

PERBANDINGAN ALGORITMA *XGBOOST* DAN *LSTM* UNTUK MEMPREDIKSI HARGA *BITCOIN* BERDASARKAN HARGA HARIAN, SENTIMEN, DAN *GOOGLE TRENDS INDEX*

Fadilla Zundina Ulya^{*1}, Siti Khomsah², Nia Annisa Ferani Tanjung³

^{1,2,3}Universitas Telkom Purwokerto, Purwokerto

Email: ¹21110001@ittelkom-pwt.ac.id, ²sitijk@telkomuniversity.ac.id, ³niaannisa@telkomuniversity.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 02 Desember 2024, diterima untuk diterbitkan: 08 Desember 2025)

Abstrak

Bitcoin merupakan salah satu jenis *cryptocurrency* yang banyak digunakan karena transaksinya yang aman, cepat, dan berpotensi memberikan keuntungan signifikan. Namun, volatilitas harga yang tinggi membuat aktivitas transaksi berisiko, karena pergerakan harga tidak hanya dipengaruhi oleh faktor internal, tetapi juga faktor eksternal seperti sentimen publik dan *Google Trends Index (GTI)*. Penelitian ini bertujuan membandingkan algoritma *XGBoost regression* dan *LSTM for regression* dalam memprediksi harga penutupan *bitcoin* dengan mengintegrasikan variabel harga harian, sentiment, dan *GTI* ke dalam model regresi yang disesuaikan dengan karakteristik data penelitian, dimana data yang digunakan bersifat non-linear, tidak berdistribusi normal, dan mengandung unsur *time series*. Berdasarkan hasil pengujian, model *XGBoost regression* terbaik diperoleh pada skenario dengan variabel eksternal. Namun menghasilkan nilai *RMSE* sebesar 5169,898 *USD* dan *R2-Score* sebesar -13%, yang menunjukkan adanya *overfitting* dan model kurang tepat untuk data *time series*. Sebaliknya, model *LSTM for regression* dengan variabel eksternal dan kombinasi *hyperparameter* terbaik menunjukkan performa yang lebih unggul dengan *RMSE* sebesar 1378,55 *USD* dan *R2-Score* sebesar 92%. Model ini tidak menunjukkan indikasi *overfitting* dan mampu mereplikasi pola pergerakan harga secara akurat. Hal ini menunjukkan bahwa *LSTM for regression* lebih mampu mengenali pola temporal dalam data historis. Selain itu, fitur harga historis, khususnya *Open* teridentifikasi sebagai variabel paling dominan berdasarkan hasil analisis menggunakan metode *SHAP*.

Kata kunci: prediksi harga bitcoin, sentimen, google trends index, *LSTM*, *XGBoost regression*

COMPARISON OF *XGBOOST* AND *LSTM* ALGORITHMS TO PREDICTE *BITCOIN* PRICE BASED ON DAILY PRICE, SENTIMENT, AND *GOOGLE TRENDS INDEX*

Abstract

Bitcoin is one type of *cryptocurrency* that is widely used because its transactions are safe, fast, and have the potential to provide significant profits. However, high price volatility makes transaction activities risky, because price movements are not only influenced by internal factors, but also external factors such as public sentiment and the *Google Trends Index (GTI)*. This study aims to compare the *XGBoost regression* and *LSTM for regression* algorithms in predicting bitcoin closing prices by integrating daily price, sentiment, and *GTI* variables into a regression model that is adjusted to the characteristics of the research data, where the data used is non-linear, not normally distributed, and contains time series elements. Based on the test results, the best *XGBoost regression* model was obtained in the scenario with external variables. However, it produces an *RMSE* value of 5169.898 *USD* and an *R2-Score* of -13%, which indicates *overfitting* and the model is less appropriate for time series data. In contrast, the *LSTM for regression* model with external variables and the best combination of *hyperparameters* shows superior performance with an *RMSE* of 1378.55 *USD* and an *R2-Score* of 92%. This model does not show any indication of *overfitting* and is able to replicate price movement patterns accurately. This shows that *LSTM for regression* is better able to recognize temporal patterns in historical data. In addition, historical price features, especially *Open*, are identified as the most dominant variables based on the results of the analysis using the *SHAP* method.

Keywords: bitcoin price prediction, sentiment, google trends index, *LSTM*, *XGBoost regression*

1. PENDAHULUAN

Investasi adalah kegiatan menanamkan aset ke dalam suatu sektor atau area tertentu (Klaudia et al.,

2018). Ada banyak cara untuk melakukan investasi antara lain, memiliki aset tanah, emas, saham, deposito, obligasi, mata uang asing hingga mata uang

kripto atau biasa dikenal dengan istilah *cryptocurrency* (Fadlurachman, 2022). *Cryptocurrency* adalah mata uang digital yang semakin populer dimanfaatkan sebagai aset dan penanaman modal oleh penggunanya (Pradana & Lestanti, 2020). Meski tidak memiliki bentuk fisik, mata uang ini dapat ditransaksikan melalui perdagangan online. Jumlah investor *cryptocurrency* atau mata uang kripto mengalami kenaikan setiap tahunnya sampai angka belasan juta pada tahun 2021 (Muhammad et al., 2023).

Salah satu jenis *cryptocurrency* yang telah meluas pemakaiannya yaitu *bitcoin*. *Bitcoin* pertama kali diperkenalkan tahun 2009 oleh seseorang (ataupun sekelompok) dengan nama samaran Satoshi Nakamoto (Rasdi Rere, Hariyanto & Rozi, 2022). Pengelola *bitcoin* bukan berasal dari bank sentral, melainkan dikelola langsung dari pengguna *bitcoin* yang identitasnya dirahasiakan. Transaksi *bitcoin* dicatat dalam blok-blok pada teknologi *blockchain* yang setiap bloknnya terhubung dan membentuk rantai yang susah untuk dimanipulasi (Pratama & Utama, 2023). Proses transaksinya yang aman, mudah, dan cepat serta memiliki potensi keuntungan yang signifikan di masa mendatang membuat *bitcoin* memiliki banyak peminat.

Walaupun demikian, perdagangan *bitcoin* merupakan aktivitas yang berisiko karena harga *cryptocurrency* yang terus berfluktuasi secara signifikan setiap harinya (Habsyi, Ramdani & Khomsah, 2023). Berdasarkan data harga *bitcoin* selama 5 tahun kebelakang (2019-2023) pada website *Yahoo finance*, harga *bitcoin* pada Bulan Maret 2021 melambung tinggi mencapai 61243,69 USD. Namun, harga *bitcoin* mengalami penurunan yang cukup tajam pada Bulan Juli 2021 mencapai 31533,07 USD. Kemudian, harga *bitcoin* kembali meningkat secara signifikan dan mencapai puncak tertingginya pada Bulan November 2021, yaitu sebesar 64469,63 USD. Berdasarkan data tersebut, dapat disimpulkan bahwa salah satu harga *cryptocurrency*, yaitu *bitcoin* sangat fluktuatif dari masa ke masa. Oleh karena itu, prediksi harga *bitcoin* sangat penting dilakukan untuk meminimalisir terjadinya kerugian dalam investasi *bitcoin*.

Pergerakan harga *bitcoin* yang sangat fluktuatif tidak hanya dipengaruhi oleh faktor internal. Namun, juga faktor eksternal seperti sentimen publik mengenai masalah politik maupun ekonomi global juga berpengaruh terhadap fluktuatifnya harga *bitcoin* (Maliki, Cholissodin & Yudistira, 2022). Sentimen publik mencerminkan pandangan dan keyakinan masyarakat terhadap *bitcoin* yang secara langsung dapat mempengaruhi permintaan dan penawaran. Sebagai contoh pada Bulan Juni 2021, harga *bitcoin* mengalami penurunan karena Elon Musk mengungkapkan ketidak-tertarikannya terhadap *cryptocurrency* dengan menambahkan tagar *bitcoin* di platform *Twitter* dan mengumumkan bahwa Tesla tidak akan menerima pembayaran

menggunakan *bitcoin* lagi. Hal tersebut memicu dugaan masyarakat bahwa nilai *cryptocurrency*, khususnya sebagai alat pembayaran sangat fleksibel dan bergantung pada penggunaannya (Jannah, 2022). Keterkaitan antara sentimen publik dengan fluktuasi harga *bitcoin* juga dibuktikan oleh penelitian (Pratama et al., 2023) yang dilakukan dengan menghitung korelasi antara harga *bitcoin* dengan sentimen data *Twitter*. Nilai koefisien ϕ yang dihasilkan dari perhitungan korelasi koefisien *pearson* sebesar 0,5 menunjukkan bahwa harga *bitcoin* dengan sentimen data *Twitter* yang membahas mengenai *bitcoin* memiliki hubungan positif kuat.

Selain sentimen publik, popularitas juga berpengaruh terhadap harga *bitcoin* yang dapat diakumulasi melalui data pencarian *Google*. *Google* mendominasi ranah mesin pencarian dan menjadi tempat utama yang mencerminkan berbagai macam pandangan dan minat masyarakat (Padhila, Cholissodin & Adikara, 2022). Salah satu layanan *Google* yang dapat digunakan untuk menampilkan nilai indeks dari kata kunci pencarian oleh pengguna internet di seluruh dunia, yaitu *Google Trends* (Riyanto, 2014). Nilai indeks pada *Google Trends* menunjukkan popularitas relatif yang berkisar antara 0 sampai 100 serta memiliki keterkaitan dengan harga *bitcoin* yang dibuktikan oleh penelitian (Indriyani & Usman, 2024), dimana hasilnya mengatakan bahwa nilai indeks *bitcoin* pada *Google Trends* memiliki hubungan yang positif dan signifikan terhadap harga *bitcoin*.

Adanya hubungan antara kedua faktor eksternal tersebut dengan harga *bitcoin* menjadi fokus pada penelitian ini untuk melihat sejauh mana akurasi terbaik dapat dihasilkan oleh model prediksi yang dibangun menggunakan teknik regresi. Teknik regresi memungkinkan suatu model dapat memproyeksikan kondisi di masa depan dengan cara membentuk pola hubungan antara satu atau lebih variabel independen dengan variabel dependen (Ayuni & Fitriana, 2019). Pemilihan teknik regresi harus mempertimbangkan karakteristik data serta didasarkan pada hasil pengujian asumsi klasik untuk memastikan validitas model yang diterapkan terhadap data yang dianalisis (Mardiatmoko, 2020).

Beberapa algoritma regresi yang biasa digunakan untuk memprediksi harga kripto antara lain yaitu, *XGBoost regression* (Huang, 2024) dan *Long Short Term Memory (LSTM) for regression* (Moch Farryz Rizkilloh & Sri Widiyanesti, 2022; Pratama dan Utama, 2023). Kedua algoritma tersebut memiliki kelebihan dan kekurangan masing-masing. Oleh karena itu, perbandingan dilakukan untuk menentukan algoritma yang paling optimal dalam memprediksi harga *bitcoin*, baik dari segi performa model maupun kemampuannya dalam menangani pola data yang kompleks (Ardian, Khomsah & Pandiya, 2024).

Keunggulan *XGBoost regression* dibuktikan pada penelitian (Huang, 2024) yang menunjukkan

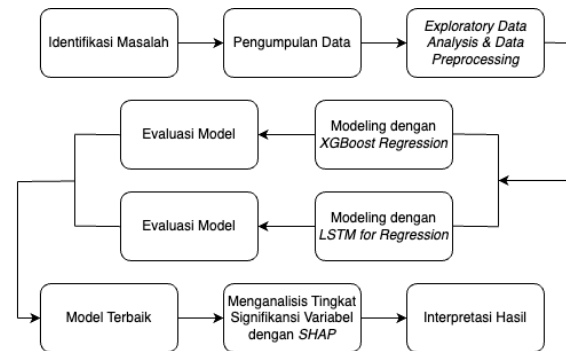
bahwa model *Grid Search XGBoost* menghasilkan *MSE* dan *R2-Score* yang terendah dibandingkan *linear regression* dan *random forest regression* dalam memprediksi harga *bitcoin* berdasarkan harga harian *bitcoin* dan sentiment pasar. Selain itu, performa *XGBoost regression* juga cukup baik dalam memprediksi harga saham Perusahaan *Amazon.com, Inc*, dimana model berhasil mencapai nilai *RMSE* sebesar 0,009437 (Ardana, 2023). Penelitian lain juga dilakukan untuk memprediksi harga *bitcoin* menggunakan *LSTM*, dimana hasil penelitiannya mengatakan bahwa model memiliki akurasi yang cukup baik dengan nilai *RMSE* dan *MAPE* berturut-turut sebesar 2033,28 dan 3,53% (Pratama et al., 2023). Keandalan *LSTM* dalam kasus prediksi terutama pada data *time series* dibuktikan dengan hasil penelitian (Moch Farryz Rizkilloh & Sri Widiyanesti, 2022) yang mampu memprediksi harga *cryptocurrency*, terutama untuk koin *DOGE* dan *ADA* dengan nilai *RMSE* berturut-turut sebesar 0,0544 dan 0,1607.

Berdasarkan kajian penelitian-penelitian tersebut, penelitian ini bertujuan membandingkan performa model regresi, yaitu *XGBoost regression* dan *LSTM for regression* dalam memprediksi harga penutupan *bitcoin*. Penelitian ini berbeda dengan penelitian (Arisandi, Ferdiansyah & Atika, 2020) dan (Padhila, Cholissodin & Adikara, 2022) yang hanya menggunakan variabel harga historis dan sentiment atau harga historis dan *Google Trends*. Pada penelitian ini, kedua variabel eksternal tersebut, yaitu sentiment dan *Google Trends* diintegrasikan sekaligus dengan harga historis *bitcoin* untuk memprediksi harga penutupan *bitcoin*. Penelitian ini tidak hanya berfokus pada pencapaian model prediksi yang akurat, tetapi juga bertujuan untuk mengidentifikasi variabel yang memiliki dampak signifikan terhadap prediksi harga *bitcoin*.

2. METODE PENELITIAN

Tahapan penelitian disusun dalam langkah-langkah terstruktur menggunakan diagram alir yang disajikan pada Gambar 1. Penelitian ini memanfaatkan *Python* sebagai bahasa pemrograman utama, sementara *Google Colaboratory* digunakan sebagai platform untuk menjalankan analisis data dan membangun model. Langkah pertama yang dilakukan pada penelitian ini yaitu identifikasi masalah dilanjutkan dengan pengumpulan data yang dibutuhkan. Setelah itu, dilakukan *EDA* untuk memeriksa karakteristik data serta mendeteksi adanya anomali. Jika ditemukan adanya indikasi data yang tidak wajar, maka dilakukan *preprocessing* untuk memperbaiki dan menyesuaikan data dengan algoritma yang diterapkan. Langkah selanjutnya yaitu modeling menggunakan algoritma *XGBoost regression* dan *LSTM for regression*. Hasilnya kemudian dievaluasi performanya menggunakan *RMSE* dan *R2-Score* untuk memperoleh model terbaik yang selanjutnya digunakan untuk

menganalisis kontribusi variabel-variabel independen terhadap harga penutupan *bitcoin* menggunakan metode *SHAP*. Setelah analisis data selesai dilakukan, langkah terakhir yaitu menginterpretasikan hasil dan menyusun kesimpulan berdasarkan temuan penelitian.

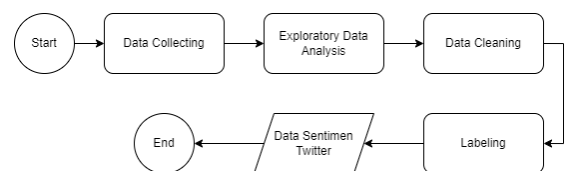


Gambar 1. Diagram alir penelitian

2.1 Pengumpulan Data

2.1.1 Data Sentimen

Alur pengumpulan data sentimen diilustrasikan pada Gambar 2. Tahap awalnya yaitu pengumpulan data yang dilakukan dengan *crawl* data *tweet* dari platform *Twitter* (sekarang menjadi *X*) dengan kata kunci “*bitcoin*” dan “*price*” sejak tanggal 1 Maret 2023 sampai 29 Februari 2024 sebanyak 35.560 *tweet*. Selanjutnya, dilakukan *EDA* berupa *data understanding*, identifikasi *missing value* dan data duplikat. Setelah itu, dilakukan *data cleaning* untuk membersihkan karakter *non-ASCII*, *mention* pengguna, tanda baca yang berlebihan, beberapa karakter khusus, dan menghapus kolom yang tidak diperlukan untuk proses *labeling*. Tahap akhir dari pengumpulan data sentimen yaitu membuat label (*labeling*) sentimen untuk kolom *tweet* menggunakan model *pre-trained*, yaitu *Large Language Model (LLM)* dengan *prompt* yang telah disesuaikan dengan kebutuhan penelitian. Setelah *labeling* selesai, maka diperoleh data sentimen yang siap digunakan pada tahap selanjutnya.



Gambar 2. Alur pengumpulan data sentimen

2.1.2 Data Google Trends Index

Selain data sentimen, penelitian ini juga menggunakan data indeks pencarian *Google* yang diperoleh melalui website <https://trends.google.com/> dengan kata kunci “*bitcoin price*” dengan periode waktu harian, sejak tanggal 1 Maret 2023 sampai 29 Februari 2024. Website tersebut menampilkan nilai indeks dari kata kunci pencarian oleh pengguna

internet di seluruh dunia dengan rentang 0 sampai 100.

2.1.3 Data Historis Harga Bitcoin

Penelitian ini juga menggunakan data historis harga *bitcoin* yang merupakan data sekunder dari website *Yahoo Finance* yang dapat diakses pada link <https://finance.yahoo.com/quote/BTC-USD/history/> dengan periode waktu satu tahun, sejak tanggal 1 Maret 2023 sampai 29 Februari 2024. Adapun variabel yang digunakan yaitu *Date*, *Open*, *High*, *Low*, *Adj. Close*, *Close*, dan *Volume*.

2.2 EDA dan Data Preprocessing

Exploratory Data Analysis (EDA) merupakan tahap awal yang esensial dalam analisis data untuk memperoleh wawasan mendalam tentang data sebelum melakukan analisis lebih lanjut atau membangun model. Pada penelitian ini, tahap *EDA* dibagi menjadi dua bagian, yaitu *data understanding* dan visualisasi data. Selanjutnya, *data preprocessing* dilakukan untuk mempersiapkan data sebelum pemodelan dengan menyesuaikan data berdasarkan karakteristik algoritma yang akan digunakan. Tahap ini terbagi menjadi *data cleaning*, transformasi, dan *splitting data*. Teknik transformasi yang digunakan yaitu *Min-Max Scaler*, karena dapat mengurangi dampak dari perbedaan skala antar variabel sehingga mampu mengatasi dampak *outlier* yang signifikan akibat volatilitas harga *bitcoin*. Selain itu, *Min-Max Scaler* memastikan semua fitur berada dalam rentang yang sama sehingga algoritma regresi dapat menangani multikolinearitas dan autokorelasi dengan lebih baik. Adapun perhitungan *MinMax Scaler* dituliskan pada persamaan (1)

$$X'_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (1)$$

dimana X'_i merupakan hasil normalisasi data, x_i merupakan nilai aktual data ke- i , $\max(X)$ merupakan nilai terbesar dari keseluruhan dataset, dan $\min(X)$ merupakan nilai terkecil dari keseluruhan dataset (Azizah et al., 2019).

Setelah proses transformasi, data dibagi menjadi data latih dan data uji menggunakan skema *time series split* dengan proporsi 80:20. Pembagian ini dilakukan secara berurutan untuk menjaga urutan kronologis data dan menghindari kebocoran informasi dari data masa depan ke masa lalu. Pendekatan ini bertujuan memastikan validitas evaluasi model dalam memprediksi data baru.

2.3 Pemodelan

Tahap pemodelan dilakukan dengan membandingkan performa dua algoritma, yaitu *XGBoost regression* dan *LSTM for regression*, dalam memprediksi harga *bitcoin*. Kedua algoritma tersebut

diuji pada dua skenario berbeda untuk menganalisis pengaruh variabel eksternal terhadap performa model. Pada skenario pertama, model dibangun hanya dengan menggunakan data harga harian sebagai fitur utama tanpa melibatkan variabel eksternal. Pada skenario kedua, kedua model melibatkan variabel eksternal seperti *volume*, *Google Trends Index*, dan sentimen sebagai fitur tambahan.

2.3.1 XGBoost Regression

XGBoost regression merupakan algoritma yang diimplementasikan untuk memprediksi data kontinu dengan mengadopsi pendekatan *ensemble* dalam *gradient tree boosting* (Sudaryanto & Sudaryanto, 2022). *XGBoost* membangun serangkaian pohon keputusan, dimana setiap pohon secara berurutan berupaya memperbaiki kesalahan prediksi dari pohon sebelumnya melalui proses pembelajaran yang mengoptimalkan parameter untuk meminimalkan fungsi loss, seperti *Mean Squared Error (MSE)* yang formulanya dituliskan pada persamaan (2)

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

dimana, y_i merupakan nilai aktual dari data ke- i , \hat{y}_i merupakan nilai hasil prediksi data ke- i , dan n merupakan banyaknya data (Wang et al., 2019).

2.3.2 LSTM for Regression

Long Short Term Memory (LSTM) merupakan jenis khusus *Recurrent Neural Network (RNN)* yang dirancang untuk memecahkan masalah dalam menangkap ketergantungan jangka panjang. Dibandingkan dengan *RNN*, model *LSTM* telah terbukti lebih efektif dalam menyimpan dan menggunakan informasi pada urutan data yang lebih panjang (Adherda, Hikmatyar & Ruuhwan, 2023). Dalam struktur jaringan *LSTM*, input dari waktu tertentu dan output dari waktu sebelumnya dimasukkan ke dalam unit *LSTM*, yang kemudian menghasilkan output yang diteruskan ke waktu berikutnya. *LSTM* mempunyai tiga *gate* utama, yaitu *input gate*, *output gate*, dan *forget gate* (Wiranda & Sadikin, 2019). *Input gate* berperan dalam pembaruan *internal state* berdasarkan input saat ini dan *internal state* sebelumnya, *output gate* berperan untuk mengatur pengaruh *internal state* pada sistem, sedangkan *forget gate* berperan dalam penentuan banyaknya *internal state* sebelumnya yang harus dilupakan. Fungsi aktivasi yang digunakan pada *forget gate* yaitu fungsi aktivasi *sigmoid* yang formulanya dituliskan pada persamaan (3).

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

dimana f_t merupakan *forget gate* pada waktu t , σ merupakan fungsi aktivasi *sigmoid* yang mengubah nilai masukan menjadi angka dalam rentang 0 sampai 1, W_f merupakan matriks bobot yang diterapkan pada

input saat ini dan output sebelumnya, h_{t-1} merupakan output dari unit *LSTM* pada waktu sebelumnya, x_t merupakan input saat ini pada waktu t , dan b_f merupakan bias dari *forget gate* (Arfan & ETP, 2019).

2.4 Evaluasi Model

Setelah tahap pemodelan selesai, maka dilakukan pengukuran matriks evaluasi untuk mengetahui performa masing-masing model menggunakan *RMSE* dan *R2-Score*. Model terbaik diperoleh apabila memenuhi kriteria, yaitu memiliki nilai *RMSE* terendah dan *R2-Score* tertinggi.

2.4.1 Root Mean Squared Error (RMSE)

Root Mean Square Error (RMSE) merupakan akar kuadrat dari kuadrat kesalahan rata-rata yang dihasilkan dari perhitungan (Yulia Hayuningtyas dan Sari, 2022). Tujuan digunakannya *RMSE* yaitu untuk mengevaluasi tingkat akurasi dari model prediksi. *RMSE* memberikan ukuran standar untuk tingkat kesalahan, dimana semakin kecil nilai *RMSE* maka semakin baik kinerja model. Hal tersebut dikarenakan nilai yang lebih kecil menunjukkan bahwa prediksi model lebih mendekati nilai sebenarnya (Zundina Ulya, Rony Wijaya & Laras Puspita, 2023). *RMSE* dapat dihitung dengan rumus pada persamaan (4).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

dimana \hat{y}_i merupakan nilai hasil prediksi, y_i merupakan nilai aktual, dan n merupakan banyaknya data (Li, 2022).

2.4.2 Goodness-of-fit (R2-Score)

R2-Score merupakan suatu nilai yang menggambarkan proporsi variabilitas dalam variabel dependen yang dapat dijelaskan oleh variabilitas dalam variabel independen. *R2-Score* digunakan untuk mengevaluasi seberapa baik model regresi dapat memprediksi variabel dependen dan berada pada kisaran nilai 0 hingga 1, dimana nilai yang mendekati 1 menunjukkan model yang lebih baik (Rhamadhani & Saputri, 2021). Adapun perhitungan *R2-Score* dituliskan pada persamaan (5).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

dimana y_i merupakan observasi respon ke- i , \hat{y}_i merupakan prediksi respon ke- i , dan \bar{y}_i merupakan rata-rata respon ke- i .

2.5 SHapley Additive Explanation (SHAP)

Setelah diperoleh model terbaik, langkah berikutnya yaitu menganalisis pengaruh signifikansi dan kontribusi variabel x terhadap variabel y menggunakan metode *SHAP* dengan menghitung

nilai *Shapley* bagi setiap variabel. *SHapley Additive Explanation (SHAP)* adalah salah satu metode yang digunakan untuk memahami model *machine learning*. Tujuan utama *SHAP* yaitu menghitung kontribusi setiap fitur untuk menjelaskan prediksi individual x berdasarkan selisih antara nilai rata-rata ketika fitur tersebut ada dan tidak ada (Primulando, 2021). Metode ini didasarkan pada teori *Shapley* yang memberikan solusi pembagian kontribusi secara adil bagi setiap pemain dalam permainan, dimana fitur diibaratkan sebagai pemain dan skor prediksi sebagai hadiah. Nilai *Shapley* dihitung dengan mengevaluasi semua kombinasi prediktor, baik dengan maupun tanpa prediktor ke- i untuk menentukan kontribusi masing-masing variabel, dimana formulanya dituliskan pada persamaan (6).

$$\varphi_j = \sum_{S \subseteq M \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} (v(S \cup \{j\}) - v(S)) \quad (6)$$

dimana φ_j merupakan nilai kontribusi peubah ke- j (nilai *Shapley*), S merupakan subset fitur yang digunakan pada model, M merupakan jumlah peubah, $v(S \cup \{j\})$ merupakan prediksi yang diperoleh dari subset fitur S dengan adanya peubah j dalam subset fitur, dan $v(S)$ merupakan prediksi yang diperoleh dari subset fitur S tanpa adanya peubah j dalam subset fitur (Mayapada, Yanti & Syarifuddin, 2022).

3. HASIL DAN PEMBAHASAN

3.1 EDA dan Data Preprocessing

Data sentimen, *google trends*, dan harga historis *bitcoin* yang telah dikumpulkan kemudian diintegrasikan, dimana cuplikan datanya dapat dilihat pada Tabel 1.

Data yang digunakan pada penelitian ini terdiri dari 366 baris dan 9 kolom dengan periode waktu satu tahun, sejak tanggal 1 Maret 2023 sampai 29 Februari 2024. Setelah dilakukan eksplorasi data, diperoleh hasil bahwa tidak terdapat duplikasi dan *missing value* pada data. Selain itu, dapat diidentifikasi bahwa data historis harga *bitcoin* memiliki tipe data kontinu dan bersifat *time series*, sedangkan data sentimen dan *Google Trends* memiliki tipe data diskrit dan kategorikal. Sebaran data yang dihasilkan memiliki *outlier* karena harga *bitcoin* yang sangat fluktuatif.

Kemudian, dilakukan uji asumsi klasik untuk menguji apakah data layak untuk dilakukan pemodelan regresi linear. Pada tabel 2, dapat dilihat bahwa dari lima uji asumsi klasik yang dilakukan, mengindikasikan bahwa data tidak berdistribusi normal, terjadi multikolinearitas antar variabel bebas, tidak adanya linearitas, terdapat autokorelasi dan terjadi gejala heteroskedastisitas pada data, dimana seluruhnya menunjukkan ketidaksesuaian dengan asumsi regresi linear. Oleh karena itu, dapat dikatakan bahwa model regresi *non-linear* lebih sesuai digunakan untuk menganalisis dataset pada penelitian ini.

Tabel 1. Integrasi data harga harian *bitcoin*, sentimen, dan *google trend*

Date	Open	High	Low	Adj_Close	Volume	Sentimen	Index	Close
2023-03-01	23150.92	23880.63	23088.62	23646.55	24662841200	positif	43	23646.55
2023-03-01	23647.01	23739.13	23245.02	23475.46	20386398516	negatif	45	23475.46
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
2024-02-29	62499.18	63585.64	60498.73	61198.38	65496611844	positif	93	61198.38

Tabel 2. Uji asumsi klasik

Jenis Pengujian	Variabel	Nilai	Keputusan	Keterangan
Uji Linearitas	Seluruh variabel x	$p\text{-value}=7,563\text{e-}16$	<0,05; maka terjadi non-linearitas pada data	Tidak terpenuhi
Uji Normalitas	Seluruh variabel x	$p\text{-value}=4,238\text{e-}72$	<0,05; maka data tidak berdistribusi normal	Tidak terpenuhi
Uji Heteroskedastisitas	Open	$p\text{-value}=0,000$	<0.05; maka terjadi gejala heteroskedastisitas pada data	Tidak terpenuhi
	Low	$p\text{-value}=0,000$		
	Adj_Close	$p\text{-value}=0,016$		
	Volume	$p\text{-value}=0,000$		
	Index	$p\text{-value}=0,000$		
Uji Multikolinearitas	High	$p\text{-value}=0,592$	>0,05, maka tidak terjadi gejala heteroskedastisitas pada data	Tidak terpenuhi
	Sentimen	$p\text{-value}=0,616$		
	Open	$VIF=535,921$	>10,00; maka terjadi multikolinearitas pada data	
	High	$VIF=1163,219$		
	Low	$VIF=803,831$		
	Adj_Close	$VIF=937,047$	<10,00; maka tidak terjadi multikolinearitas pada data	
	Volume	$VIF=4,817$		
	Sentimen	$VIF=1,176$		
Index	$VIF=1,727$			
Uji Autokorelasi	Seluruh variabel x	Nilai $\text{durbin-watson}=0.124$	<2,0; maka terjadi autokorelasi positif	Tidak Terpenuhi

Selanjutnya, kolom *Date* dihapus karena hanya berfungsi sebagai penanda waktu, sedangkan kolom *Adj. Close* juga dihapus karena memiliki korelasi sempurna dengan kolom *Close*, yang menjadikannya redundan dan tidak memberikan kontribusi tambahan pada model. Setelah itu, dilakukan transformasi data dengan teknik *scaling* menggunakan *Min-Max Scaler* untuk menyesuaikan nilai-nilai setiap fitur ke dalam rentang tertentu sehingga dapat mengurangi perbedaan skala antar fitur. Proses terakhir pada tahap *data preprocessing* ini yaitu *splitting* atau pembagian data dengan proporsi 80:20. Dimana 80% atau sekitar 292 baris menjadi data latih dan 20% atau sekitar 73 baris menjadi data uji. Setelah menyelesaikan seluruh rangkaian proses pada tahap ini, data telah siap untuk masuk dalam tahap pemodelan.

3.2 Modeling dan Evaluasi

Tahap pemodelan dilakukan dengan dua skenario di setiap algoritma yang digunakan, yaitu pemodelan yang dilakukan tanpa melibatkan variabel eksternal (*volume*, *Google Trends Indeks*, dan *sentimen*) dan pemodelan yang dilakukan dengan melibatkan variabel eksternal. Hal tersebut bertujuan untuk melihat pengaruh penambahan variabel eksternal terhadap performa model.

3.2.1 XGBoost Regression

Pembangunan model *XGBoost* dilakukan dengan pendekatan *hyperparameter tuning* menggunakan *Optuna* untuk memperoleh kombinasi parameter terbaik. Tujuannya yaitu meminimalkan kesalahan prediksi dan meningkatkan performa model. Parameter yang disesuaikan meliputi,

learning_rate, *max_depth*, *n_estimators*, *subsample*, *colsample_bytree*, *alpha*, dan *lambda*. Adapun kombinasi *hyperparameter* yang optimal untuk masing-masing skenario disajikan pada Tabel 3.

Tabel 3. *Hyperparameter algoritma XGBoost regression*

Hyperparameter	Skenario	
	Tanpa Variabel Eksternal	Dengan Variabel Eksternal
<i>learning_rate</i>	0,05038	0,02571
<i>max_depth</i>	1	2
<i>n_estimators</i>	378	494
<i>subsample</i>	0,72814	0,99729
<i>colsample_bytree</i>	0,83603	0,75292
<i>alpha</i>	0,00063	0,00027
<i>lambda</i>	0,00019	0,00030

Hyperparameter tersebut kemudian digunakan pada pelatihan model dengan skenario yang telah ditetapkan sebelumnya. Pada Tabel 4, dapat dilihat bahwa tanpa penggunaan variabel eksternal pada model *XGBoost*, rata-rata kesalahan yang dihasilkan antara nilai sebenarnya dan nilai prediksi harga penutupan *bitcoin* pada data latih dan data uji berturut-turut sebesar 540,95 dan 5407,22. Nilai *R2-Score* pada data latih cukup tinggi yaitu 99%, sedangkan pada data uji *R2-Score* bernilai negatif sebesar -24%. Selain itu, penambahan variabel eksternal pada model *XGBoost* mampu menurunkan nilai rata-rata kesalahan antara nilai sebenarnya dan nilai prediksi, baik pada data latih maupun data uji berturut-turut sebesar 439,27 dan 5169,90. Nilai *R2-Score* untuk data latih ketika menggunakan variabel eksternal sama dengan nilai *R2-Score* tanpa variabel eksternal, yaitu sebesar 99%. Namun nilai *R2-Score*

pada data uji lebih meningkat dari pemodelan dengan skenario sebelumnya, yaitu sebesar -13%.

Tabel 4. Evaluasi model *XGBoost regression*

Skenario	Evaluasi	Train Set	Test Set
Tanpa variabel eksternal	<i>RMSE</i>	540,95	5407,22
	<i>R2-Score</i>	0,99	-0,24
Dengan variabel eksternal	<i>RMSE</i>	439,27	5169,90
	<i>R2-Score</i>	0,99	-0,13

Secara keseluruhan, hasil evaluasi dari kedua skenario menunjukkan bahwa penambahan variabel eksternal, seperti sentimen, *index*, dan *volume* dapat meningkatkan performa model, terutama dalam menurunkan *RMSE* dan meningkatkan *R2-Score* pada data latih maupun data uji. Namun, kedua hasil tersebut mengindikasikan adanya *overfitting* pada model, yang dibuktikan dengan selisih nilai *RMSE* dan *R2-Score* antara data latih dan data uji cukup tinggi, sehingga dapat dikatakan bahwa model terlalu baik dalam memprediksi data latih, namun tidak mampu memprediksi data uji yang diberikan.

3.2.2 *LSTM for Regression*

Pembangunan model *LSTM* dilakukan dengan eksplorasi kombinasi *hyperparameter* berupa jumlah *layer*, *neuron*, dan *epoch* untuk meminimalkan nilai *MSE*. Pada skenario tanpa variabel eksternal, kombinasi terbaik diperoleh dengan 2 *layer*, 150 *neuron*, dan 10 *epochs* dengan *MSE* sebesar 0,001123703. Sementara itu, pada skenario dengan variabel eksternal, kombinasi terbaik yaitu menggunakan 2 *layer*, 200 *neuron*, dan 30 *epochs* dengan *MSE* sebesar 0,001001388.

Hyperparameter tersebut kemudian digunakan pada pelatihan model dengan skenario yang telah ditetapkan sebelumnya. Pada Tabel 5, dapat dilihat bahwa tanpa penggunaan variabel eksternal pada model *LSTM*, rata-rata kesalahan yang dihasilkan antara nilai sebenarnya dan nilai prediksi harga penutupan *bitcoin* pada data latih dan data uji berturut-turut sebesar 807,48 dan 1632,28. *R2-Score* pada data latih dan data uji memiliki nilai yang cukup tinggi, yaitu berturut-turut sebesar 97% dan 89%. Penambahan variabel eksternal pada model *LSTM* menghasilkan rata-rata kesalahan prediksi yang lebih rendah dari pemodelan tanpa menggunakan variabel eksternal, yaitu sebesar 734,01 untuk data latih dan 1378,55. Nilai *R2-Score* pada model *LSTM* dengan penambahan variabel sentimen untuk data latih dan data uji lebih tinggi daripada skenario sebelumnya, yaitu berturut-turut sebesar 98% dan 92%.

Tabel 5. Evaluasi model *LSTM for regression*

Skenario	Evaluasi	Train Set	Test Set
Tanpa variabel eksternal	<i>RMSE</i>	807,48	1632,28
	<i>R2-Score</i>	0,97	0,89
Dengan variabel eksternal	<i>RMSE</i>	734,01	1378,55
	<i>R2-Score</i>	0,98	0,92

Pemodelan menggunakan variabel sentimen pada model *LSTM* memiliki performa yang lebih bagus daripada pemodelan tanpa menggunakan variabel sentimen. Selain itu, selisih antara pengujian menggunakan data latih dan data uji juga memiliki selisih yang tidak jauh antar keduanya, sehingga dapat dikatakan bahwa model *LSTM* sudah cukup bagus dalam memprediksi harga penutupan *bitcoin*, baik pada data latih maupun data uji.

3.2.3 Identifikasi Model Terbaik

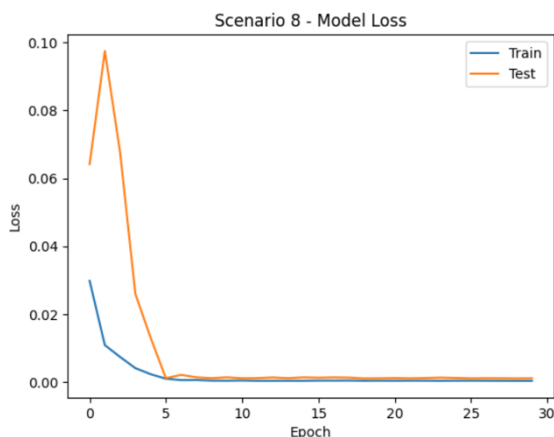
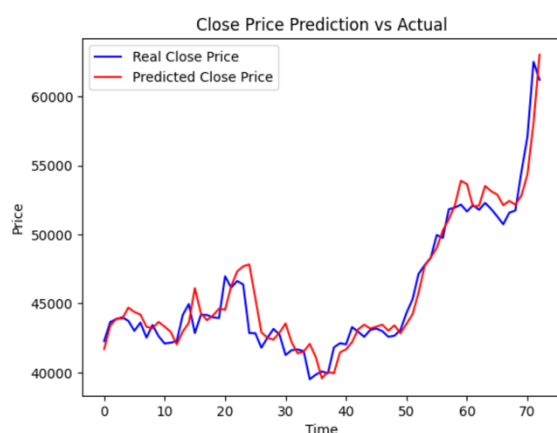
Dapat dilihat pada tabel 6, bahwa model *LSTM for regression* dengan penambahan variabel eksternal memiliki performa yang lebih unggul dibandingkan model lainnya. Hal tersebut dibuktikan dengan nilai *RMSE* yang lebih rendah, yang menunjukkan bahwa rata-rata kesalahan model dalam memprediksi harga penutupan *bitcoin* sebesar 1378,55. Selain itu nilai *R2-Score* pada skenario tersebut juga lebih tinggi daripada model dengan skenario lainnya, yaitu sebesar 92%. Nilai tersebut menunjukkan bahwa model mampu menjelaskan sekitar 92% dari variabilitas pada variabel yang diprediksi berdasarkan variabel input yang digunakan dalam model. Artinya, sebagian besar fluktuasi dalam data target dapat dijelaskan oleh variabel-variabel yang dimasukkan ke dalam model, sementara sisanya dipengaruhi oleh faktor lain yang tidak terwakili dalam model.

Tabel 6. Perbandingan performa model

Rank	Model	Matriks Evaluasi	
		<i>RMSE</i>	<i>R2-Score</i>
1	<i>LSTM</i> dengan variabel eksternal	1378,55	0,92
2	<i>LSTM</i> tanpa variabel eksternal	1632,28	0,89
3	<i>XGBoost</i> dengan variabel eksternal	5169,90	-0,13
4	<i>XGBoost</i> tanpa variabel eksternal	5407,22	-0,24

Pada Gambar 3, disajikan *learning curve* atau kurva pembelajaran kinerja model *LSTM* yang menggambarkan penurunan nilai *loss* selama proses pelatihan. Seiring bertambahnya *epochs*, nilai *loss* menurun tajam, baik untuk data latih maupun data uji yang mencerminkan peningkatan performa prediksi. Saat *epochs* mencapai iterasi ke-30, kurva stabil pada nilai *loss* yang rendah. Hal tersebut menunjukkan bahwa model telah mencapai keseimbangan dan tidak mengalami *overfitting* dengan performa yang konsisten pada data latih maupun data baru.

Pada Gambar 4 terlihat bahwa prediksi menggunakan model *LSTM* menghasilkan grafik perbandingan nilai aktual dan nilai prediksi harga penutupan *bitcoin* yang saling berdekatan. Garis berwarna biru mencerminkan nilai aktual harga penutupan *bitcoin*, sedangkan garis berwarna merah mencerminkan nilai prediksi dari model *LSTM*, dimana selisih antar keduanya tidak terlalu signifikan.

Gambar 3. Kurva pembelajaran algoritma *LSTM*

Gambar 4. Diagram garis nilai aktual dan nilai prediksi

Selain itu, model juga mampu menangkap pola *time series* pada data, dimana tren kenaikan dan penurunan harga *bitcoin* berhasil direplikasi dengan baik oleh prediksi model. Meskipun terdapat beberapa deviasi kecil antara nilai aktual dan prediksi, kesesuaian garis menunjukkan bahwa model berhasil mengidentifikasi fluktuasi harga dengan performa yang cukup baik, sehingga dapat digunakan untuk memprediksi pergerakan harga *bitcoin* di masa mendatang secara efektif.

3.3 Kontribusi Variabel Terhadap Harga Penutupan *Bitcoin*

Kontribusi variabel secara umum diidentifikasi menggunakan *SHAP* yang disajikan pada Gambar 5, dimana sumbu horizontal menunjukkan seberapa besar kontribusi setiap fitur, dengan nilai yang lebih ke kanan dapat meningkatkan prediksi, sementara yang lebih ke kiri dapat menurunkannya.

Dapat diidentifikasi bahwa fitur *Open* memiliki pengaruh paling dominan terhadap hasil prediksi. Hal ini ditunjukkan oleh area berwarna merah yang paling besar, menandakan kontribusi positif yang signifikan daripada fitur yang lain. Kemudian diikuti oleh fitur *High* dan *Low*, yang juga memberikan kontribusi positif terhadap prediksi model, meskipun dengan pengaruh yang lebih kecil. Sementara itu, fitur Sentimen, *Volume*, dan *Index* memiliki pengaruh

yang sangat kecil sehingga tidak terlihat dalam visualisasi *SHAP*. Perlu diperhatikan bahwa pengaruh fitur dalam konteks ini tidak menunjukkan kualitas model, melainkan sejauh mana fitur tersebut berkontribusi dalam menentukan nilai prediksi yang dihasilkan oleh model. Secara keseluruhan, fitur harga historis seperti *Open*, *High*, dan *Low* memiliki peran yang lebih dominan dalam membentuk hasil prediksi, sedangkan fitur eksternal memberikan kontribusi yang minimal dan cenderung menurunkan nilai prediksi.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, hasil menunjukkan bahwa model *XGBoost regression* terbaik diperoleh pada skenario penambahan variabel eksternal dengan kombinasi *hyperparameter* (*objective*: reg:squarederror, *learning_rate*: 0,02571, *max_depth*: 2, *n_estimators*: 494, *subsample*: 0,99729, *colsample_bytree*: 0,75292, *alpha*: 0,00027, *lambda*: 0,00030) yang menghasilkan nilai *RMSE* sebesar 5169,898 *USD* dan *R2-Score* sebesar -13%. Nilai tersebut mengindikasikan adanya *overfitting* pada model, karena selisih nilai *RMSE* dan *R2-Score* antara data latih dan data uji cukup tinggi, sehingga dapat dikatakan bahwa model *XGBoost regression* kurang cocok pada data penelitian ini yang bersifat *time series*. Model *LSTM for regression* pada skenario penambahan variabel eksternal dengan kombinasi *hyperparameter* (jumlah *layer*: 2, jumlah *neuron*: 150, *epoch*: 20, *optimizer*: *adam*) merupakan model terbaik dibandingkan model lainnya, yang dibuktikan dengan nilai *RMSE* terendah yaitu sebesar 1378,55 *USD* dan nilai *R2-Score* tertinggi yaitu sebesar 92%. Selain itu, model *LSTM for regression* tidak mengindikasikan adanya *overfitting* yang dibuktikan dengan kestabilan grafik *loss* pada kurva pembelajaran dengan nilai *loss* yang rendah saat *epoch* mencapai iterasi ke-20. Model juga mampu menangkap pola *time series* pada data, dimana tren kenaikan dan penurunan harga *bitcoin* berhasil direplikasi dengan baik oleh prediksi model. Hal tersebut menunjukkan bahwa *LSTM* memiliki kemampuan dalam mengenali ketergantungan temporal dan pola jangka panjang dalam data historis. Secara keseluruhan, variabel historis harga memberikan pengaruh yang signifikan dalam membentuk hasil prediksi harga penutupan *bitcoin* yang diidentifikasi menggunakan metode *SHAP*, dengan *Open* sebagai variabel paling dominan.

Untuk kedepannya, penelitian ini dapat diperluas dengan mengeksplorasi jenis algoritma regresi non-linear yang lain agar tidak berpotensi mengabaikan algoritma lain yang lebih efektif. Selain itu, dapat dilakukan penambahan dataset dengan jumlah yang lebih besar untuk meningkatkan performa model, terutama pada model *deep learning*.



Gambar 5. Diagram SHAP

DAFTAR PUSTAKA

- ADHERDA, D.T., HIKMATYAR, M. DAN RUUHWAN, 2023. Klasifikasi Gender Berdasarkan Suara Menggunakan Recurrent Neural Network (Rnn). *Jurnal Ilmiah Teknik Informatika*, 17(1), pp.111–122.
- ARDANA, A., 2023. Performance Analysis of XGBoost Algorithm to Determine the Most Optimal Parameters and Features in Predicting Stock Price Movement. *Telematika: Jurnal Informatika dan Teknologi Informasi*, 20(1), p.91. <https://doi.org/10.31315/telematika.v20i1.9329>.
- ARDIAN, M., KHOMSAH, S. DAN PANDIYA, R., 2024. Perbandingan Model Regresi Untuk Memprediksi Harga Jual Cabai Rawit Berdasarkan Iklim Harian. *Jurnal JUPITER*, [online] 16(2), pp.549–560. Tersedia di: <<https://hargajateng.org/tabel-harga->>.
- ARFAN, A. DAN ETP, L., 2019. Prediksi Harga Saham Di Indonesia Menggunakan Algoritma Long Short-Term Memory. *Universitas Gunadarma Jl. Margonda Raya No*, [online] 3(1). Tersedia di: <<https://www.ofx.com>>.
- ARISANDI, A.D., FERDIANSYAH DAN ATIKA, L., 2020. Prediksi Mata Uang Bitcoin Menggunakan LSTM Dan Sentiment Analisis Pada Sosial Media. *Jurnal Ilmiah Komputasi*, 19(4). <https://doi.org/10.32409/jikstik.19.4.370>.
- AYUNI, G.N. DAN FITRIANAH, D., 2019. Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti Pada PT XYZ. *Jurnal Telematika*, [online] 14(2), pp.79–86. Tersedia di: <<https://journal.ithb.ac.id/telematika/article/view/321>>.
- AZIZAH, S.R., HERTENO, R., FARMADI, A., KARTINI, D. DAN BUDIMAN, I., 2019. Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(6), pp.1361–1368. <https://doi.org/10.25126/jtiik.2023107467>.
- FADLURACHMAN, M.F., 2022. Prediksi Nilai Tukar Cryptocurrency Jangka Pendek dengan Menggunakan Long Short Term Memory (LSTM). *Universitas Hasanudin*.
- HABSYI, H.A., RAMDANI, C. DAN KHOMSAH, S., 2023. Perancangan Model Artificial Neural Network Backpropagation Untuk Prediksi Harga Bitcoin-US Dollar. *LEDGER: Journal Informatic and Information Technology*, 2(2).
- HUANG, Q., 2024. Bitcoin Price Prediction Based on Fear & Greed Index. In: *SHS Web of Conferences*. <https://doi.org/10.1051/shsconf/202418102015>.
- INDRIYANI, D. DAN USMAN, B., 2024. Pengaruh Google Trend dan Makroekonomi Terhadap Harga , Return , dan Volume Perdagangan Bitcoin. *JUKPEND: Jurnal Ekonomi dan Pendidikan*, 7(2), pp.19–28. <https://doi.org/10.26858/jekpend.v7i1.56193>.
- JANNAH, A.W., 2022. Perkembangan Hukum Positif Dan Hukum Islam Di Indonesia Terhadap Eksistensi Cryptocurrency. *Jatiswara*, 37(1), pp.127–140.
- KLAUDIA, S., ROHMAH, T.N., DEVI, Y.V. DAN AYU, C.R.L., 2018. Menakar Pengaruh Risiko, Return, Pemahaman Investasi, dan Modal Investasi Terhadap Minat UMKM dalam Memilih Jenis Investasi. *Jurnal Penelitian Teori & Terapan Akuntansi (PETA)*, 3(1), pp.109–124. <https://doi.org/10.51289/peta.v3i1.339>.
- LI, X., 2022. *Advances in Intelligent Automation and Soft Computing*. Springer. <https://doi.org/https://doi.org/10.1007/978-3-030-81007-8>.
- MALIKI, M.A., CHOLISSODIN, I. DAN YUDISTIRA, N., 2022. Prediksi Pergerakan Harga Cryptocurrency Bitcoin Terhadap Mata Uang Rupiah menggunakan Algoritme LSTM. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, [online] 6(7), pp.3259–3268. Tersedia di: <<https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11326%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/11326/5004>>.
- MARDIATMOKO, G., 2020. Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Berganda (Studi Kasus Penyusunan Persamaan Allometrik Kenari Muda [*Canarium Indicum L.*]). *BAREKENG: Jurnal Ilmu Matematika dan Terapan*,

- 14(3), pp.333–342.
<https://doi.org/10.30598/barekengvol14iss3pp333-342>.
- MAYAPADA, R., YANTI, R.W. DAN SYARIFUDDIN, S., 2022. Analisis Tingkat Kepentingan terhadap Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Indonesia. *Journal of Mathematics: Theory and Applications*, 4(2), pp.45–49.
- MOCH FARRYZ RIZKILLOH DAN SRI WIDIYANESTI, 2022. Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), pp.25–31.
<https://doi.org/10.29207/resti.v6i1.3630>.
- MUHAMMAD, M., LENIWATI, D., PRASETYO N.W, A.M., JUANDA, A., WAHYUNI, E.D. DAN SETYAWAN, S., 2023. Pengaruh Attitude Subjective Norms, dan Perceived Behavioural Control Terhadap Minat Investor Berinvestasi Cryptocurrency. *Jurnal Akuntansi dan Keuangan (JAK)*, 11(1), p.47.
<https://doi.org/10.29103/jak.v11i1.8372>.
- PADHILA, P.H., CHOLISSODIN, I. DAN ADIKARA, P.P., 2022. Prediksi Harga Bitcoin berdasarkan Data Historis Harian dan Google Trend Index menggunakan Algoritme Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, [online] 6(7), pp.3515–3524. Tersedia di: <<http://j-ptiik.ub.ac.id>>.
- PRADANA, N.F.B. DAN LESTANTI, S., 2020. Aplikasi Prediksi Jangka Pendek Harga Bitcoin Menggunakan Metode ARIMA. *Jurnal Ilmiah Informatika Komputer*, 25(3), pp.160–174.
<https://doi.org/10.35760/ik.2020.v25i3.3128>.
- PRATAMA, A.F., KURNIAWAN, T.B., MISINEM DAN DEWI, D.A., 2023. Implementasi Analisis Sentimen dan Model Deep Learning Untuk Prediksi Harga Bitcoin. *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, 15(1), pp.403–412.
- PRATAMA, M.L. DAN UTAMA, H., 2023. Pendekatan Deep Learning Menggunakan Metode LSTM Untuk Prediksi Harga Bitcoin. *The Indonesian Journal of Computer Science Research*, 2(2), pp.43–50.
- PRIMULANDO, R., 2021. Simulasi Karakterisasi Interaksi Wimp-Quarks di LHC dengan Menggunakan Deep Learning. In: *Prosiding Seminar Nasional Riset dan Teknologi Terapan (RITEKTRA) 2021*. pp.1–6.
- RASDI RERE, M.L., HARIYANTO DAN ROZI, 2022. Studi Prediksi Harga Bitcoin Menggunakan Recurrent Neural Network. In: *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)*. pp.149–155.
- RHAMADHANI, D.A. DAN SAPUTRI, E.E.D., 2021. Analisa Model Machine Learning dalam Memprediksi Laju Produksi Sumur Migas 15/9-F-14H. *Journal of Sustainable Energy Development*, 1(1), pp.48–55.
- RIYANTO, A.D., 2014. Pemanfaatan Google Trends Dalam Penentuan Kata Kunci Sebuah Produk untuk Meningkatkan Daya Saing Pelaku Bisnis di Dunia Internet. In: *Seminar Nasional Informatika 2014 (semnasIF 2014)*. pp.52–59.
- SUDARYANTO, S.N. DAN SUDARYANTO, 2022. Sintesis Fitur Density Based Feature Selection (DBFS) dan AdaBoots dengan XGBoost Untuk Meningkatkan Performa Model Prediksi. *Prosiding Seminar Nasional Sains dan Teknologi*, 12(1), p.305.
<https://doi.org/10.36499/psnst.v12i1.6997>.
- WANG, Y., PAN, Z., ZHENG, J., QIAN, L. DAN LI, M., 2019. A Hybrid Ensemble Method for Pulsar Candidate Classification. *Astrophysics and Space Science*, pp.1–15.
<https://doi.org/10.1007/s10509-019-3602-4>.
- WIRANDA, L. DAN SADIKIN, M., 2019. Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 8(3), pp.184–196.
- YULIA HAYUNINGTYAS, R. DAN SARI, R., 2022. Implementasi Data Mining Dengan Algoritma Multiple Linear Regression Untuk Memprediksi Penyakit Diabetes. *Jurnal Teknik Komputer AMIK BSI*, [online] 8(2), pp.174–180.
<https://doi.org/10.31294/jtk.v4i2>.
- ZUNDINA ULYA, F., RONY WIJAYA, A. DAN LARAS PUSPITA, P., 2023. Peramalan Harga Cabai dan Bawang di Pasar Tradisional Purwokerto dengan Model Autoregressive Integrated Moving Average (ARIMA). In: *Seminar Nasional Official Statistics 2023*. pp.757–766.