

ANALISIS KLASIFIKASI ARAH HARGA KRIPTO MENGGUNAKAN POHON KEPUTUSAN (DECISION TREE): PENGUJIAN KETERGANTUNGAN PADA FITUR LAGGED DAN INDIKATOR TEKNIS

Akhlis Munazilin^{*1}, M. Fakhri Cholil²

^{1,2}Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Ibrahimy Situbondo Jatim
Email: akhlistmunazilin@email.ac.id, ² fakhricholilik23@email.ac.id

SEJARAH ARTIKEL

Diterima: 15.12.2025

Direvisi: 29.12.2025

Publish: 29.12.2025



*Hak Cipta © 2025
Penulis: Ini adalah
artikel akses terbuka
yang didistribusikan
berdasarkan ketentuan
Creative Commons
Attribution 4.0
International License.*

ABSTRAK

Pasar aset kripto dicirikan oleh volatilitas tinggi, menjadikannya tantangan signifikan dalam prediksi pergerakan harga. Penelitian ini bertujuan untuk mengklasifikasikan arah pergerakan harga harian (Naik/Turun) dari data historis kripto ("btcdtr.csv") menggunakan algoritma Decision Tree. Proses Data Mining melibatkan pembersihan data, rekayasa fitur melalui penciptaan lagged features dan indikator teknis seperti Moving Average (MA), serta pembentukan variabel target Klasifikasi. Model yang dilatih menunjukkan Akurasi sebesar 60% pada data uji. Hasil evaluasi mengungkap adanya ketidakseimbangan kinerja: model sangat baik dalam memprediksi kelas Turun (Recall 0.88), namun lemah dalam memprediksi kelas Naik (Recall 0.32). Analisis Feature Importance menegaskan bahwa harga penutup hari ini (Terakhir) dan harga hari sebelumnya (Terakhir_lag1) adalah prediktor dominan, sementara indikator teknis (MA_5 dan Perubahan%) diabaikan oleh model sederhana ini. Disarankan penggunaan teknik oversampling dan model ensambel untuk meningkatkan sensitivitas prediksi kenaikan harga.

Kata Kunci: klasifikasi, data mining, pohon keputusan, kripto, prediksi harga, feature engineering.

ABSTRACT

Cryptocurrency markets are characterized by high volatility, making it a significant challenge in price movement prediction. This study aims to classify daily price movement direction (Up/Down) from historical crypto data ("btcdtr.csv") using a Decision Tree algorithm. The Data Mining process involves data cleaning, feature engineering through the creation of lagged features and technical indicators such as Moving Averages (MA), and the formation of a Classification target variable. The trained model demonstrated 60% accuracy on the test data. The evaluation results revealed a performance imbalance: the model performed very well in predicting the Down class (Recall 0.88), but poorly in predicting the Up class (Recall 0.32). Feature Importance analysis confirmed that today's closing price (Last) and the previous day's price (Last_lag1) were the dominant predictors, while technical indicators (MA_5 and Change%) were ignored by this simple model. The use of oversampling techniques and ensemble models is recommended to increase the sensitivity of price increase predictions.

Keywords: classification, data mining, decision trees, crypto, price prediction, feature engineering.

1. PENDAHULUAN

Volatilitas pasar kripto, terutama aset utama seperti Bitcoin, telah menarik perhatian luas dari akademisi dan praktisi keuangan. Fluktuasi harga yang cepat menawarkan potensi keuntungan tinggi, namun juga risiko yang substansial [1], [2], [3]. Memprediksi arah pergerakan harga merupakan tantangan yang kompleks karena pasar finansial dianggap sebagai sistem dinamis non-linear yang dipengaruhi oleh faktor internal dan eksternal, termasuk sentimen pasar dan berita global [4], [5], [6].

Dalam konteks ini, Data Mining dan Pembelajaran Mesin (Machine Learning) menawarkan pendekatan yang menjanjikan untuk mengungkap pola tersembunyi dalam data deret waktu keuangan [7], [8], [9]. Klasifikasi, sebagai

salah satu teknik utama Data Mining, bertujuan untuk memprediksi kelas diskrit—dalam hal ini, arah harga di masa depan (Naik atau Turun) [10], [11], [12].

Penelitian terdahulu telah mengaplikasikan berbagai algoritma, mulai dari Support Vector Machine (SVM) hingga Jaringan Saraf Tiruan (Neural Networks) untuk prediksi pasar saham dan mata uang [13], [14], [15]. Algoritma Decision Tree dipilih dalam studi ini karena kemampuannya yang intuitif, cepat, dan kemudahan interpretasi terhadap aturan keputusan yang dihasilkannya, meskipun seringkali memiliki risiko overfitting [10], [16], [17].

Tujuan utama dari penelitian ini adalah: (1) Mengembangkan model Klasifikasi Decision Tree untuk memprediksi arah harga harian aset kripto, (2) Mengevaluasi kinerja model menggunakan metrik standar klasifikasi, dan (3) Mengidentifikasi fitur-fitur historis yang paling berpengaruh dalam keputusan prediksi model [18], [19]. Hasil ini diharapkan dapat memberikan wawasan mengenai ketergantungan prediksi pasar kripto terhadap data harga murni dibandingkan indikator teknis yang direkayasa [14].

2. METODE PENELITIAN

Metode penelitian mengikuti tahapan Knowledge Discovery in Databases (KDD), berfokus pada pra-pemrosesan data deret waktu dan implementasi model Klasifikasi Decision Tree [20].

2.1. Sumber Data dan Pra-pemrosesan

Data yang digunakan adalah kumpulan data historis harga kripto harian ("btccidr.csv") yang mencakup kolom: Tanggal, Terakhir, Pembukaan, Tertinggi, Terendah, Vol., dan Perubahan%. Data ini harus melalui langkah pembersihan kritis [21].

1. Konversi Format Numerik: Nilai harga dikonversi dari format Indonesia (pemisah ribuan: titik, desimal: koma) menjadi format numerik standar float.
2. Penanganan Nilai Hilang: Kolom Vol. yang memiliki nilai hilang diabaikan untuk menjaga kualitas data, dan baris yang dihasilkan hilang akibat lagging dihilangkan (dropna).

2.2. Rekayasa Fitur (Feature Engineering)

Pembentukan variabel independen (Features) dan variabel dependen (Target) dilakukan untuk mendukung analisis Klasifikasi [22].

1. Variabel Target (Y): Kolom Arah_Harga dibentuk dengan membandingkan harga Terakhir hari $t+1$ dengan hari t . Nilai 1 diberikan jika harga Naik, dan 0 jika harga Turun atau Stagnan.
2. Fitur Lagged: Nilai harga penutup sebelumnya (Terakhir_lag1) dibuat untuk menangkap ketergantungan urutan waktu [23].
3. Indikator Teknis: Moving Average 5 hari (MA_5) dihitung sebagai fitur tren sederhana.
4. et Fitur (X): Terdiri dari: Terakhir, Terakhir_lag1, Pembukaan, Tertinggi, Terendah, MA_5, dan Perubahan%.

2.3. Implementasi Model dan Evaluasi

Data di bagi menjadi training Set (80%) dan Testing Set (20%) dengan menjaga urutan waktu(shuffle=False) [19].

1. Model: Digunakan DecisionTreeClassifier dengan pembatasan kompleksitas max_depth =3 untuk mencegah overfitting dan memfasilitasi interpretasi aturan [24].
2. Evaluasi: Kinerja model diukur menggunakan metrik Klasifikasi standar [25]:
 1. Akurasi: Tingkat prediksi benar secara keseluruhan.
 2. Precision: Proporsi prediksi positif yang benar.
 3. Recall (Sensitivitas): Proporsi kasus positif yang sebenarnya berhasil ditangkap.
 4. F1-Score: Rata-rata harmonis Precision dan Recall.
 5. Feature Importance: Mengidentifikasi bobot kontribusi setiap fitur dalam proses splitting keputusan.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dapat diuraikan mengenai hasil dari penelitian beserta pengujian yang telah dilakukan. Selain itu, disampaikan juga mengenai pembahasan dari penelitian maupun pengujian yang telah dilakukan.

3.1. Hasil Evaluasi Kinerja Klasifikasi

Tabel 1 menunjukkan ringkasan kinerja model Decision Tree pada Testing Set (Total Support =35):

Metrik	Kelas 0 (Turun/Stagnan)	Kelas 1 Naik	Weighted Avg
Precision	0.58	0.69	0.65
Recal	0.88	0.32	0.60
F1-Score	0.70	0.44	0.59
Accuracy	-	-	0.6000

Model mencapai Akurasi sebesar 60.00%, yang secara statistik menunjukkan kemampuan prediksi yang sedikit lebih baik daripada peluang acak (50%) [26]. Namun, hasil ini juga menyoroti masalah ketidakseimbangan kinerja kelas:

1. Kelas 0 (Turun/Stagnan): Memiliki Recall yang sangat tinggi (0.88), menandakan bahwa model sangat baik dalam mengidentifikasi sebagian besar kasus di mana harga akan tetap stabil atau turun.
2. Kelas 1 (Naik): Memiliki Recall yang sangat rendah (0.32). Hal ini menunjukkan model memiliki kecenderungan tinggi untuk False Negative (gagal memprediksi kenaikan padahal harga sebenarnya naik), yang dapat menyebabkan kerugian peluang investasi.

Tingginya Precision untuk Kelas 1 (0.69) menunjukkan bahwa meskipun model jarang memprediksi kenaikan, ketika ia memprediksinya, prediksinya cenderung benar.

3.2. Analisis Pentingnya Fitur

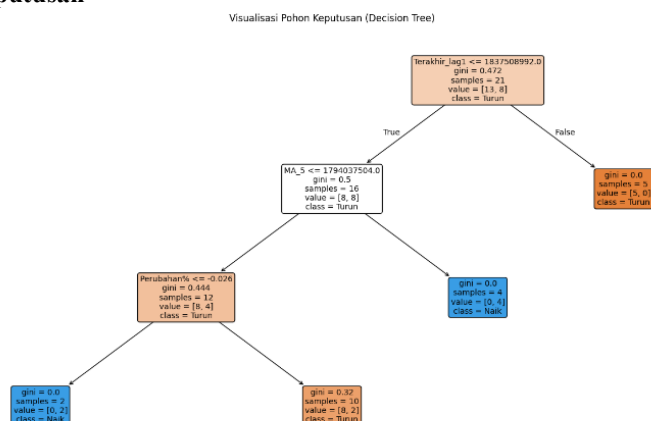
Hasil analisis Feature Importance (Tabel 2) menunjukkan fitur-fitur yang dominan dalam pohon keputusan:

Tabel 2. Pentingnya Fitur (Feature Importance)

Fitur	Pentingnya (Nilai)
Terakhir	0.5828
Terakhir_lagi	0.3013
Tertinggi	0.1159
MA_5	0.0000
Pembukaan	0.0000
Terendah	0.0000
Perubahan%	0.0000

Hasil ini menegaskan bahwa model Klasifikasi Decision Tree sangat bergantung pada nilai harga penutup saat ini (Terakhir) dan harga penutup kemarin (Terakhir_lagi) [27]. Secara kolektif, kedua fitur ini menyumbang lebih dari 88% dari total kepentingan dalam struktur pohon keputusan. Menariknya, indikator teknis seperti MA_5 dan Perubahan% memiliki nilai kepentingan 0.0000. Hal ini mengindikasikan bahwa, pada kedalaman pohon yang terbatas ($\text{max_depth}=3$), model tidak menemukan nilai diskriminatif yang cukup pada fitur-fitur yang direkayasa tersebut untuk memisahkan data, dan lebih memilih split pada harga murni. Temuan ini sejalan dengan hipotesis pasar semi-efisien, di mana informasi historis yang kompleks mungkin tidak secara langsung memberikan keunggulan prediksi yang signifikan bagi model sederhana [28].

3.3. Visualisasi Pohon Keputusan



Gambar 1. Visualisasi Pohon Keputusan (Decision Tree)

Visualisasi dari pohon keputusan yang dihasilkan dengan $\text{max_dept} = 3$ (gambar 1) memberikan interpretasi yang transparan mengenai proses pengambilan keputusan model.

- a. Node Akar dan Dominasi Harga Murni
Node Akar (tingkat 1) menunjukkan bahwa pemisahan pertama di dasarnya berdasarkan pada fitur terakhir. Hal ini secara visual mendukung hasil feature importance (Tabel 2) karena fitur ini menyumbang nilai gini impurity terbesar, menjadikannya kriteria pemisahan paling diskriminatif.
- b. Interpretasi aturan keputusan
Dengan kedalaman yang terbatas, pohon keputusan menghasilkan serangkaian aturan sederhana. Sebagai contoh interpretasi dari jalur kiri (True) pada node akar:

$IF (Terakhir \leq Threshold X) AND (Terakhir_{lag1} \leq Threshold Y) \rightarrow Prediksi\ Kelas\ 0 (TURUN)$
Aturan ini menyiratkan bahwa ketika harga penutup hari ini berada di bawah ambang batas tertentu, dan harga kemarin juga rendah, model cenderung memprediksi kesinambungan tren turun atau stagnasi.

Sebaliknya, jalur kanan (False) pada node akar sering mengarah ke prediksi Kelas 1 (Naik), namun dengan jumlah sampel yang lebih kecil dan tingkat Gini yang lebih tinggi di leaf nodes, menjelaskan mengapa Recall untuk kelas 'Naik' (0.32) sangat rendah. Model Decision Tree, dalam konfigurasi ini, cenderung memerlukan bukti yang jauh lebih kuat dan spesifik (berdasarkan perbandingan harga murni) untuk memprediksi kenaikan, dibandingkan dengan prediksi penurunan.

c. Keterbatasan Fitur Teknis

Secara visual, Decision Tree hanya menggunakan fitur Terakhir, Terakhir_lag1, dan Tertinggi untuk melakukan pemisahan data hingga mencapai kedalaman maksimum (3). Fitur-fitur seperti MA_5, Pembukaan, Terendah, dan Perubahan% tidak muncul dalam struktur pohon. Hal ini membuktikan bahwa, dalam konteks model non-ensambel dengan kedalaman dangkal, rekayasa fitur sederhana seperti Moving Average tidak memberikan nilai informasi tambahan dibandingkan data harga langsung, yang menjadi temuan signifikan terkait efektivitas Feature Engineering dalam prediksi deret waktu kripto.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Decision Tree dapat diterapkan untuk mengklasifikasikan arah pergerakan harga harian aset kripto dengan tingkat akurasi sebesar 60%, yang sedikit lebih baik dibandingkan prediksi acak. Namun demikian, hasil evaluasi mengungkap adanya ketidakseimbangan kinerja antar kelas, di mana model sangat efektif dalam memprediksi kondisi Turun/Stagnan (Recall 0.88), tetapi memiliki sensitivitas yang rendah terhadap kondisi Naik (Recall 0.32). Hal ini mengindikasikan kecenderungan model untuk menghasilkan *false negative* pada peristiwa kenaikan harga, yang berpotensi mengurangi manfaat praktisnya dalam konteks pengambilan keputusan investasi.

Analisis feature importance menunjukkan bahwa prediksi model didominasi oleh data harga murni, khususnya harga penutup saat ini (Terakhir) dan harga penutup hari sebelumnya (Terakhir_lag1), yang secara kolektif menyumbang lebih dari 88% kontribusi dalam struktur pohon keputusan. Sebaliknya, indikator teknis yang direkayasa, seperti Moving Average 5 hari (MA_5) dan Perubahan Persentase, tidak memberikan kontribusi berarti dalam proses pengambilan keputusan model pada konfigurasi pohon dengan kedalaman terbatas. Temuan ini mengindikasikan bahwa, untuk model Decision Tree sederhana, informasi historis harga langsung lebih informatif dibandingkan indikator teknis dasar.

Secara keseluruhan, penelitian ini menegaskan bahwa interpretabilitas tinggi Decision Tree memungkinkan pemahaman yang jelas terhadap mekanisme prediksi, namun efektivitasnya masih terbatas dalam menangkap pola kenaikan harga pada pasar kripto yang sangat volatil. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan kelas, memperluas rekayasa fitur, serta mengeksplorasi model ensambel atau metode non-linear yang lebih kompleks guna meningkatkan sensitivitas prediksi dan kinerja klasifikasi secara keseluruhan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Akhlis Munazilin atas bimbingan, arahan, serta masukan yang diberikan selama proses penelitian dan penyusunan artikel ini. Ucapan terima kasih juga disampaikan kepada Universitas Ibrahimy Situbondo, khususnya Fakultas Sains dan Teknologi, atas dukungan akademik dan fasilitas yang telah diberikan sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] EUGENE F. FAMA, "EFFICIENT CAPITAL MARKETS: A REVIEW OF OF THEORY AND EMPIRICAL WORK," pp. 383–417.
- [2] L. Catania, "Bitcoin at High Frequency," no. 2017, pp. 1–20, 2019, doi: 10.3390/jrfm12010036.
- [3] Z. Muhidin, N. Karim, and M. M. Efendi, "Analysis of Splicing Manipulation in Digital Images using Dyadic Wavelet Transform (DyWT) and Scale Invariant Feature Transform (SIFT) Methods," vol. 8, no. 2, pp. 408–412, 2024.
- [4] D. K. Gode and S. Sunder, "Allocative Efficiency of Markets with Zero-intelligence Traders : Market as a Partial Substitute for Individual Rationality," vol. 101, no. 1, 1993.
- [5] S. V Tharun, G. Saranya, T. Tamilvizhi, and R. Surendran, "Cryptocurrency Price Prediction using Deep Learning," pp. 1–18.
- [6] Wenti Ayu Wahyuni, M. N. Karim, and Z. Muahidin, "Penerapan algoritma xgboost dalam penentuan potensi sektor wisata lokal di lombok timur," vol. 03, pp. 155–160, 2025.
- [7] J. Han, Micheline Kamber, and Jian Pei, *Data Mining Concepts and Techniques*.

- [8] "Machine Learning for Financial Forecasting Abstract," no. M1, pp. 240–257.
- [9] R. Forest *et al.*, "A Comparative Hybrid Approach for Python Bug Detection Using Syntactic Features ," vol. 19, no. 2, pp. 141–150, 2025.
- [10] J. Ross, Q. Morgan, and K. Publishers, "Programs for Machine Learning," vol. 240, pp. 235–240, 1994.
- [11] R. Timofeev, "Classification and Regression Trees (CART) by," 2004.
- [12] B. Norma, A. Abella, M. Multazam, Z. Hadi, and Z. Muahidin, "KLASIFIKASI PENYAKIT TANAMAN TEMBAKAU MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK (CNN) BERBASIS WEB," vol. 3, no. 2, pp. 74–80, 2025.
- [13] K. Kim, "Financial time series forecasting using support vector machines," vol. 55, pp. 307–319, 2003, doi: 10.1016/S0925-2312(03)00372-2.
- [14] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep Learning for Stock Market Prediction," pp. 1–22, 2020, doi: 10.3390/e22080840.
- [15] Biva Candra Lutfi Adiatma, Z. Muahidin, and kusrini, "Sistem Pemilihan Ruang Rawat Inap Menggunakan Metode Weighted Product dan K-Nearest Neighbor," pp. 1–11.
- [16] S. Rasoul Safavian and David Landgrebe, "A Survey of Decision Wee Classifier Methodology," vol. 21, no. 3, 1991.
- [17] Z. Muahidin, "PENERAPAN ALGORITMA COLLABORATIVE FILTERING PADA SISTEM INFORMASI PENYEWAAN JASA TOUR GUIDE BERBASIS WEB DI WISATA BENANG KELAMBU," vol. 2, no. 2, pp. 83–88, 2024.
- [18] T. Lim, "A Comparison of Prediction Accuracy , Complexity , and Training Time of Thirty-three Old and New Classification Algorithms," vol. 229, pp. 203–229, 2000.
- [19] M. N. Karim, M. M. Efendi, and Z. Muahidin, "PENERAPAN ARTIFICIAL NEURAL NETWORK DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI KUALITAS MUTIARA KHAS LOMBOK BERDASARKAN CIRI VISUAL," *PENERAPAN Artif. NEURAL Netw. DAN Support VECTOR Mach. UNTUK KLASIFIKASI KUALITAS MUTIARA KHAS Lomb. BERDASARKAN CIRI Vis.*, vol. 3, no. 1, pp. 39–47, 2025.
- [20] U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in," vol. 17, no. 3, pp. 37–54, 1996.
- [21] D. Pyle, S. Editor, and D. D. Cerra, *Data Preparation for Data Mining*. 1999.
- [22] H. R. Varian, "Big Data: New Tricks for Econometrics †," vol. 28, no. 2, pp. 3–28, 2014.
- [23] S. Ataei, S. T. Ataei, and A. M. Saghiri, "Applications of Deep Learning to Cryptocurrency Trading : A Systematic Analysis," vol. 2, pp. 1–38, 2025.
- [24] Trevor Hastie, R. Tibshirani, and Jerome H. Friedman, "The Elements of Statistical Learning".
- [25] Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone, "Classification and Regression Trees," vol. 535, pp. 534–535, 1987.
- [26] Thomas G. Dietterich, "Approximate Statistical Test for Comparing Supervised Classification Learning Algorithms".
- [27] J. Li and S. Wang, "Feature Selection: A Data Perspective," pp. 1–73.
- [28] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of Technical Analysis : Computational Algorithms , Statistical Inference , and Empirical Implementation," vol. LV, no. 4, pp. 1705–1765, 2000.