

**COMPARISON OF EXTREME GRADIENT BOOSTING (XGBOOST) METHOD
WITH LONG SHORT-TERM MEMORY (LSTM) FOR STOCK PREDICTION PT.
BANK MANDIRI TBK. (BMRI)**

**PERBANDINGAN METODE EXTREME GRADIENT BOOSTING (XGBOOST)
DENGAN LONG SHORT-TERM MEMORY (LSTM) UNTUK PREDIKSI
SAHAM PT. BANK MANDIRI TBK. (BMRI)**

Bagas Pratama¹, Lintang Yuniar Banowosari²

Program Studi Manajemen Sistem Infomasi, Universitas Gunadarma, Depok, Indonesia¹

Program Studi Informatika, Universitas Gunadarma, Depok, Indonesia²

m.bagas057@gmail.com^{1*}, lintang@staff.gunadarma.ac.id²

ABSTRACT

In the continuously evolving world of investments, achieving optimal investment results is the primary goal of every investor. The high profit potential in stock investments makes it an attractive choice. However, it is difficult to predict the direction of stock price movements, but many methods and ways are used to predict in terms of buying and selling, one of which is the result of the very rapid development of computing for machine learning, namely artificial intelligence techniques. In this study, the artificial intelligence methods used are Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM). This study uses a stock dataset of PT. Bank Mandiri Tbk. (BMRI) for 20 years, especially in the open price column. After testing, the results from the XGBoost method are obtained, namely the Coefficient of Determination (R²) value of 89.09%, indicating that the results are good and the Mean Absolute Percentage Error (MAPE) is 3.21%, indicating that the error percentage is low. While in the LSTM method, the R² value is 98.44%, meaning that the prediction results have been predicted very well and the MAPE is 1.77%, indicating that the error percentage is very low.

Keywords: XGBoost, Open Price, Prediction, LSTM, Stocks

(ABSTRAK)

Dalam dunia investasi yang terus berkembang, pencapaian hasil investasi yang optimal merupakan tujuan utama setiap investor. Potensi keuntungan yang tinggi pada investasi saham menjadikannya pilihan menarik. Namun, pergerakan harga saham sulit untuk ditebak akan arah kelajuannya, tetapi banyak metode dan cara yang dilakukan untuk memprediksi dalam hal jual beli saham, salah satunya adalah akibat perkembangan komputasi untuk pembelajaran mesin yang sangat pesat, yaitu teknik kecerdasan buatan. Dalam penelitian ini metode kecerdasan buatan yang digunakan adalah Extreme Gradient Boosting (XGBoost) dan Long Short-Term Memory (LSTM). Penelitian ini menggunakan dataset saham PT. Bank Mandiri Tbk. (BMRI) selama 20 tahun, khususnya pada kolom harga pembukaan. Setelah dilakukan pengujian didapatkan hasil dari metode XGBoost, yaitu nilai Koefisien Determinasi (R²) sebesar 89.09%, menandakan bahwa hasil prediksi sudah baik dan Mean Absolute Percentage Error (MAPE) sebesar 3.21%, menandakan bahwa persentase error sudah rendah. Sedangkan pada metode LSTM, yaitu nilai R² sebesar 98.44% menandakan bahwa hasil prediksi sudah sangat baik dan MAPE sebesar 1.77%, menandakan bahwa persentase error sudah sangat rendah.

Kata Kunci: XGBoost, Harga Pembukaan, Prediksi, LSTM, Saham

PENDAHULUAN

Dalam dunia investasi yang terus berkembang, pencapaian hasil investasi yang optimal merupakan tujuan utama setiap investor. Pentingnya memprediksi harga saham dalam dunia investasi menjadi fokus berikutnya. Namun, Pergerakan harga saham sulit

untuk ditebak akan arah kelajuannya. Saat ini, perkembangan komputasi untuk pembelajaran mesin sangat pesat, yaitu machine learning dan deep learning. Dengan melihat permasalahan yang ada, maka tujuan penelitian adalah mendapatkan hasil perbandingan salah satu metode

machine learning, yaitu Extreme Gradient Boosting (XGBoost) dan salah satu metode deep learning, yaitu Long Short-Term Memory (LSTM) berupa tingkat nilai regression metrics dan nilai error, selanjutnya menentukan metode yang memiliki tingkat nilai regression metrics dan nilai error yang terbaik apabila diimplementasikan pada harga pembukaan saham [1].

Jurnal penelitian terkait yang dijadikan acuan dalam melakukan penelitian ini adalah penelitian yang dilakukan oleh Affan Ardana pada tahun 2023 berjudul Analisis Performa Algoritma XGBoost untuk Menentukan Parameter dan Fitur yang Paling Optimal dalam Memprediksi Pergerakan Harga Saham [1], hasil dari penelitian ini menunjukkan bahwa prediksi pergerakan harga saham menggunakan algoritma XGBoost memberikan hasil nilai RMSE terendah sebesar 0,009437, selanjutnya penelitian yang dilakukan Beno Jange pada tahun 2022 berjudul Prediksi Harga Saham Bank BCA Menggunakan XGBoost [2], hasil dari penelitian ini menunjukkan bahwa peramalan harga saham Bank BCA dengan metode XGBoost memperoleh nilai MAPE sebesar 4,01%, selanjutnya penelitian yang dilakukan Tita Lattifiaa, Putu Wira Buana, dan Ni Kadek Dwi Rusjayanthib pada tahun 2022 berjudul Model Prediksi Cuaca Menggunakan Metode LSTM [3], hasil dari penelitian ini menunjukkan bahwa prediksi cuaca menggunakan metode Long Short-Term Memory (LSTM) memperoleh nilai RMSE dan MAPE terbaik diperoleh yaitu 1,7444 dan 1,9499%, selanjutnya penelitian yang dilakukan Adhitio Satyo Bayangkari Karno pada tahun 2020 berjudul Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory) [4], hasil

prediksi data time series saham Bank BRI dengan menggunakan mesin belajar LSTM terlihat cukup berhasil untuk mendapatkan nilai akurasi tinggi RMSE sebesar 227,470, dan terakhir adalah penelitian yang dilakukan Prismahardi Aji Riyantoko, Tresna Maulana Fahrudin, Kartika Maulida Hindrayani, dan Eristya Maya Safitri pada tahun 2020 berjudul Analisis Prediksi Harga Saham Sektor Perbankan Menggunakan Algoritma LSTM (Long Short-Term Memory) [5], hasil prediksi data time series saham Bank Mandiri dengan menggunakan mesin belajar LSTM terlihat cukup berhasil untuk mendapatkan nilai akurasi tinggi RMSE sebesar 240,71.

Berdasarkan kelima jurnal penelitian terkait yang telah disebutkan, seluruhnya memiliki kekurangan, yaitu hanya menggunakan salah satu metode, yaitu XGBoost ataupun LSTM, sehingga tidak diketahui akurasi terbaik yang didapatkan dari metode time series lain dalam memprediksi harga saham. Oleh karena itu, pada penelitian ini akan dilakukan perbandingan metode XGBoost dengan LSTM untuk prediksi saham PT. Bank Mandiri Tbk. (BMRI), khususnya pada harga pembukaan saham.

METODE PENELITIAN

Metode penelitian yang digunakan dalam penyusunan penulisan ini meliputi 5 tahap, sebagai berikut:

1. Perencanaan

Tahap perencanaan penelitian ini adalah memperoleh bahan-bahan dan informasi yang dibutuhkan untuk penelitian, seperti buku, jurnal, dan artikel yang berkaitan dengan penelitian. Selain dari buku, jurnal, dan artikel, pengumpulan data berupa harga saham PT. Bank Mandiri Tbk. melalui

media internet sebagai koleksi dataset yang digunakan.

2. Analisis

Tahap analisis penelitian ini adalah mengamati secara detail kebutuhan yang digunakan untuk pembuatan perbandingan metode Long Short-Term Memory (LSTM) dan Extreme Gradient Boosting (XGBoost) termasuk menyiapkan dataset yang digunakan untuk melatih sistem dan efektifitas yang digunakan dalam sistem.

3. Perancangan

Tahap perancangan dalam penelitian ini adalah pembuatan alur kerja menggunakan flowchart secara umum, yaitu dimulai dari pengumpulan data, pra-pemrosesan data, pembuatan model, training model, validasi model, dan terakhir adalah evaluasi model.

4. Implementasi

Tahap implementasi dalam penelitian ini adalah menerapkan metode Long Short-Term Memory (LSTM) dan Extreme Gradient Boosting (XGBoost) dengan kode program menggunakan bahasa pemrograman Python.

5. Uji Coba

Tahap uji coba dalam penelitian ini adalah menggunakan metode black box, yaitu pengujian terhadap data, dari pengambilan sampai dengan menampilkan hasil grafik terhadap data yang diuji serta pengujian menggunakan dataset saham lain.

Penelitian ini memberikan gambaran umum tentang proses pengembangan model prediksi saham yang melibatkan beberapa tahapan penting. Tahap pertama adalah pengumpulan dataset yang akan digunakan untuk implementasi model. Dataset ini mencakup data saham yang

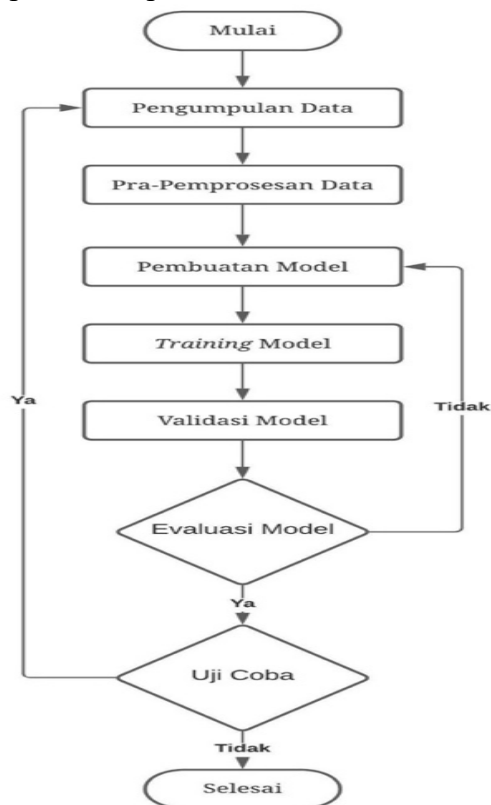
relevan dan akan digunakan sebagai bahan belajar bagi model.

Setelah dataset dikumpulkan, tahap kedua adalah pra-pemrosesan data. Pada tahap ini, data dibersihkan, dinormalisasi, dan ditransformasi untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam model. Tahap ketiga adalah perancangan model dengan metode Extreme Gradient Boosting (XGBoost) dan Long Short-Term Memory (LSTM). Proses ini melibatkan beberapa tahapan penting dan dimulai dengan penggunaan kumpulan data yang telah dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian.

Tahap keempat adalah pelatihan data. Pada tahap ini, model akan mempelajari fitur yang didapat dari dataset dan melakukan perubahan nilai parameter selama proses pelatihan berlangsung. Proses ini melibatkan penggunaan algoritma belajar mesin untuk menyesuaikan bobot dan bias dalam model berdasarkan data input dan output yang diharapkan. Tahap kelima adalah memprediksi model dengan data testing predict. Data testing predict kemudian diklasifikasikan, dan hasilnya dibandingkan dengan nilai parameter data testing real. Proses ini memungkinkan peneliti untuk mengevaluasi sejauh mana model dapat memprediksi data dengan benar dan mendapatkan tingkat akurasi dalam klasifikasi proses data testing predict dan data testing real.

Tahap terakhir, yaitu tahap keenam adalah evaluasi model pada dua metode, yaitu XGBoost dan LSTM dengan membandingkan regression metrics dan nilai error. Jika pada tahap evaluasi model didapatkan nilai regression metric dan nilai error yang tidak sesuai dengan yang diharapkan, maka dilakukan tahapan pembuatan

model kembali. Namun, jika pada tahap evaluasi model didapatkan nilai regression metrics dan nilai error yang diharapkan, maka model tersebut dapat diuji coba ataupun tidak dengan dataset saham lain. Jika dilakukan uji coba, maka akan mengulang ke tahapan pertama, yaitu pengumpulan dataset. Namun, jika tidak, maka tahapan penelitian dinyatakan selesai. Gambaran umum tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

HASIL DAN PEMBAHASAN PENELITIAN

Hasil

Seluruh dataset akan dibaca menggunakan bahasa pemrograman Python. Berikut adalah hasil dan pembahasan:

a. Prediksi Model LSTM

Perintah pertama yang dilakukan adalah membuat variabel predict yang berisi data array pada saat tahapan sebelumnya, yaitu pembuatan model.

Selanjutnya, membuat variabel, yaitu berisi dataframe hasil prediksi dan data real. Data target yang digunakan adalah data open price saham BMRI.

	Date	Real_Price	Predicted_Price
926	2023-06-20	5100.0	5075.747070
927	2023-06-21	5050.0	5074.275879
928	2023-06-22	5100.0	5037.886230
929	2023-06-23	5075.0	5072.092285
930	2023-06-26	5100.0	5056.202637

Gambar 2. Hasil Prediksi Metode LSTM

b. Prediksi Model XGBoost

Perintah pertama yang dilakukan adalah membuat variabel predict yang berisi data array pada saat tahapan sebelumnya, yaitu pembuatan model. Selanjutnya, membuat variabel, yaitu berisi dataframe hasil prediksi dan data real. Data target yang digunakan adalah data open price saham BMRI.

	Date	Real_Price	Predicted_Price
986	2023-06-19	5100.0	4561.919922
987	2023-06-20	5100.0	4470.040527
988	2023-06-21	5050.0	4465.299316
989	2023-06-22	5100.0	4497.688965
990	2023-06-23	5075.0	4390.514160

Gambar 3. Hasil Prediksi Metode XGBoost

c. Evaluasi Model

Tahapan selanjutnya adalah evaluasi model, dimana akan menghasilkan nilai Koefisien Determinasi (R^2) dan nilai Mean Absolute Percentage Error (MAPE). Koefisien Determinasi (R^2) digunakan untuk mengukur seberapa baik nilai prediksi dibandingkan dengan nilai real. Nilai koefisien determinasi adalah antara nol dan satu. Jika dalam uji empiris didapat nilai adjusted R^2 negatif, maka nilai adjusted R^2 dianggap nol. Semakin tinggi nilainya, semakin baik modelnya. Koefisien determinasi dapat dihitung dengan rumus:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Keterangan:

y_i = nilai *real*

\hat{y} = nilai prediksi

\bar{y} = nilai rata-rata

Mean Absolute Percentage Error (MAPE) adalah metode untuk menghitung perbedaan antara data asli dan data hasil prediksi. Perbedaan tersebut diabsolutkan, kemudian dihitung ke dalam bentuk persentase terhadap data asli. Hasil persentase tersebut kemudian didapatkan nilai mean-nya [6]. MAPE dapat dihitung dengan rumus:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}|}{y_i}$$

Keterangan:

n = jumlah data

y_i = nilai *real*

\hat{y} = nilai prediksi

Selanjutnya adalah evaluasi dari model LSTM dan model XGBoost.

1. Evaluasi Model LSTM

Pada program yang dibuat menghasilkan nilai Koefisien Determinasi (R^2) dan nilai *Mean Absolute Percentage Error* (MAPE) untuk model LSTM. Gambar 4 adalah hasil evaluasi model LSTM pada program.

Mean Absolute Percentage Error (MAPE): 1.7718551960356412%

Coefficient of Determination (R^2): 98.44831607458653%

Gambar 4. Hasil Evaluasi Model LSTM

2. Evaluasi Model XGBoost

Pada program yang dibuat menghasilkan nilai Koefisien Determinasi (R^2) dan nilai *Mean Absolute Percentage Error* (MAPE) untuk model XGBoost. Gambar 5 adalah hasil evaluasi model XGBoost pada program.

Mean Absolute Percentage Error (MAPE): 3.219962526293405%

Coefficient of Determination (R^2): 89.09446912563874%

Gambar 5. Hasil Evaluasi Model XGBoost

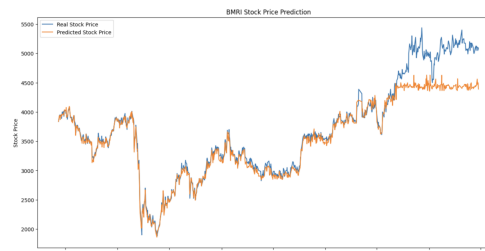
d. Visualisasi Hasil Prediksi LSTM



Gambar 6. Visualisasi Hasil Prediksi LSTM

Pada Gambar 6 merupakan visualisasi hasil prediksi LSTM, dimana garis biru merupakan nilai real data test dan garis oranye merupakan nilai predict, jika dilihat pada Gambar 6 garis biru dan garis oranye berhimpit mulai dari data pertama hingga data terakhir, ini menandakan bahwa hasil prediksi sudah sangat baik.

e. Visualisasi Hasil Prediksi XGBoost

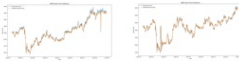
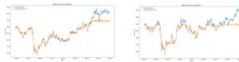


Gambar 7. Visualisasi Hasil Prediksi XGBoost

Pada Gambar 7 merupakan visualisasi hasil prediksi XGBoost, dimana garis biru merupakan nilai real data test dan garis oranye merupakan nilai predict, jika dilihat pada Gambar 7 garis biru dan garis oranye berhimpit mulai dari data pertama hingga data terakhir, namun jika dibandingkan dengan metode LSTM, garis biru dan oranye pada metode XGBoost tidak terlalu berhimpit, ini menandakan bahwa hasil prediksi sudah baik. Selajutnya dilakukan pengujian terhadap dataset lain. Perbandingan dengan menggunakan saham lain dilakukan agar mengetahui apakah program dapat dijalankan dengan saham yang berbeda. Pada Tabel 1 dijelaskan tentang hasil pengujian dan perbandingan

menggunakan data saham PT. Bank Rakyat Indonesia Tbk. (BBRI).

Tabel 1. Perbandingan Saham BMRI Dengan BBRI

Uji Coba	Hasil				Kesimpulan
	BMRI		BBRI		
Nilai R ² dan MAPE <i>Data Test</i> pada LSTM	R ²	: 98.44%	R ²	: 95.94%	BMRI lebih baik dari BBRI karena nilai R ² yang lebih tinggi dengan tingkat MAPE yang lebih rendah.
	MAPE	: 1.77%	MAPE	: 1.73%	
Nilai R ² dan MAPE <i>Data Test</i> pada <u>XGBoost</u>	R ²	: 89.09%	R ²	: 84.25%	BMRI lebih baik dari BBRI karena nilai R ² yang lebih tinggi dengan tingkat MAPE yang lebih rendah.
	MAPE	: 3.21%	MAPE	: 2.87%	
Perbandingan Data Prediksi dengan <i>Data Test</i> pada LSTM					Hasil perbandingan <i>Data Test</i> dengan Data Prediksi BMRI lebih baik dibanding dengan BBRI karena garis <i>real</i> dan <i>predict</i> saling berhimpit.
Perbandingan Data Prediksi dengan <i>Data Test</i> pada <u>XGBoost</u>					Hasil perbandingan <i>Data Test</i> dengan Data Prediksi BMRI lebih baik dibanding dengan BBRI karena garis <i>real</i> dan <i>predict</i> saling berhimpit.

Keterangan: R² adalah Koefisien Determinasi, MAPE adalah Mean Absolute Percentage Error

PENUTUP

Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dalam penelitian ini dapat diambil kesimpulan, yaitu dari 4.960 data open price saham PT. Bank Mandiri (BMRI) pada tahun 2003 s.d. 2023, disimpulkan bahwa metode Extreme Gradient Boosting (XGBoost) dan Long Short-Term Memory (LSTM) dapat digunakan dalam memprediksi dataset saham tersebut. Hasil prediksi, yaitu berupa nilai Koefisien Determinasi (R²) dan Mean Absolute Percentage Error (MAPE). Setelah dilakukan pengujian didapatkan hasil dari metode XGBoost, yaitu nilai R² sebesar 89.09 % dan MAPE sebesar 3.21 %, sedangkan pada metode LSTM, yaitu nilai R² sebesar 98.44 % dan MAPE sebesar 1.77 %. Berdasarkan nilai Koefisien

Determinasi (R²) dan nilai error MAPE yang didapatkan pada kedua metode, dapat dikatakan bahwa metode Long Short-Term Memory (LSTM) merupakan metode terbaik dalam memprediksi dataset saham. Berdasarkan kesimpulan yang telah diuraikan, maka ada beberapa saran yang dapat disampaikan, yaitu Jumlah dataset saham yang dianalisis ditingkatkan agar hasil prediksi semakin baik dan pada penelitian selanjutnya dapat menggunakan metode time series lainnya agar mendapatkan perbandingan metode yang lebih baik dalam memprediksi saham.

DAFTAR PUSTAKA

- [A. Ardana, "Performance Analysis of XGBoost Algorithm to Determine the Most Optimal Parameters and Features in Predicting Stock Price Movement," *Telematika: Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 1, pp. 91–102, 2023.
- B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan Prophet," *Journal of Trends Economics and Accounting Research*, vol. 2, no. 1, pp. 1–5, 2021.
- T. Lattifia, P. W. Buana, and N. K. D. Rusjayanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 994–1000, 2022.
- A. S. B. Karno, "Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory)," *Journal of Informatic and Information Security*, vol. 1, no. 1, 2020.
- P. A. Riyantoko, T. M. Fahrudin, K. M. Hindrayani, and E. M. Safitri, "Analisis Prediksi Harga Saham Sektor Perbankan Menggunakan Algoritma Long-Short Terms Memory (Lstm)," in *Seminar Nasional Informatika (SEMNASIF)*, 2020, pp. 427–435.