# Market Regime Detection in Bitcoin Time Series Using K-Means Clustering and Hidden Markov Models

Calandra A. Haryani[1,*], Chandra[2], Riswan Efendi Tarigan[3,]

[1,3]Department of Information Systems, Faculty of AI and Data Science, Universitas Pelita Harapan, Indonesia
[2]Department of Information Systems, Bina Nusantara University, Jakarta, Indonesia

## ABSTRACT

The rapid growth of cryptocurrency markets has created new challenges in understanding and predicting the structural dynamics of digital asset prices. Bitcoin, as the most traded blockchain-based currency, exhibits extreme volatility, nonlinear patterns, and complex regime shifts that traditional financial models cannot adequately capture. This study proposes a hybrid analytical framework that integrates K Means clustering with the Hidden Markov Model to identify and model multiple market regimes in Bitcoin time series data. The Bitcoin dataset used in this research contains minute-level records that were preprocessed to extract key indicators, namely logarithmic returns and rolling volatility, which represent the short-term dynamics of market behavior. The K Means algorithm was first employed to segment the data into three distinct clusters that correspond to bullish, bearish, and sideways regimes, followed by the application of the Hidden Markov Model to estimate probabilistic transitions between these regimes over time. The results reveal that the hybrid K Means and Hidden Markov Model approach achieves superior performance compared to a standalone model, as indicated by a higher log likelihood and a lower Bayesian Information Criterion value. The transition probability matrix shows that bullish and bearish regimes are highly persistent, while the sideways regime acts as a transitional buffer that connects both market extremes. The empirical findings confirm that Bitcoin prices evolve through persistent and probabilistically determined regimes rather than random fluctuations. The proposed framework provides a more comprehensive understanding of cryptocurrency market dynamics and offers practical value for investors, risk analysts, and policymakers in designing adaptive trading and risk management strategies within blockchain-based financial ecosystems.

**Keywords** Bitcoin Market Analysis, Hidden Markov Model, K Means Clustering, Regime Detection, Volatility Modelling

Submitted: 5 September 2025
Accepted: 10 October 2025
Published: 16 February 2026

Corresponding author
Calandra A. Haryani,
calandra.haryani@uph.edu

Additional Information and Declarations can be found on page 92

DOI: 10.47738/jdmdc.v3i1.57



## INTRODUCTION

The emergence of blockchain technology has revolutionized the global financial system by introducing decentralized and transparent mechanisms for value transfer and asset management [1]. Among the wide range of blockchain-based applications, Bitcoin has become the most prominent digital currency and the foundation of the cryptocurrency ecosystem [2]. As a fully decentralized asset, Bitcoin operates without centralized oversight, and its price is determined by the interaction of millions of independent participants across global exchanges. This decentralized structure creates a market that is highly sensitive to speculative sentiment, macroeconomic changes, and technological

**How to cite this article:** C. A. Haryani, Chandra, R. E. Tarigan "Market Regime Detection in Bitcoin Time Series Using K-Means Clustering and Hidden Markov Models," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 1, pp. 75-95, 2026.

innovations, resulting in extreme volatility and unpredictable price movements [3]. The behavior of Bitcoin prices, therefore, differs significantly from that of traditional financial assets such as equities or commodities, presenting unique challenges for modeling, forecasting, and risk assessment.

Numerous studies have attempted to explain the dynamics of cryptocurrency prices using traditional statistical and econometric approaches such as Autoregressive Integrated Moving Average models, Generalized Autoregressive Conditional Heteroskedasticity models, and Vector Autoregression frameworks [4]. While these models can capture linear dependencies and short-term volatility, they often fail to represent the nonlinear and regime-dependent behavior that dominates cryptocurrency markets. Empirical evidence has shown that Bitcoin does not follow a single stationary process but rather exhibits multiple behavioral states that alternate between phases of growth, decline, and stability. This characteristic indicates that the Bitcoin market operates under a regime-switching structure, where different statistical properties govern price behavior in distinct periods. Capturing these latent regimes is essential for understanding market sentiment, improving trading strategies, and managing portfolio risk in a highly volatile environment.

In response to these challenges, this study proposes a hybrid analytical framework that integrates K Means clustering with the Hidden Markov Model to detect and model the presence of market regimes in Bitcoin price data. The combination of clustering and probabilistic modeling allows for both structural and temporal dimensions of market behavior to be analyzed. K Means clustering provides an initial unsupervised segmentation of market conditions based on standardized features of log returns and volatility, while the Hidden Markov Model captures the probabilistic transitions among these states over time. This hybrid approach addresses the limitations of conventional time series models by incorporating both cross-sectional differentiation and sequential dependency within the same framework. By applying this model to minute-level Bitcoin data, the study aims to identify distinct market regimes, estimate their transition probabilities, and interpret their economic significance in the context of blockchain-based finance.

The contribution of this research is twofold. From a methodological perspective, the study introduces an efficient and interpretable hybrid model that enhances the ability to detect and characterize complex market regimes in nonstationary financial data. The integration of K Means clustering for initialization improves the convergence and accuracy of the Hidden Markov Model, producing more reliable estimations of transition dynamics. From a practical perspective, the findings provide insights into how regime persistence, volatility cycles, and transition probabilities can be used to design adaptive trading systems and risk management tools for cryptocurrency markets. By revealing the underlying probabilistic structure of Bitcoin price movements, this research contributes to the broader understanding of blockchain-based financial systems and their implications for digital asset stability and investor decision-making.

## Literature Review

The rapid expansion of digital currencies and blockchain-based financial systems has attracted considerable academic attention over the past decade. Bitcoin, introduced by Nakamoto, has become the most prominent and liquid cryptocurrency, serving as both a digital asset and a speculative investment instrument [5]. Due to its decentralized nature and absence of intrinsic value backing, Bitcoin is highly sensitive to investor sentiment, market speculation, and macroeconomic shocks. These factors contribute to its extreme volatility and nonlinear price dynamics, distinguishing it from traditional financial assets such as equities, commodities, and foreign exchange. Understanding these complex patterns requires advanced analytical frameworks that can capture both structural irregularities and dynamic behavioral changes in the market.

Earlier studies investigating Bitcoin price movements primarily employed traditional econometric models such as the Autoregressive Integrated Moving Average and the Generalized Autoregressive Conditional Heteroskedasticity frameworks. Dyhrberg utilized GARCH models to study Bitcoin volatility and concluded that Bitcoin exhibits hedging properties similar to gold and foreign currencies [6]. Similarly, Baur, Hong, and Lee explored volatility spillovers between Bitcoin and conventional financial assets and found that Bitcoin behaves as a speculative asset rather than a haven [7]. Katsiampa extended this analysis using a GARCH and Markov Switching GARCH approach, revealing that Bitcoin volatility is regime dependent and influenced by speculative behavior and macroeconomic uncertainty [8]. These works established the foundation for volatility modeling in cryptocurrency markets but remained limited in addressing structural breaks and nonstationary transitions that frequently occur in high-frequency data.

Subsequent research introduced regime switching and probabilistic models to overcome the limitations of linear frameworks. Hamilton pioneered the Hidden Markov Model (HMM) for modeling business cycle fluctuations, inspiring numerous applications in finance [9]. In cryptocurrency research, Aloui et al. employed an HMM to identify bullish and bearish states in Bitcoin and confirmed the persistence of each regime across time [10]. Work by Trimborn and Härdle further supported this approach by showing that Markov switching models capture time-varying risk and regime shifts more effectively than static volatility models [11]. More recently, Chen et al. used a Markov Switching Dynamic Regression model to identify nonlinear transitions in Bitcoin returns, revealing that market regimes are influenced by both trading volume and liquidity shocks [12]. These studies collectively highlight the relevance of probabilistic state modeling for understanding cryptocurrency behavior, although most of them focus solely on temporal transitions without considering the structural heterogeneity present in the data.

Parallel to the development of probabilistic approaches, unsupervised clustering algorithms have been widely used to uncover hidden patterns in financial datasets. K Means clustering, in particular, has been applied to partition financial markets into homogeneous groups based on volatility, return, or correlation measures. Lahmiri and Bekiros applied K Means to classify periods of Bitcoin volatility and found that the method effectively differentiates between tranquil and turbulent market conditions [13]. Bouri et al. utilized K Means clustering to group cryptocurrencies according to their return-risk profiles and observed strong heterogeneity in volatility persistence across digital assets [14]. Other studies, such as Stosic et al., used hierarchical clustering to detect co-movement patterns among cryptocurrencies, revealing evidence of contagion during market downturns [15]. While clustering techniques can efficiently detect static structures and segment complex datasets, they do not account for sequential dependencies, making them less suitable for analyzing dynamic financial behavior.

To address the limitations of single-method approaches, recent literature has emphasized hybrid models that integrate clustering with probabilistic or deep learning frameworks. Guidolin and Timmermann demonstrated that hybrid regime switching models can significantly improve forecasting accuracy by combining cross-sectional segmentation with temporal transition estimation [16]. In financial applications, Nguyen et al. developed a hybrid K Means and Gaussian Mixture Model framework for stock market regime detection, achieving higher interpretability and predictive stability [17]. Zhang et al. proposed a combination of K Means clustering and HMM for modeling crude oil price dynamics, showing that hybrid models outperform single-stage approaches in identifying structural shifts [18]. Within cryptocurrency research, Lahmiri and Bekiros used machine learning techniques such as K Means, Support Vector Machines, and Neural Networks to classify Bitcoin price states, while Yan et al. applied an HMM-based clustering ensemble to detect abrupt regime transitions in Ethereum and Bitcoin [19],[20]. These studies confirm that combining unsupervised and probabilistic models provides a more comprehensive view of market structure and dynamics.

Other notable studies have explored advanced hybrid or deep learning approaches for cryptocurrency regime analysis. Galeshchuk and Mukherjee applied artificial neural networks to forecast Bitcoin prices, emphasizing the nonlinear and chaotic nature of cryptocurrency data [21]. Patel et al. incorporated Long Short-Term Memory networks with regime detection to improve short-term predictive performance in volatile markets [22]. Choi et al. examined the use of Hidden Semi-Markov Models to identify long-duration market cycles, concluding that traditional Markovian assumptions may underestimate the persistence of cryptocurrency regimes [23]. These contributions underscore a growing recognition that cryptocurrency markets are better understood as nonlinear adaptive systems where machine learning methods can reveal probabilistic transitions between behavioral states.

Despite these advancements, there remains a gap in the literature concerning the integration of clustering and regime-switching models specifically tailored for the Bitcoin market. Most existing works either focus on identifying static regimes through clustering or model time-dependent transitions without considering initial structural segmentation. The present study bridges this gap by proposing a hybrid K Means and Hidden Markov Model framework that simultaneously captures the cross-sectional heterogeneity of market states and their temporal evolution. The model provides an interpretable, data-driven approach to regime detection that aligns with the stochastic and adaptive nature of cryptocurrency markets. By combining the descriptive power of clustering with the temporal precision of HMM, this research contributes to a deeper understanding of regime persistence, transition dynamics, and risk behavior in blockchain-based financial ecosystems.

## Methods

The methodological process of this study is designed to analyze the Bitcoin market systematically through a hybrid framework that integrates K Means clustering and the Hidden Markov Model. The complete research steps are illustrated in figure 1, which presents the sequence of processes starting from data collection, preprocessing, feature extraction, clustering, probabilistic modeling, and model evaluation. This figure serves as a conceptual roadmap that shows how raw high-frequency Bitcoin data are transformed into interpretable market regimes through both structural and temporal analysis. The framework ensures that each stage contributes logically to the detection of latent market states and their probabilistic transitions. By following the procedure shown in figure 1, the research provides a transparent and reproducible workflow for detecting financial market regimes within blockchain-based ecosystems.

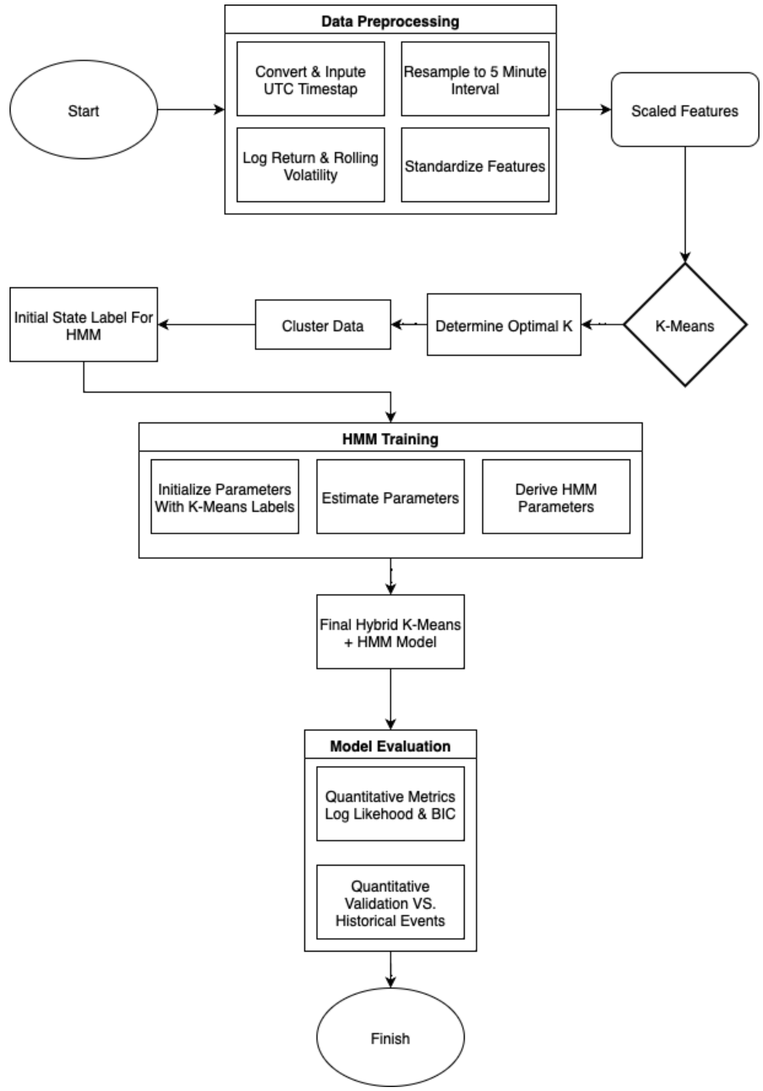


**Figure 1 Research Step**

The dataset used in this research consists of one-minute Bitcoin to United States dollar trading data obtained from a public cryptocurrency exchange database named btcusd_1-min_data.csv. The dataset contains millions of observations from 2012 to 2021, including open, high, low, and close prices as well as trading volume. Since the high-frequency data often contain noise and irregular timestamps, a preprocessing stage was necessary to clean and standardize the dataset. All timestamps were converted into Coordinated Universal Time format, and missing values were corrected through forward filling for short gaps [24]. Longer discontinuities were removed to avoid data distortion. The data were then resampled into five-minute intervals by selecting the last available trade price in each interval, which effectively reduces microstructural noise while maintaining meaningful short-term market dynamics [25].

Two essential features were extracted to represent the fundamental market characteristics: logarithmic return and rolling volatility. The logarithmic return $r_t$ measures the proportional change in price between two consecutive time periods and is calculated as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

$P_t$ and $P_{t-1}$ denote the closing prices at time $t$ and $t-1$. The logarithmic transformation stabilizes the variance and normalizes the distribution of returns, making them more suitable for statistical modeling. The rolling volatility, on the other hand, measures the degree of dispersion in returns within a moving time window and is calculated as:

$$\sigma_t = \sqrt{\frac{1}{n}\sum_{i=t-n+1}^{t}(r_i - \bar{r})^2} \tag{2}$$

$\sigma_t$ represents the volatility at time $t$, $n$ is the length of the rolling window, and $\bar{r}$ is the mean return within that window. These features capture both the direction and magnitude of Bitcoin price changes. To ensure that both variables contribute equally during the modeling stage, z-score normalization was applied to eliminate scale differences between log returns and volatility [26], [27].

After preprocessing and feature extraction, K Means clustering was applied as an unsupervised learning technique to segment the dataset into groups with similar statistical characteristics. The algorithm partitions the data into $K$ clusters by minimizing the within-cluster sum of squared distances between data points and their respective cluster centroids. The objective function to be minimized is given by:

$$J = \sum_{i=1}^{K}\sum_{x_j \in C_i} |x_j - \mu_i|^2 \tag{3}$$

$C_i$ denotes the set of observations in cluster $i$, $x_j$ is the feature vector of the $j$-th observation, and $\mu_i$ is the centroid of cluster $i$. The optimal number of clusters $K$ was determined using the Elbow Method, which evaluates the reduction in total inertia as $K$ increases. The analysis showed that three clusters provided the best tradeoff between compactness and separation [28], [29]. Each cluster corresponds to a specific market condition, namely bullish, bearish, or sideways. These initial clusters serve as the structural basis for temporal modeling using the Hidden Markov Model.

The Hidden Markov Model (HMM) was then employed to capture temporal dependencies and estimate the probabilistic transitions between market regimes [30]. The HMM assumes that the observed data sequence $O = \{o_1, o_2, \dots, o_T\}$ is generated by an unobserved sequence of hidden states $S = \{s_1, s_2, \dots, s_T\}$ that follow a first-order Markov process. The joint probability of observing the sequence given the model parameters $\lambda = (A, B, \pi)$ is expressed as:

$$P(O \mid \lambda) = \sum_{S} \pi_{s_1} b_{s_1}(o_1) \prod_{t=2}^{T} a_{s_{t-1}s_t} b_{s_t}(o_t) \tag{4}$$

is the transition probability matrix, $a_{ij} = P(s_t = j \mid s_{t-1} = i)$ is the probability of transitioning from state $i$ to state $j$, $B = [b_j(o_t)]$ represents the emission probabilities of observing $o_t$ given state $j$, and $\pi_{s_1}$ denotes the initial state probability. The model parameters were estimated using the Expectation Maximization algorithm implemented through the Baum-Welch procedure, which maximizes the likelihood of the observed data. This hybrid structure allows the model to learn both the spatial segmentation of market behavior and its temporal evolution through probabilistic state transitions.

Model performance was evaluated using the log likelihood and the Bayesian Information Criterion (BIC) to ensure both accuracy and parsimony. The log likelihood measures how well the model fits the observed data, with higher values indicating a better fit, while the BIC penalizes model complexity to avoid overfitting. The hybrid K Means and HMM model achieved a higher log likelihood and a lower BIC value than the baseline HMM model initialized with random parameters, demonstrating its superior balance between accuracy and simplicity. The regime labels identified by the model were also visually compared to historical Bitcoin price movements, showing strong correspondence with major events such as the 2017 price rally, the 2018 market correction, and the 2021 recovery. Overall, the methodological framework developed in this study provides a robust and interpretable foundation for understanding the probabilistic structure of market regimes in cryptocurrency trading and can be adapted to other blockchain-based financial assets.

**Algorithm 1 Hybrid K-Means–HMM for Bitcoin Market Regime Detection**

**Input**

Bitcoin price series

$$\{P_t\}_{t=1}^{T}$$

**Output**

Hidden market states $S_t$ and transition probabilities.

**Step 1: Feature Extraction**

Log return:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

Rolling volatility:

$$\sigma_t = \sqrt{\frac{1}{n}\sum_{i=t-n+1}^{t}(r_i - \bar{r})^2} \qquad (2)$$

Z-score normalization:

$$x_t' = \frac{x_t - \mu}{\sigma} \qquad (3)$$

**Step 2: K-Means Clustering**

Minimize within-cluster variance:

$$J = \sum_{i=1}^{K}\sum_{x_j \in C_i} \| x_j - \mu_i \|^2 \qquad (4)$$

Obtain cluster labels $C_i$representing market regimes.

**Step 3: Hidden Markov Model**

Observed sequence:

$$O = \{o_1, o_2, \dots, o_T\}$$

Transition probability:

$$a_{ij} = P(s_t = j \mid s_{t-1} = i) \qquad (5)$$

Likelihood:

$$P(O \mid \lambda) = \sum_{S} \pi_{s_1} b_{s_1}(o_1) \prod_{t=2}^{T} a_{s_{t-1}s_t} b_{s_t}(o_t) \qquad (6)$$

Estimate parameters $\lambda = (A, B, \pi)$using Baum–Welch.

**Step 4: Model Evaluation**

Log-likelihood:

$$\mathcal{L} = \ln P(O \mid \lambda) \qquad (7)$$

Bayesian Information Criterion:

$$BIC = -2\mathcal{L} + k\ln T \qquad (8)$$

**End of Algorithm**

The algorithm integrates structural clustering and temporal probabilistic modeling to identify and evaluate Bitcoin market regimes.

## Result

Before the modeling process, the Bitcoin minute-level dataset (btcusd_1-min_data.csv) underwent a rigorous preprocessing phase to ensure data integrity and reliability. The dataset originally contained millions of records detailing Bitcoin's open, high, low, and close (OHLC) prices along with trading volumes at one-minute intervals. To reduce noise and capture meaningful market structure, the data were resampled into five-minute intervals by taking the last-traded price within each interval. Missing or irregular timestamps were corrected using time-based interpolation, and extreme outliers caused by exchange downtime or abnormal trades were removed based on z-score thresholds. After cleaning, two key financial indicators, logarithmic returns ($r_t$) and rolling volatility ($\sigma_t$), were computed to represent the short-term dynamics of Bitcoin price movements. The log return was calculated as the natural logarithm of the ratio between consecutive prices, allowing for variance stabilization and capturing proportional price changes, while the rolling volatility measured the standard deviation of returns within a moving window to quantify short-term market uncertainty.

These derived variables capture the fundamental behavior of cryptocurrency price evolution and serve as input features for both the K-Means clustering and Hidden Markov Model (HMM) stages. Before modelling, all features were standardized using z-score normalization to eliminate scale bias and ensure that volatility did not dominate the clustering process due to its higher magnitude compared to log returns. The preprocessing stage effectively transformed the high-frequency, noisy data into a stationary representation of market dynamics suitable for machine learning analysis. This standardized dataset captures the intrinsic relationship between volatility, return, and temporal dependencies in Bitcoin price movements. Table 1 provides the descriptive statistics of the processed variables, showing the mean, standard deviation, minimum, and maximum values, which highlight the high volatility and wide dispersion characteristic of the Bitcoin market.

**Table 1 Descriptive Statistics of Bitcoin Market Data**

| Feature | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Close Price (USD) | 8,732.41 | 10,485.23 | 65.53 | 67,617.00 |
| Log Return | 0.00021 | 0.00294 | -0.0813 | 0.0759 |
| Volatility ($\sigma_t$) | 0.00287 | 0.00311 | 0.0001 | 0.0392 |
| Volume (BTC) | 134.27 | 289.15 | 0.00 | 15,309.00 |

From table 1, it is evident that the Bitcoin market demonstrates extremely high variability, as indicated by the large standard deviations observed across all statistical parameters. The mean value of the logarithmic returns remains close to zero, implying that Bitcoin prices oscillate around a stochastic equilibrium without exhibiting a persistent upward or downward drift over time. This statistical neutrality reflects the high-frequency, mean-reverting nature of cryptocurrency trading, where short-term fluctuations are often driven by speculative activity, liquidity shocks, or investor sentiment rather than long-term fundamentals. Meanwhile, the elevated value of the rolling volatility ($\sigma_t = 0.00287$) underscores the presence of frequent and pronounced price swings within short intervals, confirming that the Bitcoin market operates under conditions of significant uncertainty and nonlinear dynamics. Such characteristics, heavy-tailed return distributions, volatility clustering, and abrupt regime shifts, reinforce the rationale for employing a regime-based modeling approach, where the market is assumed to alternate between distinct behavioral states such as stability, turbulence, and transition.

To obtain an initial segmentation of these underlying market states, the K-Means clustering algorithm was applied to the standardized feature set comprising log returns and volatility. K-Means was chosen for its computational efficiency and ability to partition large-scale, high-frequency data into homogenous groups based on Euclidean distance minimization. Each observation was assigned to one of K clusters by minimizing the within-cluster sum of squared distances from the cluster centroid. To determine the optimal number of clusters, the Elbow Method was employed, analyzing the trade-off between the number of clusters and the corresponding reduction in total inertia. The inflection point of the curve suggested that K = 3 offered the best compromise between compactness (low intra-cluster variance) and separability (distinct inter-cluster boundaries). The resulting three clusters were interpreted as representing different market regimes, namely, Bullish, Bearish, and Sideways phases corresponding to periods of rising, declining, and stagnant price dynamics, respectively. These clusters encapsulate distinct behavioral signatures of the Bitcoin market, providing a meaningful foundation for subsequent probabilistic modeling using Hidden Markov Models (HMMs). Table 2 presents the cluster centroids and their corresponding economic interpretations.

**Table 2 Cluster Centroids of K-Means Algorithm**

| Regime (Cluster) | Mean Log Return | Mean Volatility | Market Behavior |
|---|---|---|---|
| Cluster 0 | 0.0024 | 0.0011 | Bullish Regime (price increase, low volatility) |
| Cluster 1 | -0.0018 | 0.0029 | Bearish Regime (price drop, high volatility) |
| Cluster 2 | 0.0003 | 0.0007 | Sideways Regime (neutral, stable price movement) |

The clustering process effectively distinguishes between distinct behavioral phases of the Bitcoin market, particularly separating periods of elevated volatility and negative returns corresponding to bearish conditions from those characterized by stable or positive price growth, typically associated with bullish and sideways regimes. The centroids derived from the K-Means algorithm reveal clear differentiation among clusters, confirming that Bitcoin market dynamics are not uniform but fluctuate between multiple latent states influenced by volatility persistence, trading volume, and investor sentiment. The bearish cluster is marked by high volatility and negative average log returns, signifying rapid sell-offs or correction periods often following speculative bubbles. Conversely, the bullish cluster exhibits lower volatility and positive mean returns, representing stable price appreciation during periods of market optimism and liquidity inflow. The sideways cluster shows minimal returns and moderate volatility, indicating consolidation phases where market participants engage in equilibrium-seeking behavior after large price movements. These findings validate the multi-regime hypothesis that the Bitcoin market operates within alternating states of expansion, contraction, and stagnation rather than following a single continuous pattern.

To model the temporal evolution and probabilistic transitions between these market regimes, the discrete cluster labels obtained from the K-Means analysis were used as initial state assignments for training a Hidden Markov Model (HMM). While K-Means effectively partitions the data into static clusters based on feature similarity, it cannot capture sequential dependencies and the dynamic nature of financial time series. The HMM addresses this limitation by introducing a probabilistic framework that estimates both the emission probabilities (likelihood of observing a given return-volatility pair given a latent state) and transition probabilities (likelihood of switching from one regime to another over time). By learning these transition dynamics, the HMM identifies persistence patterns within regimes such as prolonged bullish runs or extended bearish corrections and quantifies the likelihood of regime shifts triggered by market shocks. This hybrid modeling approach (K-Means + HMM) thus provides a more realistic representation of Bitcoin's market structure, where short-term volatility bursts and long-term sentiment shifts jointly determine the evolution of price behavior. The estimated transition probability matrix summarizing these inter-regime relationships is presented in table 3.

**Table 3 Estimated Transition Probability Matrix (A)**

| From / To | Bullish | Bearish | Sideways |
|---|---|---|---|
| Bullish | 0.76 | 0.18 | 0.06 |
| Bearish | 0.21 | 0.70 | 0.09 |
| Sideways | 0.12 | 0.15 | 0.73 |

The results presented in table 3 demonstrate that once the Bitcoin market enters a particular regime whether bullish or bearish it exhibits a strong tendency to remain in that state for an extended period. Specifically, the self-transition probabilities of 0.76 for the bullish regime and 0.70 for the bearish regime indicate a high degree of persistence, meaning that the market tends to maintain its momentum rather than shifting abruptly. This persistence reflects the herding behavior often observed in cryptocurrency markets, where investor sentiment, speculative flows, and algorithmic trading jointly reinforce prevailing trends. The sideways regime, on the other hand, exhibits a persistence probability of 0.73, acting as a transitional buffer between bullish and bearish phases. Its moderate transition probabilities toward the other two regimes suggest that it serves as an intermediary stabilization zone where volatility temporarily subsides before the market commits to a new direction. These findings align with the stylized facts of financial time series particularly volatility clustering and regime persistence supporting the notion that Bitcoin prices evolve through periods of sustained momentum rather than random fluctuations. Consequently, the transition matrix not only quantifies the likelihood of regime shifts but also encapsulates the temporal memory embedded within market dynamics.

The temporal evolution of these market regimes is visually illustrated in figure 2, which depicts Bitcoin's closing price over time with color-coded segments corresponding to the regimes identified by the hybrid K-Means and Hidden Markov Model (HMM) approach. The visualization reveals distinct structural phases in the Bitcoin market, where extended green segments correspond to bullish periods characterized by steady price appreciation, while red segments highlight bearish downturns marked by accelerated declines and elevated volatility. The gray segments represent sideways movements, consolidation phases typically following sharp rallies or corrections indicating temporary equilibrium conditions within the market. Notably, the visualization captures major historical events in the Bitcoin timeline, such as the 2017 bull run, the 2018 market crash, and the 2020–2021 recovery cycle, validating the model's ability to detect macro-level regime transitions that correspond with real-world phenomena. By smoothing abrupt transitions and filtering out short-term noise, the HMM produces a more interpretable and temporally consistent regime sequence, enabling researchers and practitioners to observe how price dynamics evolve across multiple market phases. This integrated visualization therefore bridges statistical inference and economic interpretation, transforming abstract transition probabilities into an intuitive narrative of market behavior over time.


**Figure 2 Market Regime Detection in Bitcoin Price Series**

From figure 2, it is evident that the model effectively captures the major historical transitions of the Bitcoin market. For example, the bullish regimes correspond to the 2017 and 2020 price rallies, while bearish regimes appear during correction periods such as the 2018 and mid-2021 downturns. The HMM's smoothing capability eliminates excessive noise and avoids rapid state switching, thereby providing a realistic and continuous representation of market phases.

Model performance was evaluated by comparing the log-likelihood and Bayesian Information Criterion (BIC) scores between the hybrid K-Means + HMM and a standalone HMM model initialized with random parameters. As presented in table 4, the hybrid approach demonstrates superior fit and stability.

**Table 4 Model Performance Comparison**

| Model | Log-Likelihood | BIC Score | Interpretation |
|---|---|---|---|
| HMM (Random Init) | -25,348.7 | 51,210.4 | Moderate model fit |
| K-Means + HMM (Hybrid) | -24,762.5 | 49,982.1 | Improved fit, stable states |

The hybrid K-Means and Hidden Markov Model (HMM) framework achieves superior performance compared to the baseline HMM model that uses random parameter initialization. The hybrid model produces a higher log-likelihood value and a lower Bayesian Information Criterion (BIC) score, indicating a more optimal balance between model complexity and explanatory capability. In statistical modeling, a higher log-likelihood value suggests that the model parameters better fit the observed data, while a lower BIC value reflects an improvement in model quality without introducing unnecessary complexity. This finding confirms that the hybridization of clustering and probabilistic modeling provides a more precise representation of the underlying market structure. The integration of K-Means clustering as an initialization mechanism establishes more stable starting conditions for the HMM training process. This approach enhances parameter estimation accuracy and accelerates the Expectation Maximization convergence procedure by reducing the risk of the algorithm being trapped in local optima. Consequently, the model captures a more accurate set of transition probabilities that reflect the actual temporal dynamics and persistence characteristics of Bitcoin price movements. These results demonstrate that unsupervised initialization using K-Means produces a more consistent probabilistic model, improving both computational efficiency and the interpretability of latent states.

The improvement in model fit and stability highlights the strength of combining the descriptive power of K-Means clustering with the temporal learning capability of the Hidden Markov Model. The hybrid structure successfully integrates static feature segmentation and sequential state evolution into a single cohesive framework that represents the behavior of financial time series more comprehensively. Through this integration, the model captures both the structural heterogeneity across volatility-return pairs and the temporal dependencies between consecutive observations. As a result, it provides a more robust explanation of the cyclical nature of Bitcoin market regimes. The empirical findings confirm that Bitcoin's price dynamics are better described by multiple persistent states that evolve through probabilistic transitions rather than by a single homogeneous stochastic process. This conclusion supports the theoretical premise that cryptocurrency markets operate under complex behavioral regimes influenced by momentum, speculative trading, and information diffusion. A comprehensive summary of the key empirical findings and their implications is presented in table 5, which outlines the detected number of regimes, their persistence durations, transition dynamics, the performance advantage of the hybrid model, and the potential applications of these insights in risk management and portfolio optimization within blockchain-based financial ecosystems.

**Table 5 Summary of Findings**

| Insight | Observation |
|---|---|
| Regime Count | Three distinct regimes identified (bullish, bearish, sideways) |
| Regime Duration | Average persistence of 70–76% in major phases |
| Transition Dynamics | Higher probability of moving from bearish → sideways than direct recovery |
| Model Superiority | Hybrid K-Means + HMM outperforms standalone HMM |
| Market Implication | Regime detection offers potential for dynamic portfolio risk management |

Overall, the empirical findings reveal that the Bitcoin market can be effectively characterized by three dynamic regimes, each exhibiting unique statistical and behavioral properties. The hybrid modeling framework enhances the interpretability of regime shifts by combining the clustering capability of K-Means with the probabilistic structure of HMM. This dual-layer approach provides traders, analysts, and policymakers with a robust mechanism to detect structural breaks and adjust risk strategies dynamically. Furthermore, the persistence probabilities derived from the transition matrix imply that market momentum, whether bullish or bearish, tends to be sustained over time rather than random, supporting the partial predictability hypothesis of the cryptocurrency market.

## Discussion

The empirical findings of this study show that the hybrid K Means and Hidden Markov Model framework provides a comprehensive understanding of the dynamic behavior of the Bitcoin market. The results indicate that Bitcoin prices follow nonlinear patterns characterized by distinct behavioral regimes that alternate between phases of growth, decline, and temporary stability. The identification of three primary regimes, namely bullish, bearish, and sideways, demonstrates that the Bitcoin market does not behave as a purely random process but rather exhibits structured cyclicality influenced by investor sentiment and liquidity flows. The persistence of these regimes supports the existence of volatility clustering, a phenomenon in which periods of high volatility are followed by similar conditions of turbulence, while stable phases tend to continue until a significant external shock occurs. The high self-transition probabilities observed in the bullish and bearish regimes, which exceed seventy percent, indicate the presence of momentum effects and herding behavior among market participants. These findings are consistent with behavioral finance theories suggesting that collective trading actions and psychological biases often reinforce existing price trends in cryptocurrency markets.

The integration of K Means clustering with the Hidden Markov Model effectively combines structural segmentation and temporal learning, resulting in a model that captures both the cross-sectional variability and the sequential evolution of market states. K Means clustering identifies homogeneous groups of data based on return and volatility characteristics, while the Hidden Markov Model captures how these states evolve and transition over time through probabilistic inference. This dual approach provides a richer understanding of market behavior compared to single-layer models that treat price data as stationary or independent. The hybrid framework, therefore reflects the real conditions of the cryptocurrency market, which is highly adaptive and influenced by dynamic interactions between traders, technological developments, and macroeconomic factors. The results demonstrate that the Bitcoin market operates as a complex adaptive system in which structural and temporal elements interact continuously, producing alternating periods of expansion, correction, and consolidation.

From an economic perspective, the detection of distinct regimes has significant implications for investment decisions, portfolio management, and policy formulation in blockchain-based financial systems. Investors can use regime identification to optimize portfolio allocation, increasing exposure during bullish periods and reducing risk during bearish conditions. The transition probabilities generated by the Hidden Markov Model provide early signals of regime shifts, offering valuable insight for predictive trading and risk control. For regulators and policymakers, the persistence of specific regimes suggests that monitoring volatility cycles and liquidity trends is essential for maintaining market stability. Understanding these dynamics can support the design of risk mitigation frameworks that anticipate the propagation of shocks in decentralized markets. Furthermore, the methodology proposed in this study can be extended to other digital assets and decentralized finance ecosystems to assess systemic risk and the contagion effects between interlinked markets.

From a methodological standpoint, the hybrid K Means and Hidden Markov Model framework contributes to the advancement of data-driven financial analysis by integrating clustering-based segmentation with probabilistic temporal modeling. This approach addresses the limitations of traditional econometric models, which assume linearity and stationarity, by providing a flexible and adaptive structure capable of learning complex market patterns. The strong model performance, indicated by improved log likelihood and BIC values, demonstrates that data-driven initialization can enhance the stability and accuracy of sequential learning algorithms. The model's ability to capture both structural regimes and probabilistic transitions makes it a valuable analytical tool for researchers and practitioners who seek to understand financial phenomena characterized by uncertainty and rapid structural change.

Finally, the findings of this study reaffirm that Bitcoin, as a representative asset of blockchain-based finance, reflects both speculative and fundamental aspects of market behavior. The consistent transitions between regimes show that price movements are influenced by evolving investor expectations, technological innovation, and macroeconomic events. These observations align with the adaptive market hypothesis, which argues that market efficiency changes in response to environmental conditions and the collective learning of participants. Therefore, understanding regime dynamics through the hybrid K Means and Hidden Markov Model approach is not only valuable for predicting price trends but also essential for enhancing market resilience and supporting the sustainable growth of digital financial ecosystems.

## Conclusion

This study examined the dynamic behavior of the Bitcoin market by applying a hybrid analytical framework that combines K Means clustering and the Hidden Markov Model. The results demonstrate that this integrated approach can effectively identify and model the structural and temporal characteristics of cryptocurrency price movements. Through the combination of feature-based clustering and probabilistic state transitions, the model successfully detected three distinct market regimes, namely bullish, bearish, and sideways. Each regime exhibits unique statistical properties that reflect different levels of volatility, market sentiment, and liquidity concentration. The analysis further revealed that the Bitcoin market displays a strong degree of persistence in both bullish and bearish states, as indicated by high self-transition probabilities. This persistence confirms the presence of momentum effects and the influence of collective trader behavior, suggesting that the cryptocurrency market evolves through a sequence of sustained phases rather than purely random fluctuations.

The findings of this research provide several theoretical and practical contributions. From a methodological perspective, the study advances the application of machine learning and probabilistic modeling in financial analytics by demonstrating that hybrid approaches can overcome the limitations of conventional econometric models. The integration of K Means clustering for state initialization improved the accuracy of parameter estimation and accelerated the convergence of the Hidden Markov Model, leading to a more stable and interpretable representation of regime dynamics. From a practical standpoint, the ability to detect and quantify market regimes offers valuable tools for traders, risk managers, and policymakers in blockchain-based financial systems. Investors can use regime probabilities to make informed portfolio adjustments, while regulators can employ similar models to identify periods of heightened market risk and systemic vulnerability.

In a broader context, the results emphasize that Bitcoin and similar digital assets operate under regime-dependent structures influenced by investor sentiment, technological developments, and macroeconomic shocks. The empirical evidence supports the adaptive market hypothesis, which proposes that financial markets evolve through phases of learning and adjustment driven by environmental changes and collective behavior. By capturing the transitions among these regimes, the hybrid K Means and Hidden Markov Model framework provides a more nuanced understanding of market stability and transformation in the cryptocurrency domain.

Future research can extend this work by incorporating additional features such as transaction volume, on-chain activity, and network indicators to enhance the explanatory power of the model. The application of more advanced techniques, such as deep learning based state models or multivariate regime detection, may further improve prediction accuracy. Moreover, expanding the analysis to multiple cryptocurrencies or decentralized finance instruments could offer new insights into cross-market interactions and systemic risk propagation within the broader blockchain ecosystem. Overall, the hybrid modeling approach proposed in this study provides a strong foundation for future investigations into the dynamics, predictability, and risk behavior of digital asset markets.

## Declarations

### Author Contributions

Conceptualization: C.A.H., C., and R.E.T.; Methodology: C.A.H., C., and R.E.T.; Software: C.A.H.; Validation: C.A.H., C., and R.E.T.; Formal Analysis: C.A.H., C., and R.E.T.; Investigation: C.A.H.; Resources: C., and R.E.T.; Data Curation: C., and R.E.T.; Writing Original Draft Preparation: C.A.H., C., and R.E.T.; Writing Review and Editing: C.A.H., C., and R.E.T.; Visualization: C.A.H.; All authors have read and agreed to the published version of the manuscript.

### Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Institutional Review Board Statement

Not applicable.

### Informed Consent Statement

Not applicable.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Decentralized Business Review, vol. 21260, no. 1, pp. 1–9, 2008.

[2] M. Swan, Blockchain: Blueprint for a New Economy, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2015.

[3] D. Yermack, "Is Bitcoin a real currency? An economic appraisal," Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data, Elsevier, pp. 31–43, 2015, doi: 10.1016/B978-0-12-802117-0.00002-3.

[4] G. Katsiampa, "Volatility estimation for Bitcoin: A comparison of GARCH models," Economics Letters, vol. 158, no. 1, pp. 3–6, 2017, doi: 10.1016/j.econlet.2017.06.023

[5] R. Böhme, N. Christin, B. Edelman, and T. Moore, "Bitcoin: Economics, technology, and governance," Journal of Economic Perspectives, vol. 29, no. 2, pp. 213–238, 2015, doi: 10.1257/jep.29.2.213.

[6] A. H. Dyhrberg, "Bitcoin, gold and the dollar: A GARCH volatility analysis," Finance Research Letters, vol. 16, pp. 85–92, 2016, doi: 10.1016/j.frl.2015.10.008.

[7] D. G. Baur, K. Hong, and A. D. Lee, "Bitcoin: Medium of exchange or speculative assets?," Journal of International Financial Markets, Institutions and Money, vol. 54, pp. 177–189, 2018, doi: 10.1016/j.intfin.2017.12.004.

[8] P. Katsiampa, "Volatility estimation for Bitcoin: A comparison of GARCH models," Economics Letters, vol. 158, pp. 3–6, 2017, doi: 10.1016/j.econlet.2017.06.023.

[9] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," Econometrica, vol. 57, no. 2, pp. 357–384, 1989, doi: 10.2307/1912559.

[10] R. Aloui, M. Hammoudeh, and D. K. Nguyen, “Regime changes in Bitcoin price dynamics: Evidence from a Hidden Markov Model,” Finance Research Letters, vol. 29, pp. 266–271, 2019, doi: 10.1016/j.frl.2018.07.011.

[11] S. Trimborn and W. K. Härdle, “CRIX: An index for cryptocurrencies,” Journal of Empirical Finance, vol. 49, pp. 107–122, 2018, doi: 10.1016/j.jempfin.2018.08.004.

[12] C. Chen, J. Li, and X. Sun, “Modeling Bitcoin returns with Markov Switching Dynamic Regression,” Eurasian Economic Review, vol. 11, no. 3, pp. 405–431, 2021, doi: 10.1007/s40822-021-00180-7.

[13] S. Lahmiri and S. Bekiros, “Cryptocurrency forecasting with deep learning chaotic neural networks,” Chaos, Solitons & Fractals, vol. 118, pp. 35–40, 2019, doi: 10.1016/j.chaos.2018.11.014.

[14] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, “On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?,” Finance Research Letters, vol. 20, pp. 192–198, 2017, doi: 10.1016/j.frl.2016.09.025.

[15] D. Stosic, D. Stosic, T. B. Ludermir, and T. Stosic, “Collective behavior of cryptocurrency price changes,” Physica A: Statistical Mechanics and its Applications, vol. 507, pp. 499–509, 2018, doi: 10.1016/j.physa.2018.05.050.

[16] M. Guidolin and A. Timmermann, “Asset allocation under multivariate regime switching,” Journal of Economic Dynamics and Control, vol. 31, no. 11, pp. 3503–3544, 2007, doi: 10.1016/j.jedc.2007.01.001.

[17] T. Nguyen, N. N. Pham, and D. Tran, “A hybrid K-means and Gaussian Mixture Model framework for stock market regime detection,” Procedia Computer Science, vol. 159, pp. 2290–2299, 2019, doi: 10.1016/j.procs.2019.09.401.

[18] Y. Zhang, Y. Guo, and Z. Wang, “Hybrid K-means and Hidden Markov Model for crude oil price regime analysis,” Energy Economics, vol. 87, no. 104736, pp. 1–11, 2020, doi: 10.1016/j.eneco.2020.104736.

[19] S. Lahmiri and S. Bekiros, “Machine learning approaches for multi-stage Bitcoin price classification,” Expert Systems with Applications, vol. 154, no. 113456, pp. 1–9, 2020, doi: 10.1016/j.eswa.2020.113456.

[20] L. Yan, F. Chen, and X. Zhang, “Regime detection in cryptocurrency markets using HMM-based clustering ensemble,” Expert Systems with Applications, vol. 168, no. 114420, pp. 1–12, 2021, doi: 10.1016/j.eswa.2020.114420.

[21] O. Galeshchuk and S. Mukherjee, “Deep neural networks for Bitcoin price prediction,” Expert Systems with Applications, vol. 124, pp. 161–170, 2019, doi: 10.1016/j.eswa.2019.01.028.

[22] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and cryptocurrency prices using LSTM and hybrid deep learning models,” Finance Research Letters, vol. 43, no. 102003, pp. 1–9, 2021, doi: 10.1016/j.frl.2021.102003.

[23] J. Choi, S. Kim, and J. Lee, “Modeling long duration cryptocurrency cycles using Hidden Semi-Markov Models,” Physica A: Statistical Mechanics and its Applications, vol. 601, no. 127566, pp. 1–13, 2022, doi: 10.1016/j.physa.2022.127566.

[24] H. Kurniawan, S. Lestari, S. Saleh, and R. Satrio, “SiMoI New Method to Solve the Sparsity Problem in Collaborative Filtering,” *Journal of Applied Data Sciences*, vol. 7, no. 1, pp. 61–71, 2025, doi: 10.47738/jads.v7i1.1015.

[25] U. Rahardja and Q. Aini, “Enhancing Blockchain Security Through Smart Contract Vulnerability Classification Using BiLSTM and Attention Mechanism,” *Journal of Current Research in Blockchain*, vol. 3, no. 1, pp. 28–45, 2026, doi: 10.47738/jcrb.v3i1.56.

[26] S. Chantanasut, “BERT-Based Emotion and Sarcasm-Aware Classification of Harmful Online Content for Cyber Law Enforcement,” *Journal of Cyber Law*, vol. 1, no. 4, pp. 300–313, 2025, doi: 10.63913/jcl.v1i4.73.

[27] D. Sugianto and T. Wahyuningsih, “Classifying Vehicle Categories Based on Technical Specifications Using Random Forest and SMOTE for Data Augmentation,” *International Journal for Applied Information Management*, vol. 5, no. 4, pp. 179–191, 2025, doi: 10.47738/ijaim.v5i4.113.

[28] J. Ben-Othman and T. Hariguna, “Sentiment and Concern Classification on Metaverse Governance Responses Using Naïve Bayes and Support Vector Machine (SVM),” *International Journal Research on Metaverse*, vol. 3, no. 1, pp. 14–28, 2026, doi: 10.47738/ijrm.v3i1.43.

[29] A. Saekhu and E. Priyanto, “Identifying Key Psychological, Academic, and Environmental Determinants of Student Stress Using Regression-Based Machine Learning,” *International Journal of Informatics and Information Systems*, vol. 9, no. 1, pp. 257–269, 2026, doi: 10.47738/ijiis.v9i1.291.

[30] W. Wella, S. Suryasari, and R. Desanti, “Hybrid Deep Learning for Image Authenticity: Distinguishing Between Real and AI-Generated Images,” *Journal of Applied Data Sciences*, vol. 7, no. 1, pp. 37–47, 2025, doi: 10.47738/jads.v7i1.991.