

# Prediksi Return Saham Index LQ45 Menggunakan Metode *Extreme Gradient Boosting* (XGBOOST)

Nesa Tita Nurlela<sup>1</sup>, Fajri Rakhmat Umbara<sup>2</sup>, Fatan Kasyidi<sup>3</sup>

Program Studi Informatika

Universitas Jenderal Achmad Yani

Kota Cimahi, Indonesia

e-mail: <sup>1</sup>nesatitan20@if.unjani.ac.id, <sup>2</sup>fajri.rakhmat@lecturer.unjani.ac.id,

<sup>3</sup>fatan.kasyidi@lecture.unjani.ac.id

Correspondence : e-mail: [nesatitan20@if.unjani.ac.id](mailto:nesatitan20@if.unjani.ac.id)

Diajukan: 13 Agustus 2024; Direvisi: 22 Agustus 2024; Diterima: 23 Agustus 2024

## Abstrak

Penelitian ini bertujuan untuk menguji kinerja algoritma *Extreme Gradient Boosting* (XGBoost) dalam memprediksi pengembalian saham, dengan penekanan khusus pada mengatasi masalah ketidakseimbangan data. Penelitian ini melihat dua metode untuk mengatasi ketidakseimbangan data SMOTE (Teknik *Oversampling Minoritas Sintetis*) dan *Random Oversampling*. Dengan menggunakan 12.572 data dari tujuh dataset saham dan faktor ekonomi makro, seperti nilai tukar Rupiah terhadap dolar AS dalam periode 29 Juni 2019 – 30 Juni 2024. Metode eksperimen membagi data menjadi 80% untuk data training dan 20% untuk data testing. Sebelum data diolah lebih lanjut, *preprocessing* dilakukan untuk membersihkan dan menyiapkannya. Untuk mengatasi ketidakseimbangan data, metode seperti *Random Oversampling* dan SMOTE digunakan. Pengujian dilakukan dengan mengoptimalkan *hyperparameter* menggunakan *Estimator Parzen Tree* untuk meningkatkan kinerja model. Hasil evaluasi dengan metrik *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) menunjukkan bahwa SMOTE lebih baik daripada *Random Oversampling*. Hasil akhir menunjukkan bahwa SMOTE menghasilkan RMSE sebesar 0,0398 dan MAE sebesar 0,0305, sedangkan *Random Oversampling* menghasilkan RMSE sebesar 0,0402 dan MAE sebesar 0,0310 menunjukkan bahwa SMOTE memiliki tingkat kesalahan yang lebih rendah dalam prediksi.

**Kata kunci:** Faktor Ekonomi, XGBoost, SMOTE, *Random Oversampling*.

## Abstract

This research aims to test the performance of the *Extreme Gradient Boosting* (XGBoost) algorithm in predicting stock returns, with a particular emphasis on addressing *Balancing Data* issues. The study examines two methods for addressing *Balancing Data*: SMOTE (*Synthetic Minority Oversampling Technique*) and *Random Oversampling*. Utilizing 12,572 data points from seven stock datasets and macroeconomic factors such as the exchange rate of the Rupiah against the US Dollar during the period from June 29, 2019, to June 30, 2024, the experimental method divides the data into 80% training data and 20% testing data. Before further processing, *preprocessing* is conducted to clean and prepare the data. To address *Balancing Data*, methods such as *Random Oversampling* and SMOTE are employed. Testing is conducted by optimizing *hyperparameters* using the *Tree Parzen Estimator* to enhance model performance. Evaluation results with metrics *Root Mean Squared Error* (RMSE) and *Mean Absolute Error* (MAE) show that SMOTE outperforms *Random Oversampling*. The final results indicate that SMOTE yields an RMSE of 0.0398 and an MAE of 0.0305, while *Random Oversampling* yields an RMSE of 0.0402 and an MAE of 0.0310, demonstrating that SMOTE has a lower error rate in predictions.

**Keywords:** Economic Factors, XGBoost, SMOTE, *Random Oversampling*.

## 1. Pendahuluan

Return saham (pengembalian saham) adalah keuntungan yang diterima investor sebagai kompensasi atas investasi mereka dalam saham. Return ini dihitung dari selisih antara harga jual dan harga beli saham, yang juga disesuaikan dengan dividen yang diterima. Saham memiliki potensi memberikan tingkat pengembalian yang menarik, sehingga banyak investor memilihnya sebagai instrumen investasi.

Saham juga memiliki keunggulan karena dapat diperjualbelikan, memungkinkan investor untuk memperoleh capital gain dan dividen[1]. Salah satu ukuran yang memiliki kekuatan untuk mempengaruhi perekonomian suatu negara adalah investasi[2]. Seorang investor harus mengetahui harga saham sebelum membeli saham. Seorang investor perlu memahami elemen-elemen yang memengaruhi fluktuasi harga saham[3].

Penelitian terdahulu telah mengeksplorasi penggunaan *Cross-Sectional Machine Learning* dalam prediksi return saham, menunjukkan bahwa pendekatan ini memiliki performa yang unggul dibandingkan dengan metode *time-series*. Sebagai contoh, *Mean Return* yang diperoleh dari pembelajaran *cross-sectional* mencapai 0.0032, dengan STD 0.0213, dan *Sharpe Ratio* 0.1540, yang lebih baik daripada hasil *time-series* dengan *Mean Return* 0.0026, STD 0.0196, dan *Sharpe Ratio* 0.1356[4]. Selain itu, penelitian lain menemukan bahwa prediksi return saham tiga hari sebelumnya menghasilkan kesalahan kuadrat terendah dan prediksi yang paling akurat, dengan fitur yang digunakan meliputi tanggal, harga pembukaan, tertinggi, terendah, penutupan, dan volume perdagangan[5].

Adapun penelitian terdahulu [6] menggunakan metode Xgboost penelitian ini menyajikan hasil temuan dari model beberapa model diantaranya XGBoost dan LSTM. Menunjukkan bahwa nilai RMSE LSTM 6,77%, nilai MAE LSTM 3,14%, dan nilai dari RMSE Xgboost 8,59 % dan nilai MAE Xgboost 13,29% dimana nilai dari Xgboost lebih baik dari yang lainnya. Penelitian [7] dimana penelitian ini membandingkan beberapa metode dan hasil prediksi BO-XGBoost yang disajikan dalam penelitian ini mencapai nilai metrik F1 93,35%, sensitivitas 90,67%, FRS 80,40%, dan AUC 0,920 berdasarkan temuan eksperimen. Jika dibandingkan dengan algoritma model lain yang menggunakan indikator yang sama, angka-angka ini adalah yang tertinggi. Penelitian ini [8] yang memprediksi harga emas menggunakan pendekatan XGBoost, peningkatan yang cukup besar dengan nilai RMSE 8,32% ditunjukkan pada hasilnya. Hal ini menunjukkan bahwa, dibandingkan dengan metode alternatif, metode XGBoost menawarkan tingkat akurasi dan presisi yang lebih tinggi dalam hal analisis dan prediksi harga emas.

Berdasarkan latar belakang ini, penelitian ini menggunakan metode *Extreme Gradient Boosting* (XGBoost) untuk memprediksi return saham dengan mempertimbangkan fitur-fitur penting seperti tanggal, harga pembukaan, harga tertinggi, harga terendah, harga penutupan, dan volume perdagangan.

## 2. Metode Penelitian

Penelitian yang digunakan dalam Prediksi Return Saham ini menggunakan beberapa tahapan, langkah-langkah tersebut dijelaskan pada Gambar 1.



Gambar 1 Metode Penelitian

### 2.1. Pengumpulan Data

Data penelitian ini diperoleh dari Kaggle dan mencakup data dari Bursa Efek Indonesia, khususnya dari indeks LQ45. Indeks LQ45 terdiri dari 45 perusahaan yang terdaftar di Bursa Efek Indonesia (BEI). Data saham yang digunakan mencakup 7 saham yaitu saham BBKA, BBNI, BBRI, BMRI, TLKM, UNTR, dan UNVR dengan total 8358 data, yang dikumpulkan dari periode 29 Juni 2019 hingga 30 Juni 2024. Selain data saham, penelitian ini juga menggunakan data faktor ekonomi berupa nilai tukar Rupiah terhadap USD, yang diperoleh dari situs Bank Indonesia dengan total 1195 data untuk periode yang sama.

### 2.2. Preprocessing

Tahap *preprocessing* merupakan langkah krusial untuk mempersiapkan data agar sesuai untuk analisis dan prediksi. Prosedur *preprocessing* meliputi:

#### 1. Data Cleaning

Pada tahap ini, data mentah dibersihkan dari noise dan variabel yang tidak relevan. Hal ini mencakup penghapusan titik, huruf, atau simbol yang tidak diperlukan dari data saham dan faktor ekonomi makro. Proses ini bertujuan untuk menghilangkan kesalahan dari data sehingga data yang digunakan lebih bersih dan relevan untuk analisis selanjutnya[9].

## 2. Missing Value

Memastikan integritas data, langkah ini mengisi nilai yang hilang dengan nilai rata-rata dari data yang tersedia. Sebagai hasilnya, data saham yang awalnya berjumlah 7490 data menjadi 11151 data setelah mengisi tanggal yang hilang. Demikian juga, data faktor ekonomi yang awalnya berjumlah 1086 data menjadi 1593 data setelah melengkapi tanggal yang hilang. Pengisian nilai yang hilang ini penting untuk menjaga konsistensi dan akurasi analisis[9].

## 3. Data Integrasi

Data saham dan data faktor ekonomi diintegrasikan untuk analisis yang lebih menyeluruh. Integrasi ini bertujuan untuk mengevaluasi korelasi antara perubahan harga saham dan faktor ekonomi seperti nilai tukar Rupiah terhadap USD. Penggabungan data ini membantu dalam memahami dampak perubahan ekonomi terhadap harga saham[10][11].

## 4. Feature Selection

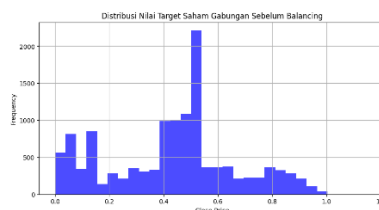
Pemilihan fitur yang relevan adalah tahap penting untuk meningkatkan performa model prediksi. Dalam penelitian ini, fitur yang digunakan mencakup harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), dan volume perdagangan untuk data saham, serta kurs jual dan beli untuk data faktor ekonomi. Pemilihan fitur yang tepat meningkatkan daya prediksi dan akurasi model. Fitur utama yang difokuskan adalah harga penutupan (*close*) berdasarkan penelitian terdahulu[5].

## 5. Data Transformasi

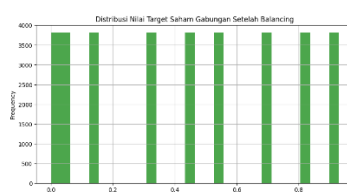
Transformasi data melibatkan perubahan data ke format yang konsisten. Pada tahap ini, normalisasi data dilakukan menggunakan metode *min-max scaling*. Normalisasi bertujuan untuk memastikan bahwa setiap variabel memiliki skala yang konsisten, yang meningkatkan kualitas data untuk analisis dan pengambilan keputusan[5].

### 2.3. Balancing Data

Untuk mengatasi masalah ketidakseimbangan data, teknik SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan[12]. Teknik ini mengurangi kemungkinan *overfitting* yang sering terjadi ketika data minoritas terlalu banyak diambil sampelnya, serta meningkatkan akurasi hasil analisis[13][14]. Data yang digunakan dalam prediksi ini memiliki distribusi kelas yang tidak seimbang, dengan kelas tertinggi mencapai 3816 data. Teknik *balancing* ini mengatur data agar lebih seimbang, seperti ditampilkan pada Gambar 2 dan Gambar 3.



Gambar 2 Data Saham Gabungan Sebelum *Balancing* Data



Gambar 3 Data Saham Gabungan Setelah *Balancing* Data

## 2.4. Splitting

Tahap penting dalam algoritme prediksi adalah membagi data menjadi dua set: data pelatihan dan data pengujian dimana pada penelitian ini dibagi menjadi 80% data pelatihan dan 20% data pengujian. Data pelatihan digunakan untuk melatih model XGBoost, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model setelah pelatihan. Pembagian data ini memastikan bahwa model dapat diterapkan pada data yang belum pernah dilihat sebelumnya dan mengevaluasi kemampuannya dalam memprediksi dengan akurat.

## 2.5. Hyperparameter

*Hyperparameter* tuning dilakukan untuk mengoptimalkan kinerja model XGBoost. Proses ini melibatkan penggunaan *Tree Parzen Estimator* (TPE) untuk menentukan kombinasi *hyperparameter* terbaik. *Hyperparameter* yang disesuaikan meliputi jumlah iterasi maksimum (*nrounds*), kedalaman maksimum pohon (*max\_depth*), jumlah subsample untuk setiap pohon (*subsample*), tingkat pembelajaran

(eta), dan jumlah variabel yang dipertimbangkan per pohon (*colsample\_bytree*). Proses ini bertujuan untuk meminimalkan nilai fungsi tujuan seperti RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*), sehingga meningkatkan akurasi model[15].

**2.6. Prediksi XGBoost**

Setelah model XGBoost dilatih dengan data pelatihan, model ini digunakan untuk membuat prediksi pada data pengujian. Model XGBoost menggunakan teknik ensemble berbasis pohon keputusan, di mana beberapa pohon keputusan dibangun selama fase pelatihan. Setiap pohon melihat sampel acak dan variabel yang dipilih secara acak. Proses ini mengoptimalkan model berdasarkan fungsi kerugian untuk meningkatkan kecocokan model dengan data. Setelah pelatihan, model XGBoost dinilai dengan menggunakan RMSE dan MAE. RMSE memberikan bobot lebih pada perbedaan besar antara prediksi dan nilai sebenarnya, sedangkan MAE mengukur rata-rata selisih absolut, memberikan gambaran umum akurasi prediksi[15].

**3. Hasil dan Pembahasan**

Berikut adalah hasil pembahasan dari penelitian ini, yang menggunakan metode XGBoost. Metode XGBoost telah digunakan untuk mengolah data dan menghasilkan model yang mampu memprediksi dengan akurasi tinggi.

**3.1. Pengujian XGBoost**

Pada tahap pengujian model algoritma menggunakan metode XGBoost ini akan di uji dengan data imbalance, *hyperparameter* tuning yang kemudian akan di evaluasi dengan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Pada pengujian ini akan dilakukan eksperimen untuk mengatasi ketidakseimbangan data dengan menggunakan dua metode yaitu SMOTE dan *Random OverSampling* dengan perbandingan data latih 80% dan data uji 20%. Kedua metode ini kemudian akan dilakukan eksperimen untuk mencari hasil terbaik. Eksperimen ini bertujuan untuk mendapatkan hasil error terbaik dari dua metode *Balancing Data* yang digunakan. Hasil akhir dari eksperimen ini akan dibandingkan berdasarkan nilai RMSE dan MAE untuk menentukan metode yang paling efektif dalam menyelesaikan masalah ketidakseimbangan data dan mencapai kinerja model yang optimal. Hasil perbandingan metode *Balancing Data* ini dapat dilihat pada Tabel 1.

Tabel 1 Perbandingan Hasil Error Metode *Balancing Data*

Metode <i>Balancing Data</i>		
	RMSE	MAE
<i>Random OverSampling</i>	0.0402	0.0310
SMOTE	0.0398	0.0305

**3.1.1. Hyperparameter Tuning & Random OverSampling**

Pada pengujian ini dilakukan dengan menggunakan metode *Random OverSampling* dengan jumlah sampel pada setiap kelas yang digunakan. Dapat dilihat pada Tabel 2.

Tabel 2 Jumlah Sampel Kelas Setelah *Random OverSampling*

Jumlah sampel pada setiap kelas Setelah <i>Random OverSampling</i> :									
1	2	3	4	5	6	7	8	9	10
3816	3816	3816	3816	3816	3816	3816	3816	3816	3816

Kemudian Parameter yang digunakan adalah Hyperparameter *Tree Parzen Estimator* untuk mendapatkan error terbaik. Dapat dilihat pada Tabel 3.

Tabel 3 *Hyperparameter Tuning TPE & Random OverSampling*

Data Latih 80% & Data Uji 20%		
<i>Hyperparameter</i> Terbaik	RMSE	MAE
<i>colsample_bytree</i>	1.0	
<i>learning_rate</i>	0.2642	
<i>max_depth</i>	3	0.0402
<i>n_estimators</i>	401	0.0310
<i>subsample</i>	1.0	

Hasil dari parameter yang digunakan didapatkan hasil RMSE 0.0402 dan MAE 0.0310.

**3.1.2. Hyperparameter Tuning & SMOTE**

Pada pengujian ini dilakukan dengan menggunakan metode SMOTE (*Synthetic Minority Over-Sampling Technique*) dengan jumlah sampel pada setiap kelas yang digunakan. Dapat dilihat pada Tabel 4.

Tabel 4 Jumlah Sampel Kelas Setelah SMOTE

Jumlah sampel pada setiap kelas Setelah SMOTE:									
1	2	3	4	5	6	7	8	9	10
3816	3816	3816	3816	3816	3816	3816	3816	3816	3816

Kemudian Parameter yang digunakan adalah Hyperparamter *Tree Parzen Estimator* untuk mendapatkan error terbaik. Dapat dilihat pada Tabel 5.

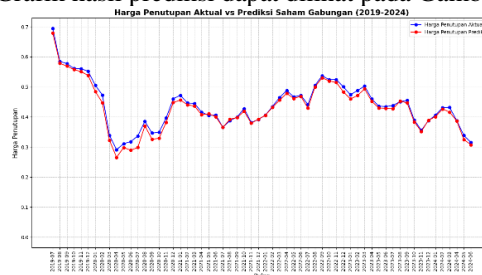
**Tabel 5 Jumlah Sampel Kelas Setelah SMOTE**  
**Data Latih 80% & Data Uji 20%**

Hyperparameter	RMSE	MAE
Terbaik		
colsample_bytree	1.0	
learning_rate	0.05	
max_depth	4	0.0398
n_estimators	299	0.0305
subsample	1.0	

Hasil dari parameter yang digunakan didapatkan hasil RMSE 0.0398 dan MAE 0.0305.

### 3.2. Hasil Prediksi

Hasil prediksi return saham menggunakan metode *Extreme Gradient Boosting* (XGBoost) untuk periode 29 Juni 2019 – 30 Juni 2024 berdasarkan gabungan data dari tujuh saham. Prediksi ini mencakup analisis tren dan evaluasi akurasi model dengan mempertimbangkan faktor ekonomi makro seperti nilai tukar Rupiah terhadap Dolar AS. Grafik hasil prediksi dapat dilihat pada Gambar 4.

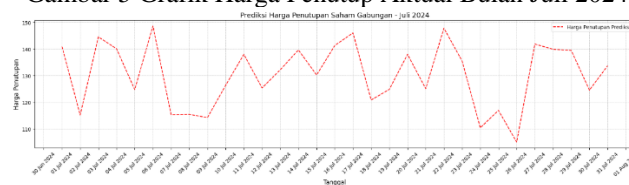


Gambar 4 Hasil Prediksi Tahun 29 Juni 2019-26 Juni 2024

Hasil analisis prediksi menunjukkan bahwa selisih antara harga penutupan aktual dan prediksi model pada tahun 2019 berkisar antara 0,01 hingga 0,03. Selisih ini sedikit meningkat menjelang 2020, namun tetap dalam kisaran toleransi 0,02 hingga 0,03. Pada tahun 2021, perbedaan rata-rata stabil antara 0,01 dan 0,02. Di tahun 2022, selisih antara harga penutupan aktual dan prediksi adalah 0,01 hingga 0,03. Pada tahun 2023, perbedaan rata-rata tetap baik, yaitu antara 0,01 dan 0,02. Model ini menunjukkan performa yang konsisten hingga Juni 2024, dengan selisih harga penutupan rata-rata tetap di kisaran 0,01 hingga 0,02.



Gambar 5 Grafik Harga Penutup Aktual Bulan Juli 2024



Gambar 6 Grafik Harga Penutup Prediksi Bulan Juli 2024

Grafik di atas menunjukkan prediksi harga penutupan saham untuk bulan Juli 2024. Meskipun bulan Juli 2024 belum terjadi, model memperkirakan rata-rata harga penutupan saham sebesar 0,31. Prediksi ini sedikit lebih rendah dibandingkan dengan estimasi rata-rata harga penutupan aktual bulan Juli 2024, yang diperkirakan mencapai 0,32. Prediksi ini didasarkan pada analisis data historis serta faktor eksternal yang mempengaruhi pergerakan pasar saham.

### 3.3. Analisis Hasil Pengujian Model XGBoost

Berdasarkan pengujian, penggunaan SMOTE dan *Random Oversampling* membantu menangani ketidakseimbangan data dalam prediksi. SMOTE menghasilkan RMSE sebesar 0.0398 dan MAE sebesar 0.0305, sementara *Random Oversampling* menghasilkan RMSE 0.0402 dan MAE 0.0310. Hasil ini menunjukkan bahwa SMOTE memberikan performa prediksi yang lebih baik dengan tingkat kesalahan

yang lebih rendah. Dengan meningkatkan sampel minoritas, SMOTE memungkinkan model mempelajari pola dari kelas yang kurang terwakili dengan lebih efektif, yang menjelaskan rendahnya nilai RMSE dan MAE dibandingkan *Random Oversampling*.

#### 4. Kesimpulan

Penelitian ini menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*), yang berhasil meningkatkan representasi kelas minoritas dan menyeimbangkan distribusi data. Penyesuaian *hyperparameter* dilakukan menggunakan *Tree Parzen Estimator*, serta penambahan faktor eksternal ekonomi makro, yaitu nilai Tukar Rupiah terhadap US Dolar. Setelah penerapan SMOTE dan optimasi *hyperparameter*, model *Extreme Gradient Boosting* menunjukkan peningkatan kinerja, dengan hasil RMSE sebesar 0.0398 dan MAE sebesar 0.0305.

#### Daftar Pustaka

- [1] D. Saputro and D. Swanjaya, "Analisa Prediksi Harga Saham Menggunakan Neural Network Dan Net Foreign Flow," *Gener. J.*, vol. 7, no. 2, pp. 96–104, 2023, doi: 10.29407/gj.v7i2.20001.
- [2] S. Sulistyowati and M. F. Rahmawati, "Analisis Pengaruh Pasar Saham Terhadap Pertumbuhan Ekonomi Di Negara Berkembang (Suatu Kajian Literatur)," *Res. Fair Unisri*, vol. 4, no. 1, 2020, doi: 10.33061/rsfu.v4i1.3392.
- [3] I. Oktavia and K. Genjar, "Sinergitas Quadruple Helix: e-Business dan Fintech sebagai Daya Dorong Pertumbuhan Ekonomi Lokal FAKTOR-FAKTOR YANG MEMPENGARUHI HARGA SAHAM," *J. Ris. Akunt. Multiparadigma*, vol. 6, no. 1, pp. 29–39, 2019.
- [4] H. Subekti and D. Saepudin, "Cross-Sectional Machine Learning Approach on Predicting Stock Return of LQ45 Index," *2022 1st Int. Conf. Softw. Eng. Inf. Technol. ICoSEIT 2022*, pp. 192–197, 2022, doi: 10.1109/ICoSEIT55604.2022.10030044.
- [5] M. G. Arfananda and D. Saepudin, "Prediksi Return Saham Pada Saham Indeks LQ45 Menggunakan Regresi Linear Univariat," *eProceedings ...*, vol. 10, no. 3, pp. 3628–3641, 2023.
- [6] S. Guan, Y. Wang, L. Liu, J. Gao, Z. Xu, and S. Kan, "Ultra-short-term wind power prediction method combining financial technology feature engineering and XGBoost algorithm," *Heliyon*, vol. 9, no. 6, p. e16938, 2023, doi: 10.1016/j.heliyon.2023.e16938.
- [7] J. Zheng *et al.*, "Metabolic syndrome prediction model using Bayesian optimization and XGBoost based on traditional Chinese medicine features," *Heliyon*, vol. 9, no. 12, p. e22727, 2023, doi: 10.1016/j.heliyon.2023.e22727.
- [8] S. Ben Jabeur, S. Mefteh-Wali, and J. L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Ann. Oper. Res.*, no. July, 2021, doi: 10.1007/s10479-021-04187-w.
- [9] M. R. Saputra and D. Saepudin, "Optimasi Portofolio Berbasis Prediksi Return Saham Menggunakan Hybrid XGBoost dan Improved Firefly Algorithm untuk Saham–Saham dalam Indeks LQ45," *eProceedings ...*, vol. 10, no. 3, pp. 3505–3514, 2023, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20622>
- [10] C. Ramonell, R. Chacón, and H. Posada, "Knowledge graph-based data integration system for digital twins of built assets," *Autom. Constr.*, vol. 156, no. October, 2023, doi: 10.1016/j.autcon.2023.105109.
- [11] B. Silva, J. Moreira, and R. L. de C. Costa, "Logical big data integration and near real-time data analytics," *Data Knowl. Eng.*, vol. 146, no. November 2022, p. 102185, 2023, doi: 10.1016/j.datak.2023.102185.
- [12] L. Wang, "Imbalanced credit risk prediction based on SMOTE and multi-kernel FCM improved by particle swarm optimization," *Appl. Soft Comput.*, vol. 114, p. 108153, 2022, doi: 10.1016/j.asoc.2021.108153.
- [13] A. M. Sapari, A. I. Hadiana, and F. R. Umbara, "Air Quality Classification Using *Extreme Gradient Boosting* (XGBOOST) Algorithm," *Innov. Res. Informatics*, vol. 5, no. 2, pp. 44–51, 2023, doi: 10.37058/innovatics.v5i2.8444.
- [14] S. Diantika, "Penerapan Teknik *Random Oversampling* Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing Menggunakan Algoritma Lightgbm," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 19–25, 2023, doi: 10.36040/jati.v7i1.6006.
- [15] S. Zedda, "Credit scoring: Does XGboost outperform logistic regression? A test on Italian SMEs," *Res. Int. Bus. Financ.*, vol. 70, no. PB, p. 102397, 2024, doi: 10.1016/j.ribaf.2024.102397.