

Hidden Markov Model untuk Deteksi Rezim dan Prediksi Kejatuhan Harga pada Bitcoin: Evaluasi Empiris Regime-Aware Stacking Ensemble

Bagus Satria Putra^{1*}, Kumara Ari Yuana²

¹Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta, Indonesia

²Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia

Corresponding Author's e-mail : bagus.satria.putra@students.amikom.ac.id



e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

<https://ejournal.45mataram.ac.id/index.php/armada>

Vol. 04, No. 10 Oktober, 2026

Page: 6104-6116

DOI:

<https://doi.org/10.55681/armada.v4i10.4073>

Article History:

Received: Agustus 02, 2026

Revised: September 08, 2026

Accepted: Oktober 18, 2026

Abstract : *Because Bitcoin swings so violently, the prospect of a sudden collapse looms large for investors, yet most forecasting schemes hold the tie between predictors and crash probability fixed regardless of market conditions. This study asks whether conditioning a predictor on latent market regimes sharpens Bitcoin crash forecasting. A regime-aware pipeline unites a three-state Gaussian Hidden Markov Model for regime detection, a stacking ensemble pairing XGBoost with a Long Short-Term Memory network for classification, and SHAP for regime-wise interpretation, evaluated on 2,884 daily records from 2018 to 2025 across seventeen technical, on-chain, and sentiment features. A crash is a cumulative fall beyond fifteen percent within seven days, with a ten-percent cutoff for sensitivity testing. Crash incidence diverges sharply across Bull, Bear, and Sideways regimes (Kruskal-Wallis $p = 0.000289$; Bear-to-Bull odds ratio = 1.70). Yet no regime-aware design beats a regime-blind baseline: where power is adequate the best variant adds only 0.0092 ± 0.0017 (mean of three runs) to the precision-recall AUC at Wilcoxon p in the 0.81–1.00 range, and a lone XGBoost outperforms every stacked variant. Regime detection is corroborated as a diagnostic risk gauge rather than a predictive booster.*

Keywords : *Bitcoin; Crash Forecasting; Hidden Markov Model; Investor Risk Management; SHAP*

Abstrak : Ayunan harga Bitcoin yang keras membuat kemungkinan keruntuhan mendadak membayangi investor, namun sebagian besar skema peramalan mematok kaitan antara prediktor dan peluang crash sebagai sesuatu yang tetap tanpa memandang kondisi pasar. Studi ini menguji apakah pengondisian prediktor pada regime pasar laten mempertajam peramalan crash Bitcoin. Alur regime-aware menyatukan Hidden Markov Model Gaussian tiga state untuk deteksi regime, stacking ensemble yang memasangkan XGBoost dengan jaringan Long Short-Term Memory untuk klasifikasi, dan SHAP untuk interpretasi per regime, dievaluasi pada 2.884 catatan harian 2018–2025 dengan tujuh belas fitur teknikal, on-chain, dan sentimen. Crash didefinisikan sebagai penurunan kumulatif di atas lima belas persen dalam tujuh hari, dengan ambang sepuluh persen untuk uji sensitivitas. Kejadian crash berbeda tajam antar regime Bull, Bear, dan Sideways (Kruskal-Wallis $p = 0,000289$; rasio odds Bear terhadap Bull = 1,70). Namun tak ada rancangan regime-aware yang mengalahkan baseline buta regime: pada kondisi berdaya memadai, varian terbaik hanya menambah $0,0092 \pm 0,0017$ (rata-rata tiga

eksekusi) pada AUPRC dengan Wilcoxon p pada rentang 0,81–1,00, dan XGBoost tunggal mengungguli seluruh varian bertumpuk. Deteksi regime terbukti sebagai pengukur diagnostik risiko, bukan penambah daya prediksi.

Kata Kunci : Bitcoin; Peramalan Crash; Hidden Markov Model; Manajemen Risiko Investor; SHAP

PENDAHULUAN

Bitcoin lahir sebagai percobaan kriptografi dan kini menjadi instrumen yang digenggam jutaan pihak, dari investor perorangan hingga lembaga besar, meski watak harganya yang liar boleh dikatakan tak bergeser. Sepanjang riwayatnya nilai aset ini pernah tergerus lebih dari delapan puluh persen pada sejumlah kesempatan; salah satu contoh mencolok ialah jatuhnya harga dari sekitar USD 69.000 pada November 2021 menuju kisaran USD 16.000 setahun kemudian (Liu et al., 2024). Bursa saham memiliki penghenti perdagangan sementara untuk meredam kepanikan, tetapi perdagangan kripto tak mengenal jeda dan berlangsung sepanjang waktu, sehingga koreksi tajam dapat meletus kapan saja dengan skala yang lebih besar. Dari sudut pandang investor, seberapa baik eksposur dapat dijaga sangat bergantung pada kesigapan membaca tanda penurunan tersebut. Persoalan ini juga menyentuh ranah kebijakan di Indonesia. Undang-Undang Nomor 4 Tahun 2022 tentang Pengembangan dan Penguatan Sektor Keuangan memindahkan kewenangan pengaturan dan pengawasan aset keuangan digital, termasuk aset kripto, ke Otoritas Jasa Keuangan; dengan perpindahan itu, melindungi investor perorangan atas aset berisiko tinggi resmi masuk agenda pengawasan, sehingga alat diagnostik risiko yang teruji empiris menjadi relevan bukan hanya bagi pelaku pasar melainkan juga otoritas.

Dalam kajian ini crash didefinisikan operasional sebagai penurunan harga kumulatif melampaui lima belas persen dalam horizon tujuh hari ke depan, mengambil posisi tengah rentang yang lazim di literatur di antara Epsilon Drawdown Method adaptif ala Liu et al. (2024) dan ambang lima belas persen berhorizon tiga minggu ala Shu dan Zhu (2024). Perlu ditegaskan pembedaan crash dari regime pasar: regime menggambarkan keadaan laten yang melatari gerak pasar bull dengan tren naik, bear dengan tren turun, dan sideways dengan volatilitas rendah tanpa tren sedangkan crash adalah kejadian diskrit yang dapat pecah di regime mana pun lewat pemicu berbeda, dari gelembung spekulatif yang meletus (Minsky, 1992) hingga gelombang latah dan panik jual (Bikhchandani & Sharma, 2001).

Sampai kini, sebagian besar riset peramalan crash kripto masih memakai satu model yang menganggap hubungan fitur dengan peluang crash bersifat baku (Chandra & He, 2024). Cara pandang ini bermasalah sebab pertanda sebelum harga runtuh saat pasar menanjak sama sekali lain dari pertanda saat pasar merosot; model yang dilatih atas seluruh data tanpa pembeda kondisi membentuk batas keputusan berupa kompromi lintas kondisi, sehingga tumpul menangkap crash pada situasi tertentu walau akurasi keseluruhannya tampak baik. Di sisi lain, Hidden Markov Model terbukti efektif memisahkan keadaan pasar kripto yang tak kasat mata (Pennoni et al., 2022; Giudici & Polinesi, 2021; Machimbo et al., 2025), sementara XGBoost tangguh mengolah data tabular (Chen & Guestrin, 2016), dan stacking memungkinkan peleburan model heterogen lewat meta-learner (Ji et al., 2019). Data on-chain merekam langsung aktivitas transaksi pada jaringan Bitcoin dan diyakini menyingkap perilaku pemegang aset yang tak tampak pada data pasar biasa; metrik MVRV, misalnya, secara historis kerap mendahului koreksi ketika melampaui 3,5 (Sentiment Academy, 2023). Namun, sejauh penelusuran penulis, belum dijumpai kajian yang serentak memadukan tiga komponen HMM untuk deteksi regime, stacking ensemble atas model heterogen berbasis pohon dan jaringan saraf, serta data on-chain dengan interpretasi SHAP per regime. Kekosongan itulah dasar kebaruan penelitian ini.

Penelitian ini merumuskan tiga pertanyaan. Pertama, bagaimana HMM mengenali tiga regime pasar Bitcoin dan bukti statistik apa yang menopang perbedaan risiko crash antar regime. Kedua, apakah pengintegrasian informasi regime ke dalam Stacking Ensemble (XGBoost + LSTM) mempertajam performa prediksi crash dibanding model tunggal dan baseline. Ketiga, sejauh mana penggerak prediksi crash bergeser antar regime menurut analisis SHAP, dan apakah pergeseran itu

memvalidasi atau menantang teori tekanan pasar. Sejalan dengan itu, tiga hipotesis diajukan: H1 menyatakan laju crash berbeda nyata antar regime; H2 menyatakan arsitektur regime-aware menaikkan AUPRC minimal 0,03 atas baseline; dan H3 menyatakan kepentingan fitur (nilai SHAP) berbeda nyata antar regime pada minimal tiga dari lima fitur teratas. Kontribusi utama artikel ini adalah pengujian empiris yang berterus terang melaporkan hasil negatif, dilengkapi analisis sensitivitas dua ambang untuk memisahkan temuan tak konklusif akibat kelangkaan kejadian dari penolakan hipotesis pada kondisi berdaya memadai.

Sebagai landasan, Hidden Markov Model memodelkan sistem sebagai proses Markov dengan state tak teramati langsung (Rabiner, 1989); pada pasar finansial, hidden state mewakili kondisi laten (bear, sideways, bull) yang memengaruhi return dan volatilitas teramati. XGBoost mengimplementasikan gradient boosting teroptimasi (Chen & Guestrin, 2016), LSTM mengatasi vanishing gradient pada data sekuensial (Hochreiter & Schmidhuber, 1997), dan SHAP menghitung kontribusi marginal fitur berbasis nilai Shapley (Lundberg & Lee, 2017; Lundberg et al., 2018). Kerangka teori tekanan pasar hipotesis instabilitas finansial Minsky (1992), keuangan keperilakuan (Baker & Wurgler, 2007), dan analisis on-chain menjadi dasar pemetaan hasil SHAP per regime.

METODE PENELITIAN

Penelitian ini bercorak kuantitatif dengan pendekatan eksperimental komparatif, tergolong applied research yang membangun model prediksi crash Bitcoin sekaligus explanatory research yang menjelaskan faktor pengaruh lewat SHAP (Lundberg & Lee, 2017). Secara metodologis, unsupervised learning untuk deteksi regime melalui HMM dipadukan dengan supervised learning untuk klasifikasi biner crash/non-crash. Target berupa label diskrit crash (1) atau non-crash (0) sehingga tugasnya klasifikasi, bukan regresi. Pilihan biner atas regresi diambil karena peramalan besaran harga kripto tenggelam dalam derau, dan biner atas multi-kelas severity karena porsi crash yang cuma 3,08 persen akan terpecah menjadi kelas beranggota di bawah 30 sampel. Base learner menghasilkan probabilitas kontinu yang di-threshold 0,5, dengan binary cross-entropy sebagai fungsi rugi dan AUPRC sebagai metrik utama. Crash ditandai sebagaimana Persamaan (1).

$$\text{crash}_t = 1 \text{ jika } (P_{t+7} - P_t) / P_t \leq \tau ; \text{selainnya } 0$$

Data dihimpun terprogram dari tiga sumber: Yahoo Finance untuk OHLCV Bitcoin (ticker BTC-USD), CoinMetrics Community API untuk metrik on-chain (MVRV, alamat aktif, jumlah transaksi, NVT adjusted, exchange outflow), dan Alternative.me untuk Fear & Greed Index skala 0–100. Periode 31 Januari 2018 hingga 23 Desember 2025 pada resolusi harian menghasilkan 2.884 observasi yang merangkum bear market 2018–2019, COVID crash 2020, bull run 2020–2021, crash 2022, dan pemulihan 2023–2025; resolusi harian dipilih karena metrik on-chain hanya tersedia pada tingkat tersebut. Rangkuman sumber ada pada Tabel 1.

Tabel 1. Sumber dan Jenis Data

Sumber	Jenis Data	Metode	Periode
Yahoo Finance	OHLCV	yfinance API	2018–2025
CoinMetrics	On-Chain Metrics	REST API	2018–2025
Alternative.me	Sentiment (FGI)	REST API	2018–2025

Sumber: Diolah penulis (2026)

Tujuh belas fitur disusun dalam tiga kategori: delapan indikator teknikal (RSI, MACD histogram, posisi Bollinger, rasio volume, log-return, volatilitas 7 dan 30 hari, perubahan volume), lima metrik on-chain (skor-z MVRV, NVT, alamat aktif, exchange flow, jumlah transaksi), dan empat indikator sentimen (FGI, rerata bergerak 7 hari FGI, momentum sentimen, interaksi sentimen-volume). Rasio observasi-per-fitur sekitar 170:1 jauh di atas batas minimum yang disarankan (Hastie et al., 2009). Normalisasi z-score dihitung hanya dari training set tiap fold untuk mencegah kebocoran data. Regime dimodelkan dengan Gaussian HMM bertiga state tersembunyi atas log-return dan volatilitas 7 hari, dengan estimasi Baum-Welch dan dekoding

Viterbi sebagaimana Persamaan (2). Jumlah state ditetapkan tiga bukan lewat minimisasi kriteria informasi (BIC/AIC menurun monoton), melainkan atas pemetaan ekonomi jelas, preseden literatur, dan keseimbangan kehematan-ketertafsiran; $k=2,4,5$ diuji sebagai sensitivitas.

$$P(o_t | s_t = i) = N(o_t; \mu_i, \Sigma_i)$$

Klasifikasi crash memakai ensemble bertingkat yang memadukan XGBoost (pembelajar pohon) dan LSTM (pembelajar sekuensial). Pada tiap fold, keduanya menghasilkan prediksi luar-lipatan berupa probabilitas crash yang menjadi masukan meta-learner berupa kombinasi cembung dengan satu bobot dicari lewat grid search; pilihan ini menggantikan Logistic Regression (Ji et al., 2019) yang rawan overfit pada kelas minoritas ekstrem. Tiga konfigurasi integrasi regime dibandingkan (Tabel 2): Stacking A tanpa regime, Strategi 1 (label regime sebagai fitur one-hot), dan Strategi 2 (submodel spesifik regime). XGBoost tunggal menjadi acuan performa terbaik, sedangkan Stacking A menjadi baseline pengujian H2.

Tabel 2. Strategi Integrasi Regime

Strategi	Deskripsi	Mekanisme
Strategi 1	Feature-Level Concatenation	One-hot regime di-concat ke 17 fitur (20 fitur) sebagai input satu Stacking Ensemble
Strategi 2	Regime-Specific Submodels	Tiga Stacking Ensemble terpisah (Bull, Bear, Sideways), routing berdasar keluaran HMM

Sumber: Diolah penulis (2026)

Pelatihan dan evaluasi memakai validasi silang jendela mengembang berjeda bersih tujuh hari untuk mencegah kebocoran label. Metrik utama AUPRC dipilih karena akurasi dan AUC-ROC menyedatkan pada kelas sangat timpang (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015), dengan ambang praktis kenaikan 0,03. Ketimpangan kelas (sekitar 30:1) ditangani lewat pembelajaran peka-biaya (He & Garcia, 2009); SMOTE dihindari karena interpolasi ruang fitur dapat melanggar urutan temporal. H1 diuji dengan Kruskal-Wallis, uji lanjut Dunn, dan rasio odds; H2 dengan Wilcoxon signed-rank atas selisih AUPRC antar fold; H3 dengan Kruskal-Wallis berkoreksi Bonferroni atas nilai SHAP. Ablation study lima konfigurasi (A: baseline; B: regime one-hot; C: submodel per regime; D: tanpa on-chain; E: tanpa sentimen) melengkapi analisis.

HASIL DAN PEMBAHASAN

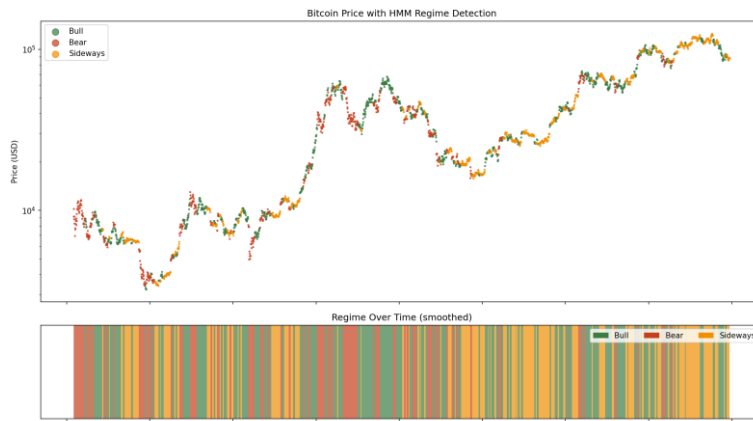
Deskripsi Data dan Deteksi Regime

Dataset final memuat 2.884 observasi harian dengan 88 crash pada ambang -15 persen (3,05 persen); sensitivitas memberi 244 crash pada -10 persen dan 44 crash pada -20 persen. HMM tiga state memisahkan observasi ke regime Bull (43 persen, return positif kecil, volatilitas terendah), Sideways (34 persen, return marginal, volatilitas medium), dan Bear (23 persen, return negatif, volatilitas tertinggi), sebagaimana Tabel 3 dan Gambar 1. Matriks transisi memperlihatkan persistensi tinggi (diagonal 0,90–0,91); perpindahan langsung Bull ke Bear amat jarang (0,0027) sehingga transisi tipikal melewati Sideways sebagai penyangga.

Tabel 3. Karakteristik Statistik Tiga Regime Bitcoin (2018–2025)

Regime	n	% Total	Mean Log Return	Std Log Return	Mean Vol-30d
Bull	1.239	42,96%	+0,0012	0,0148	0,0224
Sideways	985	34,15%	+0,0013	0,0291	0,0312
Bear	660	22,89%	−0,0010	0,0555	0,0445

Sumber: Diolah penulis (2026)



Gambar 1. Visualisasi Regime Detection Bitcoin 2018–2025
Sumber: Diolah penulis (2026)

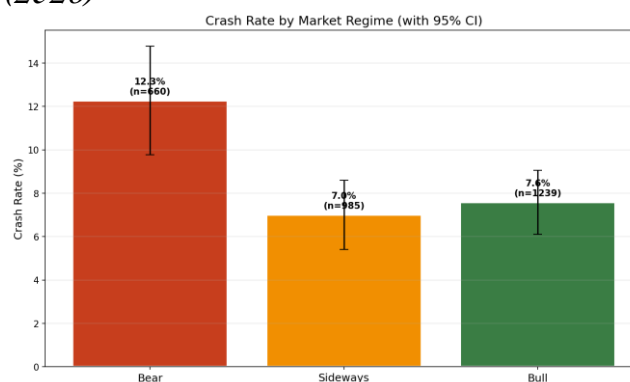
Perbedaan Crash Risk Antar Regime (H1)

Uji Kruskal-Wallis menghasilkan $H = 16,295$ dengan $p = 0,000289$, menolak hipotesis nol bahwa distribusi crash setara antar regime. Laju crash Bear (12,27 persen) jauh melampaui Bull (7,59 persen) dan Sideways (7,01 persen) yang nyaris identik, sebagaimana Tabel 4. Uji lanjut Dunn menegaskan perbedaan hanya melibatkan Bear (Bear vs Bull $p = 0,00143$; Bear vs Sideways $p = 0,00051$; Bull vs Sideways $p = 1,000$), dan rasio odds menempatkan peluang crash Bear 1,705 kali lipat Bull (IK 95 persen [1,247; 2,331]). Pola yang muncul bersifat Bear-lawan-sisanya, bukan tiga jenjang bertingkat, dan tetap signifikan pada ketiga ambang sensitivitas, bahkan menguat pada -20 persen ($p < 0,000001$). H1 diterima: nilai tambah deteksi regime terletak pada kesanggupannya mengisolasi kondisi Bear.

Tabel 4. Crash Rate dan Rasio Odds per Regime (Threshold -15%)

Regime	Crash Events	Total Hari	Crash Rate	Rasio Odds [IK 95%]
Bear	81	660	12,27%	1,705 [1,247; 2,331] vs Bull
Bull	94	1.239	7,59%	0,919 [0,666; 1,268] Sideways vs Bull
Sideways	69	985	7,01%	1,855 [1,325; 2,597] Bear vs Sideways

Sumber: Diolah penulis (2026)



Gambar 2. Crash Rate per Regime Bitcoin
Sumber: Diolah penulis (2026)

Performa Arsitektur Regime-Aware (H2)

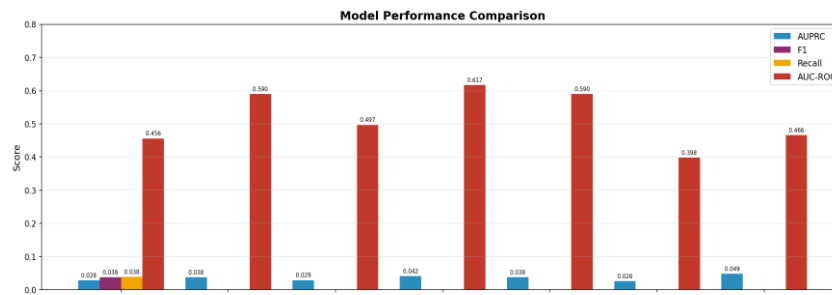
Pengujian pada ambang primer -15 persen terkendala hanya dua crash test sehingga AUPRC maksimum terbatas $\sim 0,006$ dan Wilcoxon empat fold tak pernah mencapai $0,05$ —status inconclusive akibat daya kurang. Analisis pada ambang -10 persen (26 crash test, tujuh fold) yang

berdaya memadai memberi kesimpulan definitif (Tabel 6): rata-rata atas tiga eksekusi independen, strategi terbaik menambah $\Delta\text{AUPRC} = +0,0092 \pm 0,0017$ di atas baseline—jauh di bawah ambang praktis 0,03—dengan Wilcoxon p konsisten tidak signifikan (rentang 0,81–1,00). Lebih menonjol, XGBoost tunggal ($\text{AUPRC} \approx 0,038$) mengungguli atau menyamai seluruh varian stacking, menandakan stacking ensemble sendiri tak memberi keuntungan dan pengondisian regime tak menutup defisitnya. Variasi nilai absolut antar eksekusi bersumber dari komponen LSTM yang non-deterministik, namun verdict penolakan konsisten di seluruh eksekusi. H2 ditolak dengan daya memadai.

Tabel 5. Performa Model pada Test Set

Model	AUPRC	AUC-ROC	Recall	Precision	Brier
Logistic Regression	0,0284	0,456	0,000	0,000	n/a
XGBoost Standalone	0,0380	0,590	0,000	0,000	0,0331
LSTM	0,0415	0,617	0,000	0,000	n/a
Stacking (A) Baseline	0,0380	0,590	0,000	0,000	0,0331
Strategi 1 (Concat)	0,0258	0,398	0,000	0,000	0,0690
Strategi 2 (Submodel)	0,0485	0,466	0,000	0,000	0,0362

Sumber: Diolah penulis (2026)



Gambar 3. Perbandingan Performa Model pada Test Set

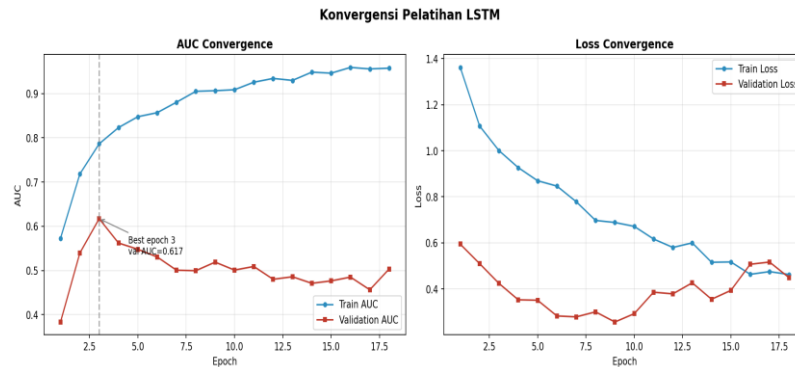
Sumber: Diolah penulis (2026)

Tabel 6. Perbandingan AUPRC pada Threshold Primer (−15%) dan Sensitivitas (−10%)

Arsitektur Model	AUPRC −15% (2 crash, 4 fold)	AUPRC −10% (26 crash, 7 fold)
XGBoost Standalone (referensi)	0,0044	0,0380
Stacking A (Baseline)	0,0059	0,0380
Strategi 1 (Feature Concat)	0,0026	0,0258
Strategi 2 (Regime Submodels)	0,0034	0,0485
Verdict H2	Inconclusive	Rejected ($\Delta +0,0092 \pm 0,0017$; $p \in [0,81-1,00]$)

Sumber: Diolah penulis (2026)

Kegagalan arsitektur regime-aware melampaui baseline paling konsisten dijelaskan tiga faktor: informasi regime sudah implisit pada fitur (terutama volatilitas dan valuasi) sehingga label eksplisit menambah derau; pergeseran distribusi antara training 2018–2023 yang sarat crash ritel dan test 2024–2025 yang didominasi arus institusional pasca-ETF (Babalos et al., 2025; Köse, 2024); serta kompleksitas stacking yang berlebihan untuk tugas bersinyal rendah, tampak pada kurva LSTM di Gambar 4 dengan AUC training $\sim 0,97$ tetapi validasi $\sim 0,56$. Perlu dicatat pula bahwa recall kelas crash bernilai nol pada seluruh model di ambang 0,5, sehingga tak ada arsitektur yang layak sebagai sistem peringatan dini operasional.



Gambar 4. Konvergensi Pelatihan LSTM (validation AUC optimum pada epoch awal $\approx 0,62$, lalu mendatar; nilai bervariasi antar eksekusi akibat non-determinisme LSTM)

Sumber: Diolah penulis (2026)

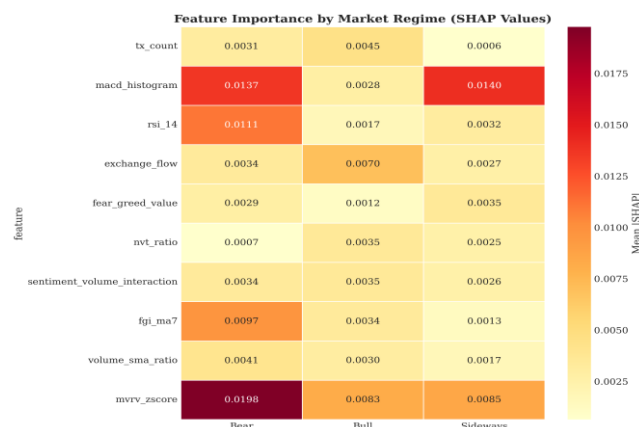
Pergeseran Kepentingan Fitur Antar Regime (H3)

Uji Kruskal-Wallis berkoreksi Bonferroni ($\alpha = 0,01$) atas lima fitur teratas menghasilkan empat fitur signifikan (Tabel 7), melampaui kriteria minimum tiga dari lima; hanya exchange_flow yang tidak signifikan ($p = 0,149$). H3 diterima secara formal, tetapi pemetaan ke teori tekanan pasar klasik tidak terkonfirmasi: di Bear fitur teratas adalah volatilitas teknikal (bukan sentimen), di Bull momentum teknikal (bukan MVRV on-chain), dan di Sideways arus exchange on-chain (bukan indikator teknikal). Temuan menyeluruh memperlihatkan indikator teknikal sebagai prediktor universal lintas regime, sedangkan fitur on-chain bersifat spesifik konteks dengan pergeseran peringkat besar menantang penerapan langsung teori Minsky, keuangan berperilaku, dan analisis teknikal murni pada kripto.

Tabel 7. Bonferroni-Corrected Test SHAP Differences pada Top-5 Features

Feature	Kruskal-Wallis H	p-value	Signifikan
macd_histogram	82,22	1,40e-18	Ya
volatility_30d	342,62	3,99e-75	Ya
exchange_flow	3,81	0,149	Tidak
mrv_v_zscore	102,84	4,66e-23	Ya
fgi_ma7	24,68	4,38e-06	Ya

Sumber: Diolah penulis (2026)



Gambar 5. Heatmap Feature Importance per Regime

Sumber: Diolah penulis (2026)

Ablation Study dan Stabilitas Fitur On-Chain

Lima konfigurasi dibandingkan (Tabel 8). Kontribusi regime conditioning marginal (Config B sedikit di bawah baseline; Config C menurun). Temuan paling substansial adalah Config D (tanpa on-chain) yang menaikkan AUPRC test sekitar 1,17 kali (0,0445 vs 0,0380) dengan AUC-ROC 0,593 bertentangan dengan ekspektasi awal bahwa on-chain menambah daya prediksi. Temuan ini tidak digeneralisasi menjadi klaim bahwa on-chain nirfaedah: pada cross-validation yang mencakup 2019–2023 fitur on-chain berkontribusi positif; yang teridentifikasi adalah masalah kestabilan lintas era akibat perubahan struktural pasar (adopsi institusional, persetujuan ETF, pasca-FTX), sehingga evaluasi stabilitas fitur diperlukan sebelum on-chain direkomendasikan untuk praktik. Ringkasan validasi ketiga hipotesis disajikan pada Tabel 9 dan Gambar 6.

Tabel 8. Hasil Ablation Study

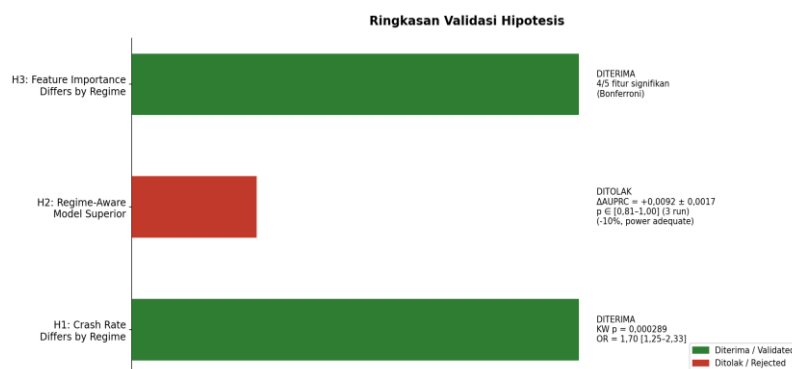
Config	Deskripsi	n Fitur	Test AUPRC	Test AUC-ROC	CV AUPRC
A	Baseline tanpa regime	17	0,0380	0,590	0,0888
B	Regime sebagai fitur	20	0,0336	0,564	0,0888
C	Separate model per regime	17	0,0288	0,497	0,0717
D	Tanpa on-chain features	12	0,0445	0,593	0,0927
E	Tanpa sentiment features	13	0,0317	0,522	0,0850

Sumber: Diolah penulis (2026)

Tabel 9. Ringkasan Validasi Hipotesis

Hipotesis	Uji Statistik	Hasil Empiris	Status
H1	Kruskal-Wallis + Dunn + OR	$p = 0,000289$; OR Bear vs Bull = 1,70 [1,25; 2,33]	Diterima
H2	Wilcoxon (4 & 7 fold)	–15%: inconclusive; –10%: $\Delta +0,0092 \pm 0,0017 < 0,03$; $p \in [0,81-1,00]$ (3 run)	Ditolak
H3	KW + Bonferroni ($\alpha=0,01$)	4/5 fitur signifikan; theory mapping 0/3 confirmed	Diterima

Sumber: Diolah penulis (2026)



Gambar 6. Visualisasi Ringkasan Validasi Hipotesis

Sumber: Diolah penulis (2026)

Pembahasan

Hasil penelitian menunjukkan bahwa dinamika pasar Bitcoin selama periode 2018–2025 dapat diidentifikasi melalui tiga kondisi pasar, yaitu *Bull*, *Sideways*, dan *Bear*, dengan karakteristik return, volatilitas, dan tingkat persistensi yang berbeda. Identifikasi menggunakan *Hidden Markov Model* (HMM) menunjukkan bahwa regime *Bull* mendominasi 42,96% observasi dengan volatilitas terendah, diikuti *Sideways* sebesar 34,15%, sedangkan *Bear* mencakup 22,89% observasi dengan volatilitas tertinggi dan rata-rata return negatif. Persistensi regime yang tinggi, sebagaimana ditunjukkan oleh probabilitas transisi diagonal sebesar 0,90–0,91, mengindikasikan bahwa kondisi

pasar cenderung bertahan selama beberapa periode sebelum mengalami perubahan. Temuan ini sejalan dengan Pennoni et al. (2022), yang menunjukkan bahwa pendekatan HMM dapat mengidentifikasi perbedaan karakteristik return dan volatilitas antar-regime pasar cryptocurrency. Hasil tersebut juga relevan dengan penelitian mengenai pergeseran regime harga Bitcoin menggunakan pendekatan Bayesian dan HMM (2025), yang menekankan pentingnya perubahan kondisi ekonomi dan pasar dalam menjelaskan dinamika harga aset kripto. Secara konseptual, temuan penelitian ini memperkuat pandangan bahwa risiko Bitcoin tidak bersifat konstan sepanjang waktu, melainkan dipengaruhi oleh perubahan keadaan pasar yang dapat diidentifikasi secara probabilistik.

Pengujian hipotesis pertama menunjukkan adanya perbedaan risiko crash antar-regime berdasarkan uji Kruskal–Wallis ($H = 16,295$; $p = 0,000289$). Hasil uji lanjut Dunn memperlihatkan bahwa perbedaan signifikan terutama terjadi antara regime Bear dengan Bull dan Sideways, sedangkan kedua regime terakhir tidak menunjukkan perbedaan yang bermakna. Rasio odds sebesar 1,705 dengan interval kepercayaan 95% sebesar 1,247–2,331 menunjukkan bahwa odds kejadian crash pada regime Bear lebih tinggi dibandingkan Bull. Temuan ini mengindikasikan bahwa kemampuan HMM dalam mengidentifikasi regime berisiko tinggi lebih menonjol dibandingkan kemampuannya membentuk hierarki risiko yang sepenuhnya berbeda pada ketiga regime. Pola tersebut sejalan dengan penelitian mengenai regime switching dan volatilitas Bitcoin yang memperlihatkan adanya perubahan intensitas risiko berdasarkan kondisi pasar. Namun, hubungan antara regime Bear dan peningkatan crash risk perlu dipahami sebagai asosiasi statistik, bukan hubungan sebab-akibat. Hal ini penting karena karakteristik volatilitas yang digunakan dalam pembentukan regime juga dapat berkaitan dengan indikator risiko crash yang dianalisis. Dengan demikian, kontribusi utama pengujian H1 terletak pada identifikasi konsentrasi risiko pada kondisi Bear, yang berpotensi mendukung pengembangan sistem pemantauan risiko pasar secara dinamis.

Berbeda dengan hasil H1, pengujian hipotesis kedua menunjukkan bahwa integrasi informasi regime ke dalam arsitektur prediksi belum memberikan peningkatan kinerja yang signifikan secara statistik maupun substantif. Pada ambang primer -15% , terbatasnya jumlah kejadian crash dalam data pengujian menyebabkan hasil evaluasi belum cukup kuat untuk menghasilkan kesimpulan yang pasti. Sementara itu, pengujian sensitivitas pada ambang -10% menghasilkan peningkatan rata-rata AUPRC sebesar $0,0092 \pm 0,0017$, tetapi tidak mencapai batas peningkatan praktis $0,03$ yang telah ditetapkan dan tidak signifikan berdasarkan uji Wilcoxon. Hasil ini menunjukkan bahwa kemampuan suatu model dalam mengidentifikasi kondisi pasar tidak secara otomatis menghasilkan peningkatan akurasi prediksi kejadian ekstrem. Temuan tersebut memberikan perspektif berbeda terhadap penelitian terdahulu yang mengembangkan pendekatan prediksi berbasis regime, termasuk penelitian *Regime Switching Forecasting for Cryptocurrencies* (2025), yang mengeksplorasi pemanfaatan informasi regime dalam pengambilan keputusan investasi. Perbedaan hasil dapat berkaitan dengan tujuan prediksi, karakteristik data, horizon pengamatan, dan kompleksitas arsitektur yang digunakan. Dalam penelitian ini, informasi regime diduga telah tercermin dalam variabel teknikal, khususnya volatilitas dan momentum, sehingga penambahan label regime secara eksplisit belum menghasilkan informasi prediktif yang cukup besar untuk meningkatkan kinerja model secara konsisten.

Rendahnya kinerja prediksi juga menunjukkan bahwa kompleksitas model tidak selalu berbanding lurus dengan kemampuan generalisasi. Model XGBoost tunggal menghasilkan AUPRC sebesar $0,0380$ dan mampu menyaingi beberapa konfigurasi stacking, sedangkan strategi *regime submodels* menghasilkan AUPRC sebesar $0,0485$ pada ambang -10% . Meskipun terdapat peningkatan numerik pada konfigurasi tertentu, hasil tersebut belum memberikan bukti keunggulan yang konsisten. Perbedaan AUC antara data pelatihan dan validasi pada model LSTM mengindikasikan potensi *overfitting*, yaitu kondisi ketika model mampu mempelajari pola historis tetapi kesulitan mempertahankan kinerjanya pada data baru. Hasil ini relevan dengan Berger (2024), yang membandingkan pendekatan ekonometrika dan pembelajaran mesin dalam peramalan return Bitcoin, serta Park dan Yang (2024), yang menunjukkan bahwa prediksi crash Bitcoin memerlukan pemahaman terhadap pola gelembung harga dan informasi pasar yang

berkaitan dengan kejadian ekstrem. Selain itu, seluruh model menghasilkan recall kelas crash sebesar nol pada ambang klasifikasi 0,5. Kondisi tersebut menunjukkan bahwa model belum dapat diandalkan sebagai sistem peringatan dini operasional berdasarkan konfigurasi keputusan yang diuji. Oleh karena itu, evaluasi model prediksi crash perlu mempertimbangkan ketidakseimbangan kelas, kalibrasi probabilitas, dan pemilihan ambang keputusan sesuai biaya kesalahan prediksi, bukan hanya membandingkan nilai AUC atau AUPRC.

Hasil pengujian hipotesis ketiga menunjukkan bahwa kepentingan fitur prediktif mengalami perubahan signifikan antar-regime. Empat dari lima fitur utama, yaitu *MACD histogram*, *volatility_30d*, *MVRV Z-score*, dan *FGI moving average*, memperlihatkan perbedaan signifikan berdasarkan pengujian Kruskal–Wallis dengan koreksi Bonferroni, sedangkan *exchange flow* tidak menunjukkan perbedaan signifikan. Temuan ini mengindikasikan bahwa kontribusi relatif suatu indikator terhadap keluaran model dipengaruhi oleh kondisi pasar. Indikator volatilitas menjadi dominan pada Bear, momentum teknikal pada Bull, dan informasi arus pertukaran aset pada Sideways. Hasil tersebut menunjukkan bahwa hubungan antara indikator teknikal, sentimen, dan data on-chain lebih kompleks daripada asumsi bahwa satu kelompok indikator akan selalu dominan pada regime tertentu. Penelitian mengenai prediksi Bitcoin menggunakan data on-chain dan seleksi fitur (2025) memperlihatkan bahwa indikator blockchain memiliki potensi untuk menjelaskan pergerakan harga, sedangkan temuan penelitian ini menunjukkan perlunya mempertimbangkan stabilitas kontribusinya pada kondisi pasar yang berbeda. Meskipun demikian, hasil SHAP harus ditafsirkan sebagai kontribusi fitur terhadap prediksi model, bukan sebagai ukuran pengaruh kausal terhadap pembentukan crash. Dengan demikian, penerimaan H3 mendukung adanya heterogenitas kepentingan fitur antar-regime, tetapi belum membuktikan kesesuaian seluruh pola tersebut dengan teori tekanan pasar, keuangan berperilaku, maupun analisis teknikal.

Hasil *ablation study* memperkuat temuan bahwa tambahan kompleksitas dan jumlah fitur tidak selalu meningkatkan kemampuan prediksi. Konfigurasi tanpa fitur on-chain menghasilkan AUPRC sebesar 0,0445, lebih tinggi dibandingkan baseline sebesar 0,0380, sedangkan konfigurasi yang menambahkan informasi regime justru menunjukkan penurunan kinerja pada data pengujian. Namun, penghapusan fitur on-chain tidak dapat langsung ditafsirkan sebagai bukti bahwa informasi blockchain tidak relevan dalam analisis Bitcoin. Perbedaan kontribusi fitur antara periode validasi silang dan data pengujian mengindikasikan kemungkinan perubahan hubungan antara indikator pasar dan kejadian crash dari waktu ke waktu. Perubahan struktur pasar, termasuk perkembangan partisipasi investor institusional dan instrumen investasi berbasis Bitcoin, merupakan penjelasan kontekstual yang perlu diuji lebih lanjut. Temuan ini menunjukkan bahwa kualitas fitur prediktif tidak hanya ditentukan oleh kekuatan hubungan historisnya dengan target, tetapi juga oleh kestabilan hubungan tersebut ketika kondisi pasar berubah. Oleh karena itu, seleksi fitur adaptif, pengujian stabilitas temporal, dan evaluasi model pada beberapa periode pasar menjadi penting dalam pengembangan sistem prediksi risiko cryptocurrency.

Secara teoretis, penelitian ini memberikan kontribusi melalui perbedaan antara kemampuan identifikasi regime, kemampuan prediksi crash, dan perubahan kepentingan fitur antar-kondisi pasar. Hasil penelitian memperlihatkan bahwa HMM dapat memberikan informasi mengenai segmentasi risiko, tetapi informasi tersebut belum tentu menghasilkan peningkatan performa ketika diintegrasikan ke dalam arsitektur pembelajaran mesin yang lebih kompleks. Dari sisi praktis, investor dan pengelola risiko dapat memanfaatkan identifikasi regime Bear sebagai indikator pendukung untuk meningkatkan kewaspadaan terhadap volatilitas dan kejadian pasar ekstrem. Namun, keluaran model tidak sebaiknya digunakan sebagai dasar tunggal dalam pengambilan keputusan investasi karena kemampuan identifikasi crash pada ambang klasifikasi yang diuji masih terbatas. Bagi pengembang sistem analitik keuangan, temuan ini menegaskan pentingnya keseimbangan antara kompleksitas model, interpretabilitas, kestabilan fitur, dan kemampuan generalisasi. Pengembangan model yang lebih sederhana, terkalibrasi, dan diuji secara berkelanjutan dapat menjadi arah yang relevan untuk meningkatkan keandalan prediksi risiko Bitcoin.

Penelitian ini memiliki beberapa keterbatasan yang perlu dipertimbangkan dalam interpretasi hasil. Pertama, kejadian crash merupakan peristiwa yang relatif jarang sehingga jumlah observasi positif, khususnya pada ambang -15% , membatasi ketepatan estimasi kinerja dan kekuatan pengujian statistik. Kedua, penelitian menggunakan data historis 2018–2025 sehingga hasilnya merepresentasikan karakteristik pasar pada periode tersebut dan belum tentu berlaku pada kondisi pasar berikutnya. Ketiga, penggunaan tiga regime HMM bergantung pada spesifikasi model dan proses estimasi, sehingga klasifikasi dapat berubah apabila digunakan jumlah state atau pendekatan identifikasi yang berbeda. Keempat, perbedaan kepentingan fitur berdasarkan SHAP belum sepenuhnya menjelaskan mekanisme ekonomi yang mendasari perubahan kontribusi indikator. Keterbatasan tersebut memberikan arah pengembangan penelitian selanjutnya melalui perluasan periode pengamatan, evaluasi *walk-forward* yang lebih panjang, pengujian perubahan struktural, kalibrasi probabilitas, serta penerapan metode penanganan ketidakseimbangan kelas. Penelitian berikutnya juga dapat membandingkan berbagai spesifikasi HMM, mengembangkan pemilihan fitur adaptif, dan mengevaluasi manfaat model berdasarkan biaya ekonomi kesalahan prediksi. Dengan pendekatan tersebut, penelitian ini menyediakan landasan empiris untuk pengembangan sistem analisis risiko Bitcoin yang lebih adaptif, interpretabel, dan teruji pada perubahan kondisi pasar.

KESIMPULAN DAN SARAN

HMM tiga state berhasil mengenali regime Bull, Bear, dan Sideways dengan laju crash berbeda nyata (Kruskal-Wallis $p = 0,000289$); Bear berpeluang crash 1,7 kali lipat Bull, sementara Bull dan Sideways berisiko serupa, sehingga deteksi regime valid sebagai alat diagnostik (H1 diterima). Namun integrasi regime ke arsitektur ensemble tidak menghasilkan keunggulan prediktif: pada ambang berdaya memadai peningkatan terbaik hanya $+0,0092 \pm 0,0017$ (rata-rata tiga eksekusi) dengan Wilcoxon p tidak signifikan (rentang 0,81–1,00), dan XGBoost tunggal mengungguli seluruh varian stacking (H2 ditolak). Analisis SHAP menunjukkan empat dari lima fitur teratas berbeda nyata antar regime, tetapi pola empiris berlawanan dengan teori klasik karena indikator teknikal mendominasi secara universal sedangkan fitur on-chain bersifat konteks-spesifik (H3 diterima secara formal). Dengan recall nol pada seluruh arsitektur, tak ada model yang layak sebagai sistem peringatan dini operasional.

Secara praktis, keluaran deteksi regime sebaiknya dipakai sebagai indikator diagnostik, bukan sinyal prediksi: saat probabilitas posterior Bear naik, eksposur patut dikurangi karena peluang crash nyata lebih tinggi. Penelitian selanjutnya disarankan menempuh evaluasi rolling window lintas siklus dengan minimal sepuluh crash per fold; estimasi HMM secara maju-berjalan untuk menyingkirkan informasi melihat ke depan; deteksi pergeseran kovariat dengan pelatihan ulang adaptif; perluasan fitur sentimen lewat NLP atas data media sosial mentah; replikasi lintas aset; serta pergeseran perumusan tugas dari klasifikasi biner kejadian langka menuju peramalan volatilitas atau analisis survival waktu-hingga-crash yang lebih serasi dengan karakteristik data kripto.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Pascasarjana dan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta atas dukungan selama penelitian ini, serta kepada rekan sejawat atas diskusi dan masukan yang membangun. Para penulis menyatakan tidak terdapat konflik kepentingan dalam penelitian ini.

DAFTAR PUSTAKA

- Agakishiev, I., Härdle, W. K., Becker, D., & Zuo, X. (2025). Regime switching forecasting for cryptocurrencies. *Digital Finance*, 7, 107–131. <https://doi.org/10.1007/s42521-024-00123-2>
- Babalos, V., Bouri, E., & Gupta, R. (2025). Does the introduction of US spot Bitcoin ETFs affect spot returns and volatility of major cryptocurrencies? *The Quarterly Review of Economics and Finance*, 102, 102006. <https://doi.org/10.1016/j.qref.2025.102006>

- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129–152. <https://doi.org/10.1257/jep.21.2.129>
- Berger, T., & Koubová, J. (2024). Forecasting Bitcoin returns: Econometric time series analysis vs. machine learning. *Journal of Forecasting*, 43(7), 2904–2916. <https://doi.org/10.1002/for.3165>
- Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets. *IMF Staff Papers*, 47(3), 279–310.
- Chandra, R., & He, Y. (2024). Review of deep learning models for crypto price prediction: Implementation and evaluation. *arXiv*. <https://arxiv.org/abs/2405.11431>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. <https://doi.org/10.1145/1143844.1143874>
- Dubey, R., & Enke, D. (2025). Bitcoin price direction prediction using on-chain data and feature selection. *Machine Learning with Applications*, 100674. <https://doi.org/10.1016/j.mlwa.2025.100674>
- Giudici, P., & Polinesi, G. (2021). A hidden Markov model to detect regime changes in cryptoasset markets. *Quality and Reliability Engineering International*, 37(7), 3021–3033. <https://doi.org/10.1002/qre.2960>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). *Springer*. <https://doi.org/10.1007/978-0-387-84858-7>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Ji, S., Kim, J., & Im, H. (2019). A comparative study of Bitcoin price prediction using deep learning. *Mathematics*, 7(10), 898. <https://doi.org/10.3390/math7100898>
- Köse, M. (2024). Evolving strategies in machine learning: A systematic review of concept drift detection. *Information*, 15(12), 786. <https://doi.org/10.3390/info15120786>
- Liu, Z., Shu, M., & Zhu, W. (2024). Contrastive learning framework for Bitcoin crash prediction. *Stats*, 7(2), 402–433. <https://doi.org/10.3390/stats7020025>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv*. <https://arxiv.org/abs/1802.03888>
- Machimbo, E. W., Cheruiyot, W., Mocheche, G. M., & Mutegi, J. (2025). Applications of hidden Markov models in detecting regime changes in Bitcoin markets. *Asian Journal of Probability and Statistics*, 27(7), 127–138. <https://doi.org/10.9734/ajpas/2025/v27i7781>
- Minsky, H. P. (1992). *The financial instability hypothesis (Working Paper No. 74)*. The Jerome Levy Economics Institute.
- Pakštaitė, V., Filatovas, E., Juodis, M., & Paulavičius, R. (2025). Bitcoin price regime shifts: A Bayesian MCMC and hidden Markov model analysis of macroeconomic influence. *Mathematics*, 13(10), Article 1577. <https://doi.org/10.3390/math13101577>
- Park, S., & Yang, J. S. (2024). Machine learning models based on bubble analysis for Bitcoin market crash prediction. *Engineering Applications of Artificial Intelligence*, 135, Article 108857. <https://doi.org/10.1016/j.engappai.2024.108857>
- Pennoni, F., Bartolucci, F., Ferrante, L., & Pandolfi, S. (2022). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. *Economic Notes*, 51(1), e12193. <https://doi.org/10.1111/ecno.12193>

- Penloni, F., Bartolucci, F., Forte, G., & Ametrano, F. (2022). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. *Economic Notes*, 51(1), Article e12193. <https://doi.org/10.1111/ecno.12193>
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Santiment Academy. (2023). *MVRV ratio: On-chain valuation metrics for cryptocurrency*. Santiment.
- Shu, M., & Zhu, W. (2024). Machine learning models based on bubble analysis for Bitcoin market crash prediction. arXiv. <https://arxiv.org/abs/2312.00268>
- Undang-Undang Republik Indonesia Nomor 4 Tahun 2022 tentang Pengembangan dan Penguatan Sektor Keuangan (2022).