

Prediksi Harga Saham Bank BCA Menggunakan XGBoost

Beno Jange

Program Studi Komputerasi Akuntansi, STMIK Dharmapala Riau, Pekanbaru

Jl. KH Samanhudi No.13, Sago, Kec. Senapelan, Kota Pekanbaru, Riau

Email: beno.jange@lecturer.stmikdharmapalariau.ac.id

Email Penulis Korespondensi: beno.jange@lecturer.stmikdharmapalariau.ac.id

Abstrak—Penelitian ini bertujuan untuk mengetahui Prediksi Harga Saham Bank BCA menggunakan XGBoost. XGBoost adalah implementasi sumber terbuka dari Gradient Boosting untuk menghasilkan prakiraan berdasarkan data historis yang cepat dan dapat diskalakan. Data dalam penelitian ini adalah data harga saham Bank BCA selama 4 (empat) tahun yaitu dari tanggal 01-01-2017 sampai 31-12-2020. Indikator teknikal yang digunakan dalam penelitian ini adalah Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD) dan Relative Strength Index (RSI). Hasil penelitian menunjukkan bahwa indikator teknikal Exponential Moving Average sangat mempengaruhi hasil prediksi. Hasil penelitian ini juga menunjukkan akurasi prediksi yang cukup baik dengan MAPE sebesar 4.01 persen dengan penyetulan hiper parameter; prediksi sedikit kurang baik pada bulan Maret tahun 2020 karena ada kasus pandemi Covid-19.

Kata Kunci: Prediksi; Harga Saham; XGBoost; Indikator Teknikal; Rekayasa Fitur

Abstract—This study aims to determine the prediction of Bank BCA's stock price using XGBoost. XGBoost is an open source implementation of Gradient Boosting for generating forecasts based on historical data that is fast and scalable. The data in this study is stock price data of Bank BCA for 4 (four) years, namely from 01-01-2017 to 31-12-2020. The technical indicators used in this study are the Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD) and Relative Strength Index (RSI). The results show that the Exponential Moving Average technical indicator greatly influences the prediction results. The results of this study also show a fairly good prediction accuracy with MAPE of 4.01 percent with hyper parameter settings; but the predictions are slightly less good in March 2020 due to the Covid-19 pandemic case.

Keywords: Prediction; Stock Price; XGBoost; Technical Indicator; Feature Engineering

1. PENDAHULUAN

Eugene Fama menemukan bahwa ada banyak bukti yang mendukung Hipotesis Pasar Efisien (Efficient Market Hypothesis), yang menunjukkan bahwa harga sekuritas “sebenarnya mencerminkan” informasi yang relevan pada waktu tertentu: Oleh karena itu, baik analisis teknikal maupun fundamental tidak efektif dalam mencari abnormal return secara konsisten di bawah kondisi pasar yang efisien (Malkiel, 2003).

Namun pada kenyataannya, investor memperoleh sejumlah pengembalian tertentu dengan menerapkan berbagai metode prediksi, biasanya analisis teknikal, yang diterapkan di berbagai pasar aset (Stankovic et al, 2015). Oriani & Coelho (2016) menemukan bahwa indikator teknikal bila digunakan sebagai input dapat meningkatkan keakuratan ramalan saham dan kombinasi berbagai indikator sebagai input dapat lebih meningkatkan kinerja peramalan.

Investasi di pasar saham dianggap memiliki risiko yang tinggi dengan keuntungan yang tinggi pula sehingga menarik perhatian banyak investor dan ekonom. Namun, informasi mengenai saham biasanya tidak lengkap, rumit, tidak pasti dan tidak jelas, sehingga menjadikan tantangan untuk memprediksi kinerja ekonomi di masa depan. Analisis data deret waktu membantu dalam mengidentifikasi pola, tren, dan periode atau siklus yang ada dalam data di pasar saham (Jange, 2021; 2022).

Hal ini membuat analisis deret waktu dan peramalan menjadi area penelitian yang penting. Evolusi berkelanjutan di bidang Pembelajaran Mesin dan Kecerdasan Buatan dan fakta bahwa informasi pasar keuangan menjadi lebih mudah diakses oleh sejumlah besar investor menghasilkan munculnya algoritma perdagangan canggih yang akibatnya mulai memiliki pengaruh signifikan terhadap perilaku pasar (Nobre & Neves, 2019).

Extreme Gradient Boosting, atau singkatnya XGBoost adalah implementasi sumber terbuka yang efisien dari algoritma peningkatan gradien (Gradient Boosting) yang dikembangkan oleh Tianqi Chen dan Carlos Guestrin. Faktor terpenting di balik kesuksesan XGBoost adalah skalabilitasnya dan kecepatannya di semua skenario. Sistem berjalan sepuluh kali lebih cepat daripada solusi populer yang ada pada satu mesin dan menskalakan hingga miliaran contoh dalam pengaturan terdistribusi atau terbatas memori.

XGBoost mendominasi kumpulan data terstruktur atau tabular pada masalah pemodelan prediktif klasifikasi dan regresi yang dibuktikan bahwa algoritma ini selalu digunakan oleh pemenang kompetisi di platform ilmu data kompetitif Kaggle (Chen & Guestrin, 2016). Meskipun XGBoost dirancang untuk klasifikasi dan regresi pada kumpulan data tabular, namun dapat juga digunakan untuk peramalan deret waktu.

Sudah banyak penelitian tentang prediksi pasar saham menggunakan XGBoost seperti pada pasar saham di Cina (Chen et al, 2021; Long et al, 2019; Yun et al, 2021), Amerika Serikat (Beyaz et al, 2018; Nobre & Neves, 2019), Iran (Nabipour et al, 2020) dan Korea Selatan (Kim, 2018). Penelitian ini bertujuan untuk memprediksi harga saham Bank BCA dengan menggunakan XGBoost dan indikator teknikal.

2. METODE PENELITIAN

2.1 XGBoost dan Algoritma Gradient Boosting

Dalam melakukan prediksi harga saham sebagai studi kasus penelitian ini, digunakan pustaka XGBoost yang merupakan implementasi dari algoritma Gradient Boosting yang dirancang untuk kecepatan komputasi dan kinerja model (Chen & Guestrin, 2016). Algoritma Gradient Boosting menggabungkan pembelajar lemah (weak learner) menjadi satu pembelajar kuat (strong learner) tunggal secara iteratif yang dapat dijelaskan sebagai berikut (Friedman, 2001):

a. Masukan: training set $\{(x_i, y_i)\}_{i=1}^n$, fungsi differential loss $L(y, F(x))$, sejumlah iterasi M .

b. Proses:

1. Inisialisasi model dengan nilai tetap:

$$F_0(x) = \arg \min \sum_{i=1}^n L(y_i, \gamma) \quad (1)$$

2. Untuk $m = 1$ sampai M :

a) Hitung pseudo-residual:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right] \quad (2)$$

b) Sesuaikan base learner (atau weak learner, misalnya tree) ditutup di bawah penskalaan $h_m(x)$ ke pseudo-residual, yaitu melatihnya menggunakan training set $\{(x_i, r_{im})\}_{i=1}^n$.

c) Hitung pengali γ_m dengan memecahkan masalah optimisasi satu dimensi berikut:

$$\gamma_m = \arg \min \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h(x_i)) \quad (3)$$

d) Perbarui modelnya

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (4)$$

c. Keluaran: $F_M(x)$.

2.2 Indikator Teknikal

Dalam analisis teknikal di bidang keuangan, indikator teknikal adalah perhitungan matematis berdasarkan harga historis, volume, atau informasi lainnya yang bertujuan untuk meramalkan arah harga pada pasar keuangan di mana yang digunakan pada penelitian ini adalah (Taran-Morosan, 2009):

a. Simple Moving Average (SMA)

$$SMA_k = \frac{P_{n-k+1} + P_{n-k+2} + \dots + P_n}{k} \quad (5)$$

$$SMA_k = \frac{1}{k} \sum_{i=n-k+1}^n P_i \quad (6)$$

Dengan P sebagai harga, k sebagai titik data dan n sebagai jumlah data.

b. Exponential Moving Average (EMA)

$$EMA_k = \alpha x_k + (1 - \alpha) S_{k-1} \quad (7)$$

$$EMA_k = S_{k-1} + \alpha(x_k - S_{k-1}) \quad (8)$$

Dengan α sebagai faktor pemulusan (smoothing factor) dan $0 \leq \alpha \leq 1$, S_k adalah rata-rata tertimbang sederhana dari pengamatan saat ini x_k dan statistik sebelumnya S_{k-1} .

c. Moving Average Convergence/Divergence (MACD)

$$MACD = EMA_{12} - EMA_{26} \quad (9)$$

$$Signal = EMA_9(MACD) \quad (10)$$

d. Relative Strength Index (RSI)

$$RS = \frac{EMA(U,n)}{EMA(D,n)} \quad (11)$$

$$RSI = 100 - \frac{100}{1+RS} \quad (12)$$

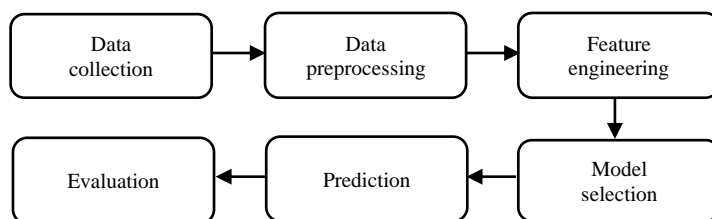
Dengan U sebagai periode kenaikan, D sebagai periode penurunan, RS sebagai rasio rata-rata yang disebut relative strength factor dan RSI sebagai konversi dari RS menjadi indeks antara 0 dan 100.

2.3 Parameter hiper pada XGBoost

Berikut beberapa parameter hiper (hyperparameter) yang harus dipertimbangkan untuk menyetel ansambel Gradient Boosting dan pengaruhnya terhadap kinerja model (Chen & Guestrin, 2016):

- a. γ yaitu pengurangan kerugian minimum yang diperlukan untuk membuat partisi lebih lanjut pada simpul daun pohon. Semakin besar γ , semakin konservatif algoritmanya.
- b. $learning_rate$ yaitu penyusutan ukuran langkah digunakan dalam pembaruan untuk mencegah overfitting. Setelah setiap langkah boosting, bisa langsung didapatkan bobot fitur baru, dan $learning_rate$ mengecilkan bobot fitur agar proses boosting lebih konservatif.
- c. max_depth yaitu kedalaman maksimum pohon. Meningkatkan nilai ini akan membuat model lebih kompleks dan lebih cenderung overfit. 0 menunjukkan tidak ada batas kedalaman. Berhati-hatilah karena XGBoost secara agresif menghabiskan memori saat melatih pohon yang dalam.
- d. $n_estimators$ yaitu jumlah pohon keputusan. Ingatlah bahwa pohon keputusan ditambahkan ke model secara berurutan dalam upaya untuk mengoreksi dan memperbaiki prediksi yang dibuat oleh pohon sebelumnya. Dengan demikian, lebih banyak pohon seringkali lebih baik.

2.4 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Model XGBoost untuk memprediksi harga saham Bank BCA dilakukan dalam 6 (enam) tahap sebagai berikut:

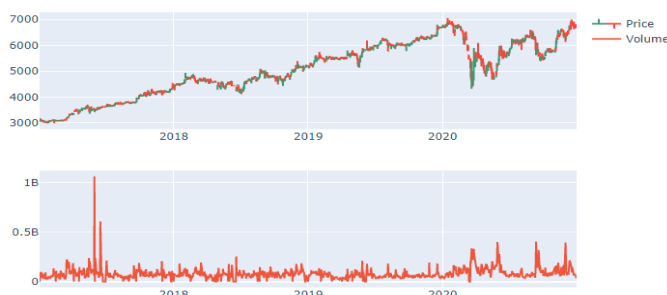
- a. Pengumpulan data (data collection) yaitu mengumpulkan data harga saham Bank BCA dari tahun 2017 sampai 2020 yang diunduh dari situs finance.yahoo.com dan memvisualisasikannya menggunakan pustaka matplotlib dan plotly.
- b. Prapemrosesan data (data preprocessing) yaitu mengubah data menjadi format yang lebih mudah dan efektif untuk diproses selanjutnya menggunakan pustaka pandas dan memvisualisasikannya menggunakan pustaka matplotlib dan plotly.
- c. Rekayasa fitur (feature engineering) yaitu membuat fitur relevan yaitu indikator teknikal yang digunakan dalam kumpulan data (data set) menggunakan pustaka pandas dan memvisualisasikannya menggunakan pustaka matplotlib dan plotly.
- d. Pemilihan model (model selection) yaitu memasukkan data ke dalam model XGBoost dengan cara memecah data menjadi data latih (train) dan data uji (test) sekaligus juga memasukkan parameter hiper untuk menentukan hasil terbaik menggunakan fungsi XGBRegressor pada pustaka XGBoost.
- e. Prediksi (prediction) yaitu memprediksi data menggunakan fungsi predict pada pustaka XGBoost dan memvisualisasikannya menggunakan pustaka matplotlib dan plotly.
- e. Evaluasi (evaluation) yaitu menentukan fitur indikator teknikal yang paling berperan penting dalam pemodelan (feature importance) dan mengukur seberapa akurat hasil prediksi yang dibuat oleh model (performance evaluation) menggunakan Mean Absolute Percentage Error (MAPE) dengan persamaan:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \tag{13}$$

dimana A_t adalah nilai aktual (actual value) pada waktu t sedangkan F_t adalah nilai prediksi (forecast value) pada waktu t . Nilai aktual dikurangi dengan nilai prediksi kemudian dibagi dengan nilai aktual dan dibuat menjadi nilai absolut. Hasil kemudian divisualisasikan menggunakan pustaka matplotlib dan plotly.

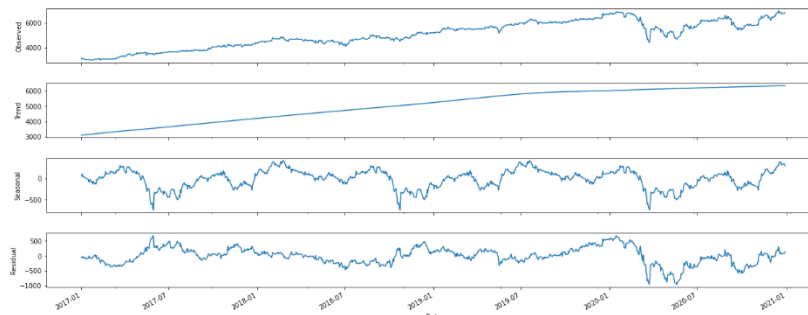
3. HASIL DAN PEMBAHASAN

Data yang diambil adalah data harga saham harian (open, high, low, close dan adj close) dan volume selama periode 01/01/2017 hingga 31/12/2020 yang ditunjukkan pada gambar berikut:



Gambar 2. Grafik Harga Penutupan Saham BBCA periode 2017 sampai 2020

Gambar 3 mengilustrasikan harga penutupan harian dari satu set saham dan dekomposisi penutupan harian saham dengan mengambil fungsi dekomposisi, menghasilkan plot tren, musiman, dan residu dari harga saham. Dapat dilihat bahwa tren kelompok harga saham ini mula-mula naik dan kemudian cenderung turun, yang konsisten dengan aturan deret waktu, dan harga saham memiliki musiman yang jelas dan residunya cenderung halus. Oleh karena itu, kumpulan data saham ini cocok digunakan untuk pelatihan model prediksi harga.



Gambar 3. Grafik Dekomposisi (Observed, Trend, Seasonal, Residual) Harga Saham BCA Periode 2017 sampai 2020

Sebelum melakukan prediksi model, perlu dilakukan rekayasa fitur menggunakan indikator teknikal seperti yang ditunjukkan pada tabel 1.

Tabel 1. Indikator Teknikal untuk Rekayasa Fitur

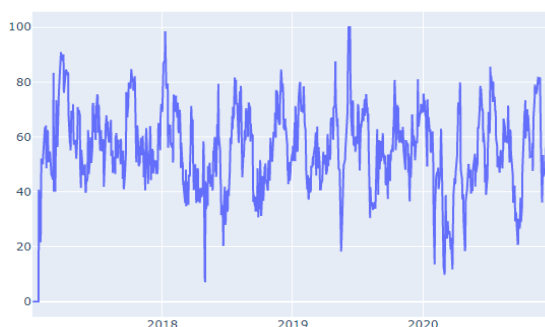
Indikator Teknikal	Deskripsi
EMA_9	Data 9 hari harga saham rata-rata bergerak tertimbang
SMA_5	Data 5 hari pergerakan harga saham rata-rata jangka pendek
SMA_15	Data 15 hari pergerakan harga saham rata-rata jangka pendek
SMA_30	Data 30 hari pergerakan harga saham rata-rata jangka pendek
RSI	$SMA(MAX(Close-LastClose,0),N,1)/SMA(ABS(Close-LastClose),N,1)*100$
MACD	EMA12-EMA26
MACD Signal	Rata-rata pergerakan bobot indeks EWM untuk MACD dengan rentang waktu 9 hari

Berikut adalah visualisasi dari rekayasa fitur menggunakan indikator teknikal EMA (Exponential Moving Average) dan SMA (Simple Moving Average) terhadap harga saham Bank BCA periode 2017 sampai 2020 seperti yang ditunjukkan pada gambar 4.



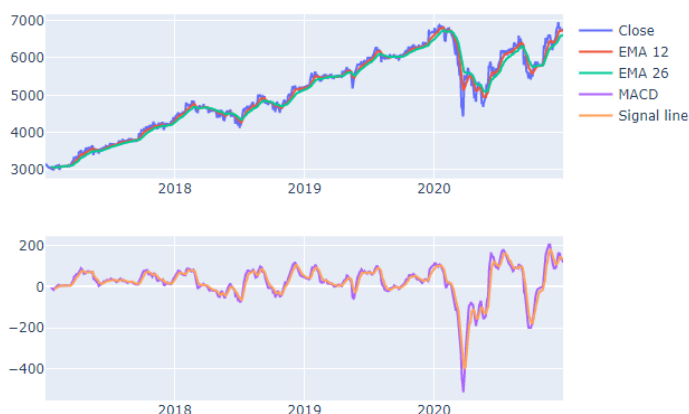
Gambar 4. Grafik Indikator Teknikal (EMA dan SMA) Harga Saham Bank BCA periode 2017 sampai 2020

Berikut adalah visualisasi rekayasa fitur menggunakan indikator teknikal RSI (Relative Strength Index) terhadap harga saham Bank BCA periode 2017 sampai 2020 seperti yang ditunjukkan pada gambar 5.



Gambar 5. Grafik Indikator Teknikal RSI Harga Saham Bank BCA periode 2017 sampai 2020

Berikut visualisasi dari rekayasa fitur menggunakan indikator teknikal MACD (Moving Average Convergence Divergence) terhadap harga saham Bank BCA periode 2017 sampai 2020 seperti yang ditunjukkan pada gambar 6.



Gambar 6. Grafik Indikator Teknikal MACD Harga Saham Bank BCA periode 2017 sampai 2020

Data selama 3 (tiga) tahun yaitu periode 2017 sampai 2019 digunakan sebagai data latih (train) untuk melatih sistem untuk menghasilkan semua model prediksi dan data 1 (satu) tahun terakhir yaitu periode 2020 digunakan sebagai data uji (test) untuk menguji model akhir yang diperoleh setelah semua pelatihan dilakukan yang ditunjukkan pada gambar 7.



Gambar 7. Grafik Pemecahan Data (Train, Test) Harga Saham Bank BCA periode 2017 sampai 2020

Ketika menggunakan algoritma XGBoost, perlu untuk dicari parameter hiper sesuai dengan karakteristik dari harga saham. Tugas penyetelan parameter hiper diserahkan pada Grid SearchCV yang akan mencoba setiap parameter yang mungkin di antara semua kandidat, dan parameter dengan kinerja terbaik adalah hasil akhirnya seperti ditunjukkan pada tabel 3. Kerugian utama dari penggunaan Grid SearchCV adalah lebih memakan waktu, tetapi menjamin keakuratan dan validitas hasil (Zhang, 2022).

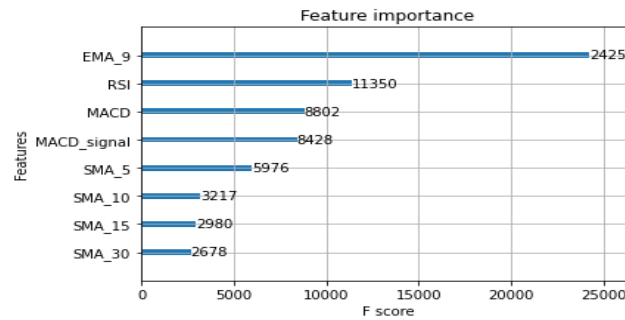
Tabel 2. Pengaturan Pencarian Parameter Hiper

Parameter Hiper	Nilai
n_estimators	[100, 200, 300, 400]
learning_rate	[0.001, 0.005, 0.01, 0.05]
max_depth	[8, 10, 12, 15]
gamma	[0.001, 0.005, 0.01, 0.02]
random_state	[42]

Tabel 3. Hasil Pencarian Parameter Hiper Terbaik

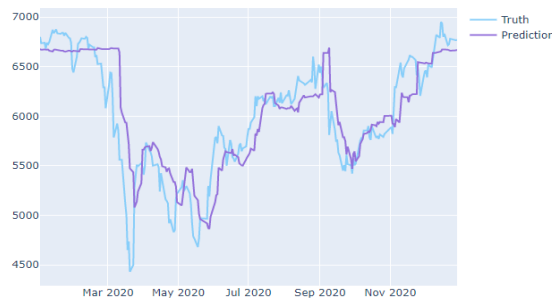
Parameter Hiper	Nilai
n_estimators	400
learning_rate	0.05
max_depth	15
gamma	0.001
random_state	42

Berikut adalah hasil dari penilaian seberapa penting fitur indikator teknikal dalam rekayasa fitur digunakan untuk memprediksi harga saham Bank BCA tahun 2020 seperti yang ditunjukkan pada gambar 8 di mana terlihat bahwa EMA_9 memiliki peran paling penting dengan nilai yang diperoleh sangat tinggi melampaui indikator teknikal lainnya (RSI, MACD, sinyal, SMA).



Gambar 8. Grafik Feature Importance Prediksi Harga Saham BBCA Tahun 2020

Berikut adalah hasil dari pemuatan (fit) dan prediksi (predict) data harga saham Bank BCA periode 2020 seperti yang ditunjukkan pada gambar 9. Pada gambar terlihat bahwa prediksi terlihat cukup akurat setelah melalui proses pengaturan parameter hiper terbaik dengan 'gamma': 0.001, 'learning_rate': 0.05, 'max_depth': 15, 'n_estimators': 400 dan diperoleh nilai MAPE sebesar 4.01 persen. Prediksi sedikit kurang baik pada bulan Maret tahun 2020 karena ada kasus pandemi Covid-19.



Gambar 9. Grafik Prediksi Harga Saham BBCA periode 2020

Jika dibandingkan dengan penelitian sebelumnya (Jange, 2021) yang memprediksi menggunakan Prophet dapat dilihat bahwa hasil prediksi XGBoost berhasil melampaui prediksi Prophet. Prediksi XGBoost dengan nilai MAPE sebesar 4.01 persen berhasil melampaui prediksi Prophet dengan nilai MAPE sebesar 5.37 persen. Meski begitu, ada banyak faktor yang mempengaruhi perbedaan tingkat akurasi prediksi untuk masing-masing algoritma. Salah satu faktornya adalah banyaknya parameter hiper dan variasi nilai parameter hiper yang digunakan. Faktor lainnya adalah prediksi XGBoost menggunakan indikator teknikal dalam rekayasa fiturnya sedangkan prediksi XGBoost pada penelitian sebelumnya tidak menggunakannya sehingga prapemrosesan datanya menggunakan data asli yaitu harga penutupan saham.

4. KESIMPULAN

Berdasarkan hasil pembahasan yang telah dilakukan maka dapat disimpulkan bahwa metode XGBoost cukup baik dalam memprediksi harga saham Bank BCA. Melalui beberapa penyetelan parameter hiper pada metode XGBoost akan didapatkan prediksi yang lebih baik yaitu MAPE sebesar 4.01 persen. Dapat dilihat bahwa pemakaian indikator teknikal dalam rekayasa fitur dan penyetelan parameter hiper cukup signifikan di dalam menurunkan nilai MAPE. Namun perlu diingat bahwa semakin banyak indikator teknikal yang digunakan dalam rekayasa fitur dan jumlah parameter yang digunakan dalam seleksi parameter hiper maka waktu untuk memprosesnya menjadi semakin lama. Jika dibandingkan dengan penelitian sebelumnya menggunakan Prophet bisa dilihat bahwa prediksi XGBoost berhasil melampaui prediksi Prophet namun ada banyak faktor penyebabnya yaitu jumlah parameter yang digunakan dan bahwa Prophet tidak menggunakan indikator teknikal dalam melakukan rekayasa fitur. Untuk penelitian selanjutnya disarankan untuk menggunakan data historis harga saham yang lebih panjang, menggunakan lebih banyak parameter hiper; membandingkan metode XGBoost dengan metode peramalan saham lainnya (seperti ARIMA, Regresi Linear, LSTM, KNN); menggunakan lebih banyak saham untuk dibandingkan; memecah data yang sebelumnya data latih (train) dan uji (test) menjadi data latih (train), validasi (validation) dan uji (test).

REFERENCES

- Beyaz, Erhan & Tekiner, Firat & Zeng, Xiao-jun & Keane, John A. (2018). Stock price forecasting incorporating market state. *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 1614-1619.
- Chen, Tianqi & Guestrin, Carlos (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785-794.

- Friedman, Jerome H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Jange, Beno. (2021). Prediksi Harga Saham Bank BCA Menggunakan Prophet. *Journal of Trends Economics and Accounting Research*, 2(1), 1-5.
- Jange, Beno. (2022). Prediksi Indeks Harga Saham Gabungan Menggunakan Prophet. *JOTIKA Journal In Management and Entrepreneurship*, 1(2), 53-59.
- Kim, HyoungDo (2018). Prediction of the Movement Directions of Index and Stock Prices Using Extreme Gradient Boosting. *The Journal of the Korea Contents Association*, 18(9), 623-632.
- Malkiel, Burton G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1), 59-82.
- Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (2020). Deep learning for stock market prediction. *Entropy*, 22(8), 840.
- Nobre, João & Neves, Rui Ferreira (2019). Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. *Expert Systems with Applications*, 125, 181-194.
- Oriani, Felipe Barboza & Coelho, Guilherme P. (2016). Evaluating the impact of technical indicators on stock forecasting. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE. 1-8.
- Stanković, Jelena & Marković, Ivana & Stojanović, Miloš (2015). Investment Strategy Optimization Using Technical Analysis and Predictive Modeling in Emerging Markets. *Procedia Economics and Finance*, 19, 51–62.
- Taran-Morosan, Adrian (2009). Some Technical Analysis Indicators. *Revista Economica*, 46(3), 116-121.
- Yun, Kyung Keun & Yoon, Sang Won & Won, Daehan (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, 186, 115716.
- Zhang, Yifan (2022). Stock Price Prediction Method Based on XGBoost Algorithm. *2022 International Conference on Big Data Blockchain and Economy Management (ICBBEM 2022)*, 595-603.