b}{a} \right|}{n} \times 100 \quad (3)$$

Keterangan:

- $n$  adalah jumlah data poin yang dievaluasi.
- $a$  adalah nilai sebenarnya.
- $b$  adalah nilai hasil.

### Uji Signifikansi Metode (Uji Statistik -t)

Uji statistik -t digunakan untuk menentukan apakah terdapat perbedaan signifikan antara dua metode yang sedang diuji. Dalam penelitian ini, metode uji statistik -t yang diterapkan adalah uji-t berpasangan (*Paired T-test*) [18].

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \quad (4)$$

Dalam konteks tersebut  $n$  adalah jumlah pasangan sampel,  $d$  adalah selisih dari sampel yang berpasangan,  $\bar{d}$  adalah rataan selisih dari sampel yang berpasangan,  $\mu_d$  adalah rataan selisih populasi, dan  $s_d$  adalah simpangan baku selisih sampel. Simpangan baku selisih sampel dinyatakan dengan Persamaan 5 [18,19]:

$$S_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n - 1}} \quad (5)$$

Tingkat signifikansi diuji dengan nilai pasti tingkat signifikansi  $\alpha$  sebesar 0.05. Hipotesis  $H_0$  dan  $H_1$  digunakan untuk pengambilan keputusan pada uji-t berpasangan. Hipotesis nol ( $H_0$ ) akan diabaikan jika nilai uji statistik lebih kecil atau sama dengan tingkat signifikansi  $\alpha$ , yang menunjukkan penerimaan hipotesis alternatif ( $H_1$ ). Pengujian hipotesis ini dilakukan seperti berikut:

$H_0 : D = 0$ , menunjukkan bahwa kinerja kedua metode tidak berbeda secara signifikan.

$H_1 : D \neq 0$ , menunjukkan bahwa kinerja kedua metode berbeda secara signifikan.

## HASIL DAN PEMBAHASAN

### Skenario Percobaan Model

Skenario yang akan dilakukan terdiri dari tiga jenis: menggunakan semua fitur yang tersedia, menggunakan fitur-fitur dengan korelasi linier positif tinggi dan fitur dengan korelasi linier rendah, dapat dilihat dalam Tabel 4. Dalam setiap skenario, akan ada optimasi *hyperparameter* untuk kedua model yang dapat dilihat dalam Tabel 6.

**Tabel 6** Hyperparameter

Model	Hyperparameter	Nilai
LSTM	<i>Epoch</i>	50, 100, 150, 200, 250
	<i>Learning rate</i>	0.001, 0.0025, 0.005
DRF	<i>max_depth</i>	100, 200, 300, 400, 500
	<i>n_estimators</i>	1, 2, 3, 4, 5, 6, 7

### Evaluasi Model

Dalam penelitian ini, model dievaluasi berdasarkan nilai RMSE, MAE, dan MAPE. Parameter optimal untuk LSTM dan DRF dalam setiap skenario dapat ditemukan dalam Tabel 7.

**Tabel 7** Optimal Parameter

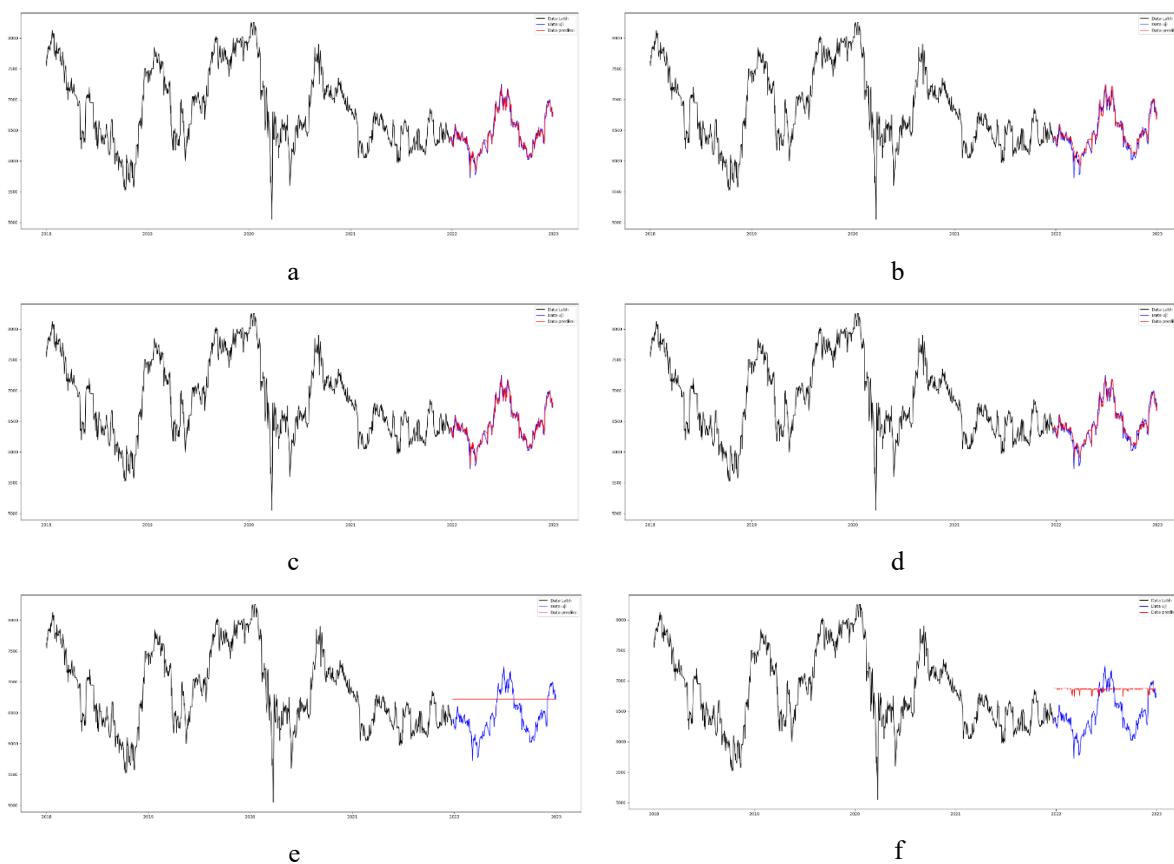
Fitur	Model	Hyperparameter	Nilai
Semua fitur	LSTM	<i>Learning rate</i>	0.001
		<i>Epoch</i>	200
	DRF	<i>Node</i>	6
		<i>n-tree</i>	100
		<i>Learning rate</i>	0.001
Fitur dengan korelasi linier positif tinggi	LSTM	<i>Epoch</i>	200
		<i>Node</i>	7
		<i>n-tree</i>	300
	DRF	<i>Learning rate</i>	0.1
		<i>Epoch</i>	50
		<i>Node</i>	1
Fitur dengan korelasi linier rendah	LSTM	<i>n-tree</i>	100
		<i>Learning rate</i>	0.1
		<i>Epoch</i>	50
	DRF	<i>Node</i>	1
		<i>n-tree</i>	100
		<i>Learning rate</i>	0.1

Tabel 8 merupakan perbandingan hasil terbaik RMSE, MAE, dan MAPE dari tiga skenario yang dilakukan. Model LSTM dalam prosesnya menggunakan *Graphics Processing Unit* (GPU) agar proses yang dijalankan semakin cepat.

**Tabel 8** Hasil Prediksi

Fitur	Model	RMSE	MAE	MAPE	Waktu (detik)
Semua Fitur	LSTM	80.98	61.08	0.94	44.62
	DRF	89.78	68.45	1.06	41.19
Fitur dengan korelasi linier positif tinggi	LSTM	81.53	61.49	0.95	47.55
	DRF	88.04	66.69	1.03	93.35
Fitur dengan korelasi linier rendah	LSTM	412.73	359.67	5.7	14.6
	DRF	510.99	439.65	7	19.75

Visualisasi prediksi terbaik dari LSTM dan DRF dapat dilihat pada Gambar 5. Warna hitam dalam gambar menunjukkan data latih, warna biru menunjukkan data uji, dan warna merah menunjukkan hasil prediksi.



**Gambar 5** Percobaan LSTM semua fitur (a), DRF semua fitur (b), fitur hubungan linier kuat atau di atasnya (c), DRF fitur hubungan linier kuat atau di atasnya (d), fitur hubungan linier sangat rendah (e), DRF fitur hubungan linier sangat rendah (f)

Hasil dari nilai RMSE, MAE dan MAPE pada kedua model dapat dilihat bahwa LSTM lebih unggul dibandingkan dengan DRF dalam semua skenario yang dilakukan.

### Uji Signifikansi Metode (Uji Statistik -t)

Uji statistik -t pada penelitian ini digunakan untuk mengetahui apakah ada perbedaan yang signifikan pada hasil metode LSTM dan DRF. Hasil dari uji-t berpasangan dapat dilihat pada Tabel 9.

**Tabel 9** Hasil uji-t

Fitur yang digunakan	Hasil Uji-t
Semua Fitur	5,19921E-12
Fitur dengan korelasi linier positif tinggi	0,000669865
Fitur dengan korelasi linier rendah	3,6273E-191

Hasil uji statistik pada Tabel 9 memperoleh hasil bahwa metode LSTM dengan DRF pada semua fitur, fitur dengan hubungan linier kuat, dan fitur dengan hubungan linier lemah berbeda secara signifikan. Perbedaan ini ditunjukkan dari hasil uji-t yang bernilai lebih kecil dari  $\alpha = 0.05$ . Hasil ini menunjukkan bahwa  $H_0$  (kinerja kedua metode tidak berbeda signifikan) ditolak yang berarti menerima  $H_1$  (kinerja kedua metode berbeda signifikan), yang artinya metode LSTM berbeda nyata dengan metode DRF. Dari hasil akurasi yang dihasilkan LSTM lebih unggul dibandingkan DRF.

## KESIMPULAN

Kesimpulan dari hasil penelitian adalah metode LSTM dengan *epoch* 200 dan *learning rate* 0,001 menggunakan semua fitur yaitu *open*, *high*, *low*, *close*, *adjusted close*, dan *volume* mendapatkan hasil prediksi terbaik untuk harga penutupan saham INDF. Hasilnya mencatat sebesar RMSE (80.98), MAE (61.08), dan MAPE (0.944). LSTM juga memiliki hasil yang lebih baik daripada DRF pada data yang memiliki hubungan linier kuat maupun rendah.

Penelitian ini memberikan beberapa saran lain yang dapat digunakan yaitu memakai data saham di luar indeks LQ45. Selain itu, hasil prediksi data latih dari model DRF sebelumnya dapat digunakan menjadi fitur baru yang akan dimasukan ke dalam model DRF baru untuk mengevaluasi apakah dapat meningkatkan akurasi prediksi dari model DRF.

## DAFTAR PUSTAKA

1. Cahya BT. Pengaruh Motivasi dan Kemajuan Teknologi Terhadap Minat Investasi Saham. Al-Masharif: Jurnal Ilmu Ekonomi dan Keislaman. 2019 Dec 31;7(2):192-207.
2. Devaki A. Faktor-faktor yang Mempengaruhi Return Saham pada Perusahaan LQ45 di Bursa Efek Indonesia. Jurnal Benefita: Ekonomi Pembangunan, Manajemen Bisnis & Akuntansi. 2017 Jul 24;2(2):157-68.
3. Permata D, Suryawati RF. Analisis Portofolio Optimal Saham Syariah Jakarta Islamic Index (JII) Periode 2015-2017. Jurnal Manajemen Dan Organisasi. 2020 Apr 29;11(1):8-21.
4. Ramadhan VP, Pamuji FY. Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI). Jurnal Teknologi dan Manajemen Informatika. 2022 Jul 4;8(1):39-45.
5. Ratnasari AP, Susetyo B, Notodiputro KA. Comparison Of Double Random Forest And Long Short-term Memory Methods For Analyzing Economic Indicator Data. Barekeng: Jurnal Ilmu Matematika dan Terapan. 2023 Jun 11;17(2):0757-66.
6. Faturrohman F, Irawan B, Setianingsih C. Analisis Sentimen Pada Bpjs Kesehatan Menggunakan Recurrent Neural Network. eProceedings of Engineering. 2020 Aug 1;7(2).
7. Lattifia T, Buana PW, Rusjayanti NK. Model Prediksi Cuaca Menggunakan Metode LSTM. JITTER J. Ilm. Teknol. dan Komput. 2022;3(1):994-1000.

8. Han S, Kim H, Lee YS. Double random forest. Machine Learning. 2020 Aug;109:1569-86.
9. Ramang GD, Tumbel TM, Rogahang JJ. Analisis Rasio Keuangan Untuk Menilai Kinerja Keuangan pada PT. Indonesia Prima Property Tbk Jakarta Pusat. Jurnal Administrasi Bisnis (JAB). 2019 Oct 11;9(3):122-30.
10. PT Bursa Efek Indonesia. [Daring]. Jakarta: PT Bursa Efek Indonesia; [diakses pada 27 Mei 2023]. Tersedia pada: <https://www.idx.co.id/>
11. Historical Stock Prices of PT Indofood Sukses Makmur Tbk (INDF.JK) Yahoo Finance. [Daring]. [diakses pada 20 Mei 2023]. Tersedia pada: <https://finance.yahoo.com/quote/INDF.JK/history?period1=1514764800&period2=1672444800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true>.
12. Zaid MA. Correlation and Regression Analysis. Ankara – Turkey: The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRIC); 2015.
13. Miftahuddin M, Sitanggang AP, Setiawan I. Analisis Hubungan Antara Kelembaban Relatif Dengan Beberapa Variabel Iklim Dengan Pendekatan Korelasi Pearson Di Samudera Hindia. Jurnal Siger Matematika. 2021 Mar 31;2(1):25-33.
14. Drianto LA, Valentino MN, Wibowo SI, Setiawan SB, Santoso TJ, Christano MS, Andika PLR, Pradiva RA, Chandranama VAD, Gehakhala HLF. Inovasi Teknologi Kecerdasan Komputer. SIEGA Publisher.
15. Le XH, Ho HV, Lee G, Jung S. Application of long short-term memory (LSTM) neural network for flood forecasting. Water. 2019 Jul 5;11(7):1387.
16. Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development. 2022 Jul 19;15(14):5481-7.
17. Prayudani S, Hizriadi A, Lase YY, Fatmi Y. Analysis accuracy of forecasting measurement technique on random K-nearest neighbor (RKNN) using MAPE and MSE. InJournal of Physics: Conference Series 2019 Nov 1 (Vol. 1361, No. 1, p. 012089). IOP Publishing.
18. Harinaldi. Prinsip-Prinsip Statistik untuk Teknik dan Sains. Erlangga; 2005
19. Laxmi GF. Optimasi Pemilihan Parameter Dan Operator Pada Fuzzy Local Binary Pattern Menggunakan Multi Objective Genetic Algorithm. Tesis. Institut Pertanian Bogor; 2012.
20. Santoni MM, Chamidah N, Prasvita DS. Strategi Ensemble Deep Learning pada Global Multi-Scale dan Local Attention Features pada Pengenalan Ekspresi Wajah. Krea-TIF [Internet]. 31 Mei 2024 [dikutip 18 Oktober 2024];12(1):12-23. Tersedia pada: <https://ejournal.uika-bogor.ac.id/index.php/krea-tif/article/view/16287>