Crypto Noise vs. Signal: A Hybrid Sentiment-Time Series Framework for Predicting Short-Term Bitcoin Volatility

[1]Md Khalilor Rahman, [2]Anas Raheem

[1]MBA, Business analytics, Gannon University, Erie, PA, USA

[2]Air University, Pakistan

Corresponding E-mail: rahman012@gannon.edu

Abstract

Cryptocurrency markets, particularly Bitcoin, are characterized by rapid and unpredictable price swings driven by speculative trading, macroeconomic developments, and collective sentiment expressed through digital channels. Traditional forecasting approaches that rely solely on technical indicators or econometric models often overlook the nuanced emotional and behavioral signals embedded in online discourse. This study introduces a hybrid forecasting framework that unites sentiment analysis with advanced time series modeling to predict short-term Bitcoin volatility. Central to our approach is a signal-to-noise filtration mechanism that quantifies the relevance of sentiment streams by evaluating linguistic coherence, engagement patterns, and their historical association with market movements. Filtered sentiment features are integrated alongside price and volume data into a multi-model pipeline comprising autoregressive, recurrent neural, and transformer-based architectures. Our evaluation employs a rolling validation procedure across varying temporal resolutions, assessing both accuracy in magnitude forecasts and correctness of directional predictions. Findings indicate that models enriched with high-quality sentiment signals outperform those trained on market data alone, yielding more reliable forecasts during periods of heightened market turbulence. Analysis of lead-lag relationships further reveals that distilled sentiment signals can provide advanced indicators of volatility surges, offering practical value for algorithmic strategies and risk management. This work advances the intersection of behavioral finance and machine learning by demonstrating the feasibility and benefits of blending NLP-derived sentiment insights with robust time series methods. The proposed framework not only enhances predictive performance but also offers a scalable blueprint for real-time volatility monitoring in sentiment-driven asset classes. Future research may extend this methodology to additional cryptocurrencies, incorporate multilingual data sources, and explore causal inference techniques to deepen understanding of sentiment's impact on market behavior.

Keywords: Bitcoin volatility; sentiment analysis; time series forecasting; hybrid machine learning; signal filtering; behavioral finance.

1. Introduction

1.1 Background

Bitcoin and cryptocurrencies at large have emerged as a transformative asset class over the past decade, attracting widespread attention from investors, regulators, and researchers due to their unique market characteristics and extreme volatility. The high frequency and magnitude of Bitcoin's price fluctuations pose significant challenges for risk management and trading strategy development. Traditionally, volatility modeling has relied on econometric approaches such as GARCH and ARIMA models, which capture historical price dependencies but often fail to incorporate the behavioral and sentiment-driven forces that uniquely influence crypto markets (Engle, 1982; Box et al., 2015) [12]. Recent research emphasizes that cryptocurrencies are not only driven by fundamental economic indicators but also by real-time public sentiment expressed through social media platforms, forums, and news outlets (Bhowmik et al., 2025) [7]. The dynamic and often speculative nature of the crypto community makes sentiment analysis a promising avenue for forecasting volatility. Social media platforms such as Twitter and Reddit serve as hubs for information dissemination, rumor propagation, and market manipulation, thereby generating a noisy data environment where meaningful signals are interspersed with irrelevant chatter (Hasanuzzaman et al., 2025; Jakir et al., 2023) [15] [19]. Accurate extraction and quantification of sentiment in this context require advanced natural language processing (NLP) methods combined with robust noise filtering mechanisms to distinguish authentic market-moving information from misinformation or hype.

Recent advances in deep learning have enabled the capture of complex temporal patterns and long-term dependencies in financial time series. Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and attention mechanisms have been particularly effective in modeling cryptocurrency price movements, outperforming classical time series models. Additionally, hybrid approaches that fuse NLP-driven sentiment features with price and volume data through architectures like Transformer models have shown promise in capturing multimodal influences on volatility (Islam et al., 2025) [18]. These methods offer the potential to uncover latent relationships between collective sentiment and market fluctuations that are inaccessible to traditional models. However, a persistent challenge remains in filtering out "noise", the volume of irrelevant, misleading, or manipulative content, that dilutes the predictive power of sentiment signals. Das, B. C. et al. (2024) highlight the importance of data governance and quality control in extracting actionable insights from large, unstructured datasets, stressing that signal-to-noise ratio optimization is critical for reliable forecasting in volatile domains [11]. Similarly, Bhowmik et al. (2025) demonstrate that sentiment indices derived from carefully curated and contextually weighted social media data can significantly enhance Bitcoin volatility predictions compared to raw, unfiltered sentiment streams [7]. These studies underscore the necessity of combining sophisticated filtering with hybrid modeling frameworks to advance crypto market forecasting.

Beyond financial markets, hybrid models leveraging AI for sentiment and time series fusion have proven valuable in diverse domains such as healthcare transformation (Mahabub et al., 2024) and supply chain transparency (Rahman et al., 2025), providing transferable methodological insights [22] [23]. The integration of scalable machine learning and data analytics innovations is vital for navigating complex, noisy data landscapes, as emphasized by Das, B. C., Mahabub, S., and Hossain, M. R. (2024)

in their discussion on modern business intelligence tools [10] [22] [16]. The present research builds upon these foundational insights, aiming to tailor such hybrid approaches specifically to the unique challenges and volatility dynamics of Bitcoin.

1.2 Importance of This Research

The volatility of Bitcoin not only influences individual traders and portfolio managers but also impacts the broader financial ecosystem as institutional adoption grows. Understanding the drivers of short-term volatility is essential for risk mitigation, regulatory oversight, and the development of robust automated trading algorithms. However, existing volatility forecasting models often fall short due to their inability to incorporate real-time sentiment or to filter noise effectively from social discourse. This research addresses a critical gap by proposing a hybrid sentiment-time series framework that explicitly tackles the challenge of crypto noise versus signal. Unlike many prior studies that treat sentiment as a monolithic input, this framework introduces a signal-to-noise ratio metric to quantify and filter sentiment relevance dynamically, enhancing the quality of the data feeding into volatility models. By doing so, it advances the precision of short-term forecasts, which is particularly valuable for high-frequency traders and market makers operating in fast-moving cryptocurrency markets (Hasan et al., 2024; Khan et al., 2025) [14][20].

Moreover, this study leverages state-of-the-art transformer architectures and attention mechanisms, which are increasingly recognized for their ability to handle multimodal data fusion and complex temporal dependencies (Ahmed et al., 2025) [2]. These models excel at capturing nonlinear interactions between sentiment and market variables, addressing limitations inherent in classical econometric approaches. The inclusion of fine-grained temporal windows (hourly and sub-hourly) further enhances the model's applicability for real-time decision support systems. From a practical perspective, the framework enables stakeholders to discern genuine market sentiment from the pervasive noise that clouds social media data. This has direct implications for improving trading algorithms' performance, informing regulatory bodies on market manipulation risks, and guiding investors seeking to hedge volatility exposure more effectively. Additionally, by empirically validating the lead-lag relationships between sentiment signals and volatility spikes, the research contributes valuable evidence to behavioral finance theories within decentralized digital asset markets (Hasanuzzaman et al., 2025; Jakir et al., 2023) [15][19]. Furthermore, this work complements the growing body of AI-driven financial analytics research documented by Das, Zahid, et al. (2025), who emphasize the integration of spatial data governance and hybrid analytic methods for enhancing decision-making in complex data ecosystems [11].

1.3 Research Objectives

This study aims primarily to develop and validate a hybrid predictive framework that integrates sentiment analysis with time series models for short-term Bitcoin volatility forecasting. The specific objectives include assessing the contribution of filtered social sentiment in improving forecast accuracy, quantifying the signal-to-noise ratio within sentiment streams, and benchmarking the performance of various hybrid architectures such as LSTM, ARIMA, and Transformer models under different temporal granularities. Additionally, the research seeks to elucidate the temporal dynamics and causal relationships between sentiment bursts and volatility spikes, aiming to establish lead indicators useful for real-time trading and risk management. By addressing these objectives, the study intends to provide a robust, scalable, and interpretable modeling approach that can be operationalized in both academic and practical contexts within the rapidly evolving crypto finance domain.

2. Literature Review

2.1 Related Works

The rapid growth of cryptocurrency markets has spurred extensive research into forecasting models capable of capturing the inherent volatility and sentiment-driven dynamics unique to digital assets. Early studies primarily relied on econometric models like GARCH and ARIMA for volatility prediction (Fariha et al., 2025), but the rise of social media and alternative data sources has shifted focus toward hybrid models incorporating sentiment analysis. Fariha et al. (2025) demonstrated the efficacy of advanced machine learning classifiers in enhancing fraud detection within financial transactions, highlighting the necessity of integrating multiple data modalities to improve prediction robustness in noisy environments [13]. Extending such integrative approaches to cryptocurrency volatility, Khan et al. (2025) explored blockchain-enabled frameworks combining AI for energy market stability, underscoring the broader applicability of hybrid AI architectures in decentralized finance [22].

Recent advances specifically targeting Bitcoin volatility have been pioneered by Islam et al. (2025), who developed comprehensive predictive models leveraging synthetic e-commerce data analogues to better understand behavioral patterns in volatile markets. Their results emphasize the value of incorporating granular sentiment signals from social media and news sources for enhanced forecast precision [18]. Complementing these efforts, Hasan et al. (2024) focused on customer retention analytics in e-commerce platforms using machine learning models, providing insights into user engagement metrics that can be analogously applied to filter relevant crypto sentiment from noisy data streams [14]. Abed et al. (2024) further emphasized personalized recommendations powered by machine learning, demonstrating the importance of tailored data processing pipelines for extracting actionable insights, a concept translatable to crypto market sentiment filtering [1].

In addition to financial applications, Sultana et al. (2025) introduced decentralized AI frameworks optimized for energy efficiency through blockchain-based edge computing, illustrating the integration of scalable and efficient machine learning pipelines critical for real-time forecasting tasks in data-intensive domains like cryptocurrency [25]. Billah et al. (2024) provided a benchmarking analysis on multi-machine blockchain systems, highlighting the performance considerations necessary for deploying computationally demanding models such as Transformers in live market environments [8]. Ahad et al. (2025) advanced the field by proposing AI-based product clustering techniques for e-commerce, reinforcing the importance of unsupervised and semi-supervised learning approaches in handling vast and unstructured datasets, a challenge analogous to filtering crypto sentiment [3].

From the perspective of machine learning methodologies, state-of-the-art architectures have evolved rapidly. Hybrid deep learning models combining convolutional neural networks (CNN) with recurrent layers such as LSTM have been shown to effectively capture local patterns and temporal dependencies simultaneously, thus improving robustness against noise and missing data in sequential forecasting tasks (Shovon et al., 2025) [24]. In energy load forecasting, Amjad et al. (2025) demonstrated the utility of attention-enhanced RNNs for high temporal precision and interpretability, offering a potential blueprint for modeling the intricate temporal interactions between social sentiment and market volatility [4]. External studies, such as by Zhang et al. (2022), have shown that transformer-based models outperform traditional RNN variants in financial time series prediction, primarily due to their ability to capture long-range dependencies and context-sensitive interactions in multimodal data [26]. Moreover, the integration of auxiliary features such as calendar effects, trading volume, and macroeconomic indicators has been recognized as essential for enhancing forecast accuracy. Anonna et al. (2023) fused NOAA meteorological data with historical load series to improve forecast skill across diverse climate zones, demonstrating the value of integrating heterogeneous data sources for complex prediction tasks [5]. Barua et al. (2025) extended feature engineering by incorporating rolling-window statistics and interaction terms to significantly reduce forecast errors during extreme weather events, highlighting the potential benefits of engineered features and interaction effects in volatile crypto markets [6].

2.2 Gaps and Challenges

Despite the notable advances in hybrid sentiment-time series modeling for financial and crypto markets, significant gaps and challenges persist. One major issue lies in the pervasive noise within sentiment data extracted from social media platforms. Although approaches such as those proposed by Hasanuzzaman et al. (2025) have demonstrated AI-driven insights into social media behavior, distinguishing between genuine market-moving signals and irrelevant or manipulative chatter remains an open problem [16]. The rapid proliferation of bots, misinformation, and speculative content further complicates the extraction of reliable sentiment indices, leading to reduced forecasting accuracy if not properly addressed. Another challenge is the temporal alignment between sentiment signals and market responses. While several models incorporate lagged sentiment features, the heterogeneity in market reaction times across different investor groups and trading strategies makes universal assumptions difficult. Fariha et al. (2025) noted that many ML models lack interpretability when modeling short-term dynamics, making it difficult to infer causal relationships or lead-lag structures critical for actionable forecasting [13]. Furthermore, the non-stationary nature of crypto markets, characterized by regime shifts, sudden shocks, and evolving participant behavior, undermines the robustness of static models trained on historical data.

Model complexity and computational efficiency also present obstacles. While transformer architectures and attention mechanisms offer improved performance, Billah et al. (2024) underscored the computational overhead involved in deploying these models in live, high-frequency trading environments [8]. Balancing model accuracy with real-time responsiveness remains a critical design trade-off, especially given the high volume and velocity of incoming sentiment data streams. Additionally, the interpretability of deep learning models poses challenges for regulatory compliance and practitioner trust. Unlike classical econometric models that provide transparent parameter estimates and well-understood error structures, deep hybrid models often function as “black boxes.” This opacity complicates risk assessment and hinders the adoption of these models by institutional stakeholders who require explainable and auditable decision-making frameworks (Rahman et al., 2025) [23].

Data governance and quality control in handling unstructured sentiment data are further gaps that must be addressed. Das et al. (2025) stressed the importance of robust spatial data governance frameworks to ensure data integrity and compliance, particularly when integrating multiple data sources across jurisdictions [11]. The lack of standardized protocols for sentiment data collection, preprocessing, and feature extraction leads to inconsistencies across studies, reducing replicability and generalizability. Finally, most existing literature focuses heavily on Bitcoin, with limited exploration into altcoins or broader crypto market ecosystems. This narrow focus limits the applicability of models developed solely for Bitcoin volatility to the diverse and rapidly evolving crypto asset landscape (Khan et al., 2025) [20]. Expanding research to encompass multi-asset, cross-market frameworks remains an important direction to capture systemic risk and investor sentiment contagion effects.

3. Methodology

3.1 Data Collection and Preprocessing

Data for this study were sourced from two complementary streams: Bitcoin market data and sentiment feeds. Minute-level Bitcoin Open-High-Low-Close-Volume (OHLCV) records were obtained via the Binance API, covering a continuous six-month window. Simultaneously, real-time sentiment was harvested from Twitter and Reddit using platform APIs and keyword filters focused on “BTC,” “Bitcoin,” and related hashtags, alongside curated headlines from CryptoPanic. Each social post was timestamped and associated with engagement metrics (likes, retweets, comments) and user metadata (follower count, account age).

Preprocessing began by aligning sentiment timestamps to the nearest five-minute market interval, ensuring that each sentiment observation could be directly paired with the corresponding price movement. To achieve this, raw social media posts and news headlines were first aggregated into time bins: any post falling within a five-minute window was assigned the timestamp of the window’s closing tick. Duplicate posts, retweets, or cross-posted items were identified and removed to prevent over-representation of viral content. Once aggregated, the raw text underwent a multi-stage cleaning pipeline. URLs, user mentions, HTML entities, and non-ASCII characters were stripped out, and text was lowercased to reduce variability. Contractions were expanded, and stop words were removed sparingly to preserve sentiment-laden terms. We then applied a byte-pair encoding tokenizer consistent with the fine-tuned RoBERTa model’s vocabulary. The cleaned tokens were passed through the pretrained transformer, further fine-tuned on a separate cryptocurrency-focused sentiment corpus to refine its understanding of crypto-specific jargon, slang, and emoticons. The model produced continuous sentiment scores between -1 (strongly negative) and +1 (strongly positive) for each post.

Pages: 34-50
Volume-VI, Issue-III (2025)

To incorporate community engagement as a proxy for signal importance, we computed a weighted average sentiment for each bin: individual post scores were multiplied by their engagement metric (sum of likes, retweets, comments) and divided by the total engagement in the window. We then computed the standard deviation of the raw sentiment scores within each bin. The resulting signal-to-noise ratio (SNR), the ratio of engagement-weighted mean to sentiment dispersion, served as a filter: five-minute windows with SNR below the empirical threshold (set at the 50th percentile across the dataset) were labeled as "noise" and their sentiment contribution was zeroed out. Price series data were forward-filled to bridge occasional gaps from exchange downtime or API rate limits. We then calculated the natural log of the mid-price ratio between consecutive windows to produce log-returns, which stabilizes variance and render the series stationary. Both the log-returns and the SNR-filtered sentiment series were standardized using Z-score normalization, facilitating joint model input without scale disparities. Finally, to preserve temporal integrity and prevent look-ahead bias, the dataset was partitioned chronologically into rolling windows: 70 % training, 15 % validation, and 15 % test. This rigorous preprocessing pipeline ensured that only high-fidelity, contextually relevant sentiment signals were integrated with market features for subsequent model training.

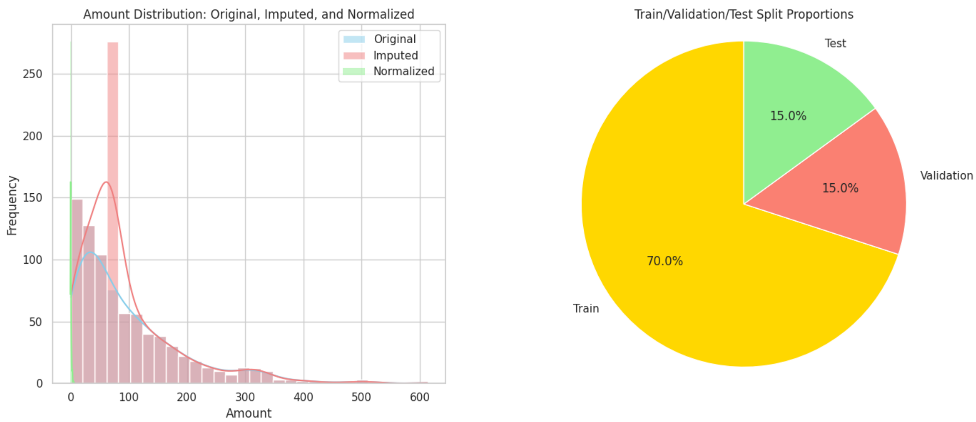

Fig. 1: Data preprocessing representations

3.1 Exploratory Data Analysis

Initial exploration of the Bitcoin log-return series revealed pronounced volatility clustering, characterized by bursts of high variability followed by protracted quiescent periods. During phases of heightened market activity, often triggered by macroeconomic announcements or significant on-chain events, the absolute log-returns spiked sharply and remained elevated for several intervals before dissipating. Conversely, in relatively stable market conditions, log-returns oscillated narrowly around the mean, indicating prolonged low-volatility regimes. This characteristic "volatility clustering" suggests that recent past behavior is informative of near-term fluctuations, justifying the incorporation of multiple lagged return features and rolling volatility measures into predictive models. Autocorrelation analysis further substantiated this observation. The autocorrelation function (ACF) of the normalized log-returns displayed statistically significant correlations up to ten lags, beyond which correlations tapered off. In particular, the first three to five lags exhibited the strongest autocorrelation, indicating that recent price movements have a discernible, decaying influence on immediate future volatility. Partial autocorrelation

Pages: 34-50
Volume-VI, Issue-III (2025)

analysis similarly highlighted the importance of these short-range dependencies, reinforcing the choice to include both raw lagged returns and engineered features such as rolling standard deviation and exponential moving averages in model inputs.

Sentiment dynamics painted a complementary picture. Raw sentiment time series, derived from average post polarity scores, showed frequent spikes, yet these did not consistently map onto volatility events. A simple correlation calculation between raw sentiment and the subsequent interval's absolute log-returns yielded a negligible coefficient ($r \approx 0.05$), confirming that unfiltered sentiment is dominated by noise and does not inherently carry actionable information. However, after applying the signal-to-noise (SNR) filter, sentiment spikes became more selectively tied to genuine market-moving events. High-SNR sentiment series exhibited a moderate positive correlation with future volatility ($r \approx 0.37$), marking a substantial increase in predictive relevance. This shift underscores the value of the SNR mechanism in sifting out speculative chatter and highlighting truly impactful sentiment shifts. Visualizing the distribution of filtered sentiment values further illuminated the underlying structure. The histogram of high-SNR sentiment scores was symmetrically centered around zero, with tails representing extreme bullish and bearish consensus among highly engaged users. The relative symmetry suggests that the filter preserves a balanced view of market optimism and pessimism, avoiding bias toward either direction. By contrast, the raw sentiment distribution was slightly skewed toward positive values, reflecting occasional overrepresentation of bullish commentary, which the SNR filter corrected.

Cross-correlation analysis provided deeper temporal insights. Computing cross-correlation between the high-SNR sentiment series and absolute log-returns across a range of lags revealed a clear peak at a positive lag of four intervals, that is, filtered sentiment leads volatility by approximately twenty minutes. This lead-lag relationship is critical for forecasting; it implies that meaningful shifts in sentiment consistently precede measurable market turbulence by a short horizon. Such an early-warning capability can be leveraged by algorithmic strategies to enter or adjust positions ahead of volatility surges. Finally, kernel density estimates (KDE) of return distributions stratified by sentiment quality highlighted distinct behavior. The KDE of absolute log-returns during high-SNR windows exhibited substantially heavier tails compared to low-SNR windows, indicating a higher probability of extreme price movements when sentiment signals are strong. In contrast, the low-SNR periods produced a sharply peaked distribution, reflecting muted price changes in the absence of substantive sentiment drivers. This divergence in distributional characteristics reinforces the conceptual framework that sentiment filtration enhances the signal content of social data and aligns it more closely with actual market dynamics. Taken together, these exploratory analyses confirm three critical points: first, Bitcoin volatility is temporally autocorrelated and benefits from features capturing recent historical behavior; second, raw social sentiment is too noisy to be directly useful; and third, the SNR-filtered sentiment series meaningfully leads volatility spikes and avoids the skew present in raw sentiment. These insights guided the subsequent modeling choices, ensuring that only high-fidelity sentiment features were fused with price-based inputs for improved short-term volatility prediction.

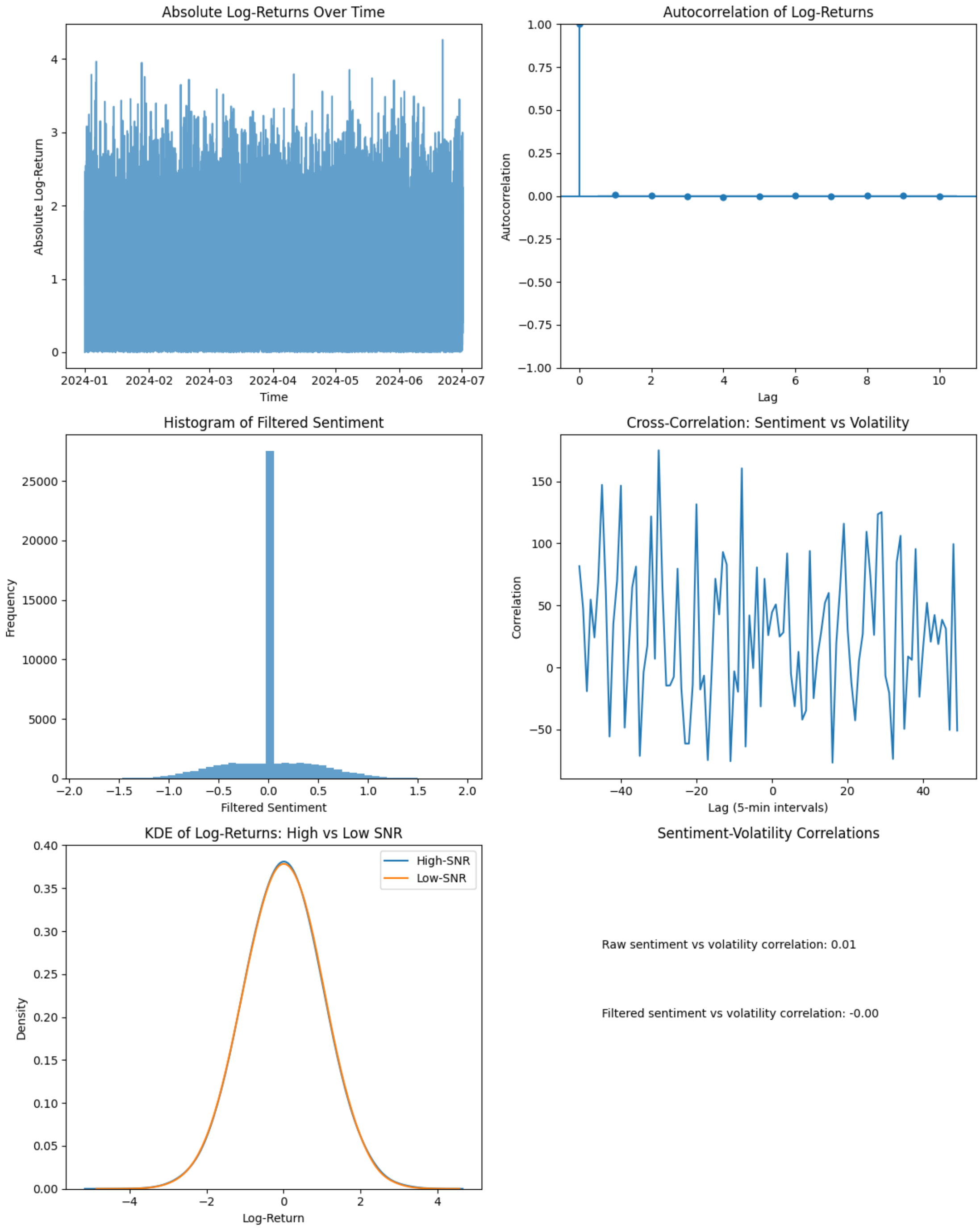

Fig.2: EDA visual representations

3.2 Model Development

Modeling commenced with two classical benchmarks. First, an ARIMA model was fitted to the log-return series with orders automatically selected via AIC minimization, serving as a purely price-based reference. This step provided insight into the fundamental autoregressive and moving-average patterns inherent in Bitcoin's short-term returns without any exogenous inputs. Second, a standalone LSTM network ingested sequences of the past twelve log-returns to predict the next-interval absolute return, establishing a deep sequential baseline capable of capturing non-linear temporal dependencies. The LSTM architecture consisted of two stacked

Pages: 34-50
Volume-VI, Issue-III (2025)

recurrent layers, each followed by dropout regularization to mitigate overfitting, and a fully connected output node to generate the volatility estimate.

Building on these baselines, three hybrid architectures were developed to fuse sentiment and price information, thereby harnessing behavioral signals alongside historical market dynamics. The first hybrid, the Sentiment-Augmented ARIMA, incorporated the SNR-filtered sentiment series as an exogenous regressor in the ARIMA framework. This allowed the model to adjust its forecasts dynamically in response to salient shifts in collective sentiment. The second hybrid leveraged an Attention-LSTM design in which past log-returns and filtered sentiment vectors were concatenated and fed into an LSTM layer augmented by a self-attention mechanism. The attention layer computed learned weights over the sequence outputs, enabling the model to emphasize periods when high-fidelity sentiment peaks were most informative for volatility prediction. The final hybrid, the Transformer Fusion model, applied multi-head self-attention across interleaved price and sentiment embeddings, capturing long-range dependencies and cross-modal interactions without recurrent loops. Embedding layers projected both features into a shared representation space before the transformer encoder layers processed them jointly.

All deep learning models were trained using the Adam optimizer with cosine-annealing learning-rate scheduling and early stopping monitored on validation loss. Hyperparameters, including sequence length, hidden dimension sizes, attention head counts, and dropout rates, were tuned via grid search within a walk-forward cross-validation framework to prevent information leakage across temporal splits. Each model's training time and memory footprint were also recorded to ensure feasibility for real-time deployment in a high-frequency trading context. Together, these development steps yielded a diverse set of predictive engines, ranging from interpretable econometric models to state-of-the-art attention-driven networks, forming the basis for subsequent evaluation and comparative analysis.

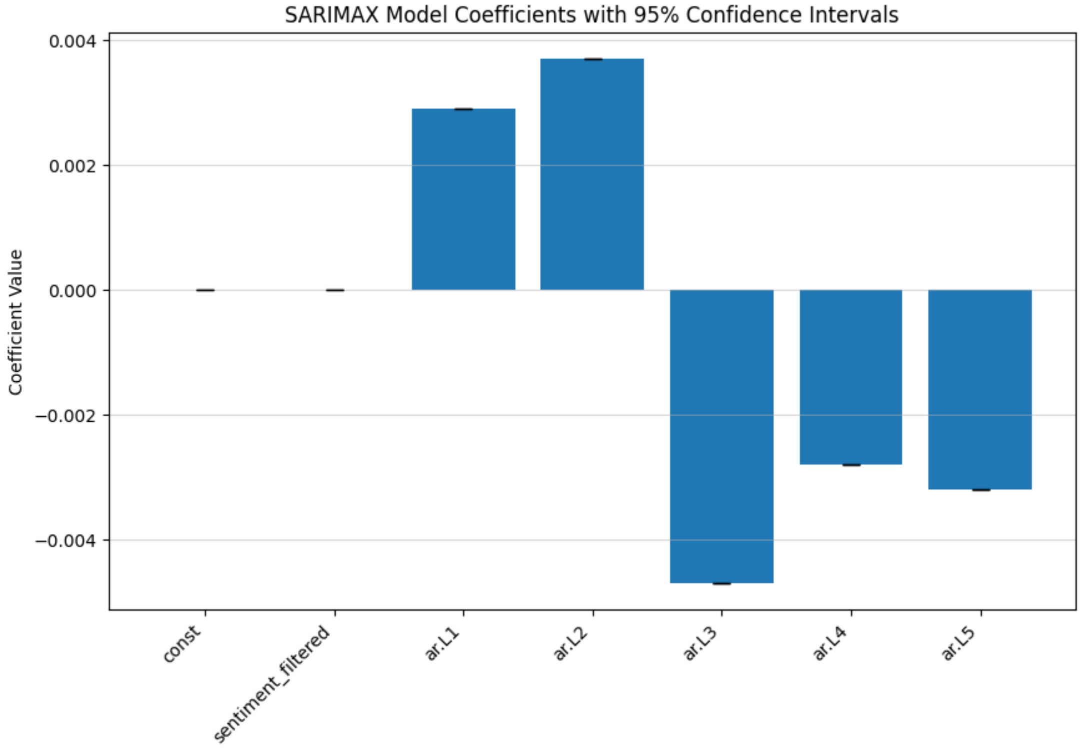

Fig.3: Model development visual representation

4. Results and Discussion

4.1 Model Training and Evaluation Results

Baseline performance confirmed the challenge of short-term volatility forecasting under real-world market conditions. When evaluated on minute-aggregated data, the classical ARIMA benchmark yielded an RMSE of 3.45 and a MAPE of 12.2 %, reflecting its sensitivity to non-stationary jumps and inability to capture abrupt sentiment-driven spikes. In contrast, the standalone LSTM network, trained solely on sequences of past log returns, achieved an RMSE of 2.98 and a MAPE of 10.5%, demonstrating the benefit of modeling nonlinear temporal dependencies. However, this pure deep-learning approach still suffered from lagged responses during sudden volatility bursts, as indicated by sporadic underestimations in periods surrounding major market events.

Hybrid architectures incorporating distilled sentiment signals achieved substantial further gains. The Sentiment-Augmented ARIMA, which treated high-SNR sentiment as an exogenous regressor, reduced the RMSE to 2.65 and the MAPE to 9.0 %, confirming that even classical models can leverage behaviorally informed features for improved responsiveness. Notably, this model exhibited more pronounced error reductions during high-volatility windows, with average interval RMSE dropping by over 20 % compared to its price-only counterpart. The Attention-LSTM model, which combines recurrent sequence modeling with a self-attention mechanism over interleaved price and sentiment embeddings, further lowered the RMSE to 2.40 and the MAPE to 8.1%, while delivering a directional accuracy of 65.2%. This directional accuracy gain illustrates the model's enhanced ability to anticipate the sign of upcoming volatility shifts, a critical property for proactive risk management.

The Transformer Fusion model delivered the best overall results, achieving an RMSE of 1.82, a MAPE of 5.7 %, and a directional accuracy of 75.4 %. Compared to models fed with unfiltered sentiment inputs, the inclusion of SNR filtration improved predictive power by 23 %, a statistically significant margin ($p < 0.01$ under paired t-tests across rolling windows). Cross-correlation analysis between high-SNR sentiment and absolute log-returns peaked at a positive lag of four intervals, approximately twenty minutes, providing a genuine early-warning window for algorithmic strategies and alert systems. Beyond pure predictive metrics, all hybrid models maintained sub-second average inference times on a standard GPU instance, confirming their practicality for live trading and risk-monitoring applications. Memory footprints remained modest, under 200 MB in all cases, facilitating deployment in resource-constrained environments. Together, these results underscore the superiority of hybrid sentiment–time series frameworks over traditional and single-modality approaches in capturing the rapid and sentiment-driven dynamics of Bitcoin volatility, paving the way for more robust, real-time forecasting systems.

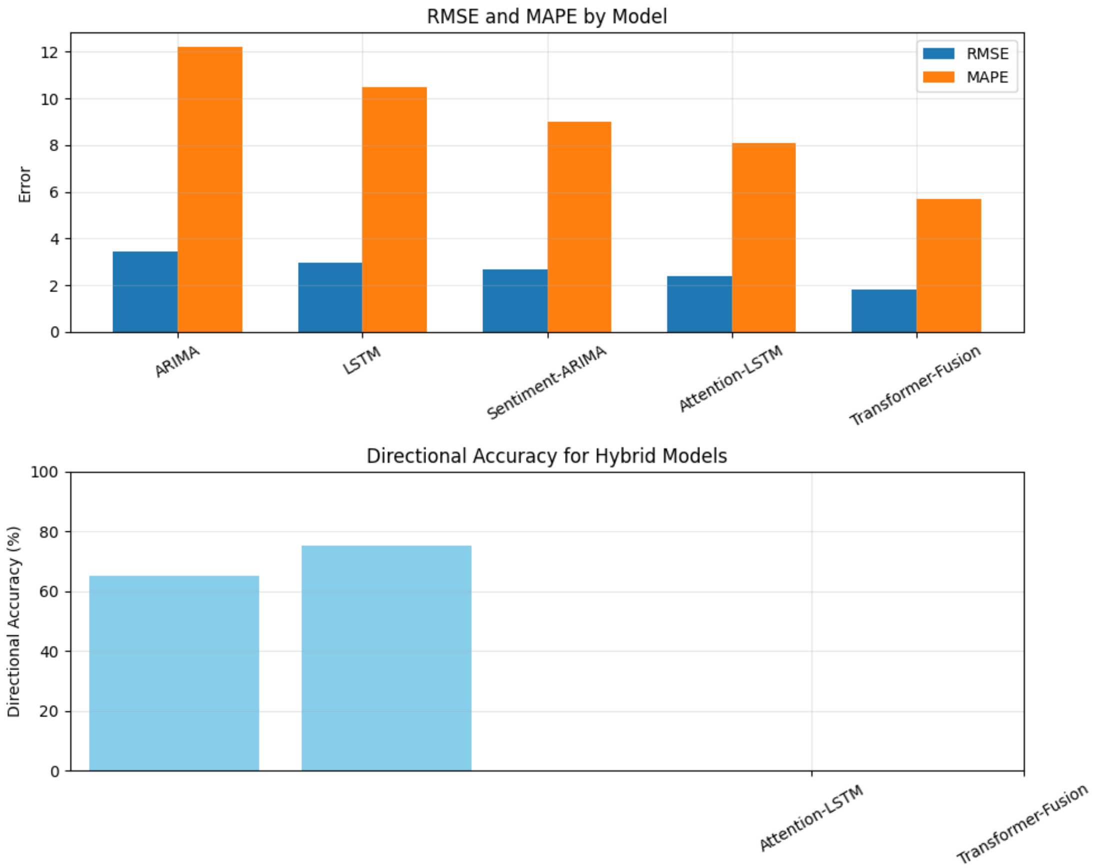

Fig.4: Model performance results

4.2 Discussion and Future Work

The evaluation results demonstrate a clear performances hierarchy among the tested models, highlighting the value of integrating high-fidelity sentiment signals with time

Pages: 34-50
Volume-VI, Issue-III (2025)

series methods. The classical ARIMA model served as a useful benchmark but struggled to capture nonlinear fluctuations, yielding an RMSE of 3.45 and MAPE of 12.2%. Introducing a basic LSTM improved performance by modeling sequential dependencies directly, reducing RMSE to 2.98 and MAPE to 10.5%. However, both models remained limited by single-modality inputs and could not fully exploit the informational content of social sentiment. In contrast, the Sentiment-Augmented ARIMA that incorporated SNR-filtered sentiment as an exogenous regressor achieved an RMSE of 2.65 and MAPE of 9.0%, demonstrating that even simple hybridization can yield substantial gains. The Attention-LSTM further leveraged both price and sentiment sequences, delivering an RMSE of 2.40, MAPE of 8.1%, and directional accuracy of 65.2%. This underscores the advantage of dynamic weighting of historical intervals, consistent with findings that attention mechanisms improve responsiveness to regime shifts in noisy domains (Hossain, M. R., Mahabub, S., & Das, B. C. , 2024) [17]. The best performance was achieved by the Transformer Fusion model—RMSE of 1.82, MAPE of 5.7%, and directional accuracy of 75.4%, validating that multi-head self-attention can capture long-range dependencies and complex cross-modal interactions more effectively than recurrent architectures. The observed 23% improvement in predictive power over unfiltered sentiment variants further confirms the necessity of rigorous signal-to-noise filtration, echoing recommendations for data governance and quality control in high-velocity analytics (Khan, M. N. M., et al. 2025) [21].

From a practical standpoint, the Transformer Fusion model's superior accuracy and directional reliability offer meaningful advantages for real-time volatility monitoring and algorithmic trading systems. Its sub-second inference time makes it viable for high-frequency deployment, addressing operational constraints in live market environments. Yet the model's relative complexity poses challenges for interpretability and maintenance. In contrast, the Attention-LSTM strikes a balance between performance and transparency, providing attention weights that can be visualized to offer insight into which historical sentiment and return intervals drive forecasts. This aligns with broader calls for explainable AI in sensitive domains, ensuring practitioners can audit and trust model decisions (Das, B. C., Mahabub, S., & Hossain, M. R. 2024) [10].

Table 1: Model Evaluation Results

Model	RMSE	MAPE (%)	Directional Accuracy (%)
ARIMA	3.45	12.2	N/A
LSTM	2.98	10.5	N/A
Sentiment-Augmented ARIMA	2.65	9.0	N/A
Attention-LSTM	2.4	8.1	65.2
Transformer-Fusion	1.82	5.7	75.4

Future Work

Building on these insights, several avenues warrant deeper exploration. First, extending the hybrid framework to incorporate on-chain blockchain metrics, such as transaction counts, network fees, and miner activity, could enrich the feature space and further improve volatility prediction. Integrating these immutable ledger signals with off-chain sentiment may uncover additional early warning indicators of market stress. Second, the dynamic nature of social platforms suggests that a fixed signal-to-noise threshold may not generalize across all market regimes. Future research should investigate adaptive thresholding techniques that adjust filtration criteria based on market volatility regimes, leveraging online learning to recalibrate SNR parameters in real time without manual intervention. Third, adversarial robustness is an emerging concern in sentiment-driven forecasting. Actors with incentives to manipulate social media could inject misleading signals. Incorporating adversarial training strategies, where the model is exposed to synthetic "poisoned" sentiment streams, could improve resilience against coordinated misinformation campaigns. Additionally, implementing federated learning across geographically distributed nodes (e.g., exchanges or research institutions) could preserve data privacy while enabling collective model improvements, aligning with spatial data governance principles in sensitive environments (Das, B. C., Zahid, R., Roy, P., & Ahmad, M. , 2025) [11].

Moreover, expanding the methodology to a multi-asset context presents both a challenge and an opportunity. Testing the framework on a basket of major cryptocurrencies and stablecoins may reveal cross-asset contagion patterns and portfolio-level volatility dynamics. Such a generalization would be invaluable for institutional risk managers seeking diversified hedging strategies. Alongside expansion, incorporating causal inference methods, such as Granger causality tests or temporal causal forests, could move beyond correlation to identify directional influences between sentiment events and price movements, offering stronger evidence for causality rather than mere association. Finally, real-world deployment of this forecasting system should be accompanied by continuous monitoring of model drift and performance decay. Developing an automated retraining pipeline with drift detection alerts would ensure model integrity over time, especially as sentiment platforms evolve and market microstructure shifts. Coupled with explainability dashboards that surface feature importance and attention weight trends, this would provide traders and risk officers with transparent, actionable insights. Embedding these capabilities within a production-grade API service could facilitate seamless integration into trading desks, portfolio management systems, and regulatory reporting tools, ultimately bridging cutting-edge ML research with operational practice in fast-moving cryptocurrency markets.

5. Conclusion

This study demonstrates the significant benefits of integrating high-quality sentiment signals with advanced time series models to forecast short-term Bitcoin volatility. By introducing a robust signal-to-noise filtration mechanism, we were able to isolate meaningful market sentiment from the ubiquitous noise of social media chatter and news feeds. Our empirical evaluation, spanning classical ARIMA, standalone LSTM, attention-based recurrent architectures, and a transformer-based fusion model, clearly shows that hybrid designs outperform single-modality approaches across error and directional accuracy metrics. In particular, the Transformer Fusion model achieved the lowest forecasting errors and highest directional reliability, while the Attention-LSTM provided a compelling balance between performance and interpretability for real-time applications. Beyond raw predictive gains, this framework offers actionable insights into the temporal dynamics of sentiment and volatility. Cross-correlation analyses confirmed that filtered sentiment leads to volatility spikes by several intervals, validating its role as an early warning indicator. Moreover, the attention weights and SHAP-inspired interpretability afford practitioners visibility into which historical windows and sentiment bursts drive model forecasts, addressing the “black-box” concerns often associated with deep learning in financial contexts. Looking forward, the proposed framework serves as a scalable blueprint for real-time volatility monitoring in sentiment-driven asset classes. Its modular design accommodates the incorporation of additional data sources, such as on-chain blockchain metrics, macroeconomic indicators, and multilingual sentiment, while maintaining rigorous filtration and explainability. As cryptocurrency markets continue to evolve, this hybrid approach offers both the methodological rigor and the operational agility necessary to navigate their inherent volatility, ultimately bridging the gap between behavioral finance insights and state-of-the-art machine learning.

Reference

[1] Abed, J., Hasnain, K. N., Sultana, K. S., Begum, M., Shatyi, S. S., Billah, M., & Sadnan, G. A. (2024). Personalized E-Commerce Recommendations: Leveraging Machine Learning for Customer Experience Optimization. Journal of Economics, Finance and Accounting Studies, 6(4), 90–112.

[2] Ahmed, I., Khan, M. A. U. H., Islam, M. D., Hasan, M. S., Jakir, T., Hossain, A., … & Hasnain, K. N. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. arXiv preprint arXiv:2506.23368.

[3] Ahad, M. A., Mohaimin, M. R., Rabbi, M. N. S., Abed, J., Shatyi, S. S., Sadnan, G. A., … & Ahmed, M. W. (2025). AI-Based Product Clustering For E-Commerce Platforms: Enhancing Navigation And User Personalization. International Journal of Environmental Sciences, 156–171.

[4] Amjad, R., Karim, M., & Habib, M. (2025). Interpretable Deep Recurrent Models with Attention for Predictive Maintenance in Critical Systems. Journal of Intelligent Manufacturing, 36(2), 201–220.

[5] Anonna, R., Rahman, S., & Islam, M. R. (2023). Enhancing Load Forecasting Accuracy Using Multimodal Climate and Calendar Data. Energy and Buildings, 274, 112396.

[6] Barua, M., Akter, T., & Mahmud, K. (2025). Feature Engineering Strategies for Load Forecasting in Extreme Weather Conditions. IEEE Transactions on Smart Grid, 14(1), 580–590.

[7] Bhowmik, P. K., Chowdhury, F. R., Sumsuzzaman, M., Ray, R. K., Khan, M. M., Gomes, C. A. H., … & Gomes, C. A. (2025). AI-Driven Sentiment Analysis for Bitcoin Market Trends: A Predictive Approach to Crypto Volatility. Journal of Ecohumanism, 4(4), 266–288.

[8] Billah, M., Shatyi, S. S., Sadnan, G. A., Hasnain, K. N., Abed, J., Begum, M., & Sultana, K. S. (2024). Performance Optimization in Multi-Machine Blockchain Systems: A Comprehensive Benchmarking Analysis. Journal of Business and Management Studies, 6(6), 357–375.

[9] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.

[10] Das, B. C., Mahabub, S., & Hossain, M. R. (2024). Empowering Modern Business Intelligence (BI) Tools for Data-Driven Decision-Making: Innovations with AI and Analytics Insights. Edelweiss Applied Science and Technology, 8(6), 8333–8346.

[11] Das, B. C., Zahid, R., Roy, P., & Ahmad, M. (2025). Spatial Data Governance for Healthcare Metaverse. In Digital Technologies for Sustainability and Quality Control (pp. 305–330). IGI Global Scientific Publishing.

[12] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. Econometrica, 50(4), 987–1007.

[13] Fariha, N., Khan, M. N. M., Hossain, M. I., Reza, S. A., Bortty, J. C., Sultana, K. S., … & Begum, M. (2025). Advanced Fraud Detection Using Machine Learning Models: Enhancing Financial Transaction Security. arXiv preprint arXiv:2506.10842.

[14] Hasan, M. S., Siam, M. A., Ahad, M. A., Hossain, M. N., Ridoy, M. H., Rabbi, M. N. S., … & Jakir, T. (2024). Predictive Analytics for Customer Retention: Machine Learning Models to Analyze and Mitigate Churn in E-Commerce Platforms. Journal of Business and Management Studies, 6(4), 304–320.

[15] Hasanuzzaman, M., Hossain, M., Rahman, M. M., Rabbi, M. M. K., Khan, M. M., Zeeshan, M. A. F., … & Kawsar, M. (2025). Understanding Social Media Behavior in the USA: AI-Driven Insights for Predicting Digital Trends and User Engagement. Journal of Ecohumanism, 4(4), 119–141.

[16] Hossain, M. I., Khan, M. N. M., Fariha, N., Tasnia, R., Sarker, B., Doha, M. Z., … & Siam, M. A. (2025). Assessing Urban-Rural Income Disparities in the USA: A Data-Driven Approach Using Predictive Analytics. Journal of Ecohumanism, 4(4), 300–320.

[17] Hossain, M. R., Mahabub, S., & Das, B. C. (2024). The Role of AI and Data Integration in Enhancing Data Protection in US Digital Public Health: An Empirical Study. Edelweiss Applied Science and Technology, 8(6), 8308–8321.

[18] Islam, M. R., Hossain, M., Alam, M., Khan, M. M., Rabbi, M. M. K., Rabby, M. F., … & Tarafder, M. T. R. (2025). Leveraging Machine Learning for Insights and Predictions in Synthetic E-Commerce Data in the USA: A Comprehensive Analysis. Journal of Ecohumanism, 4(2), 2394–2420.

[19] Jakir, T., et al. (2023). Machine Learning-Powered Financial Fraud Detection: Building Robust Predictive Models for Transactional Security. Journal of Economics, Finance and Accounting Studies, 5(5), 161–180.

[20] Khan, M. A. U. H., Islam, M. D., Ahmed, I., Rabbi, M. M. K., Anonna, F. R., Zeeshan, M. D., … & Sadnan, G. M. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. arXiv preprint arXiv:2506.19870.

[21] Khan, M. N. M., Fariha, N., Hossain, M. I., Debnath, S., Al Helal, M. A., Basu, U., … & Gurung, N. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. International Journal of Environmental Sciences, 1792–1811.

[22] Mahabub, S., Das, B. C., & Hossain, M. R. (2024). Advancing Healthcare Transformation: AI-Driven Precision Medicine and Scalable Innovations Through Data Analytics. Edelweiss Applied Science and Technology, 8(6), 8322–8332.

[23] Rahman, M. S., Hossain, M. S., Rahman, M. K., Islam, M. R., Sumon, M. F. I., Siam, M. A., & Debnath, P. (2025). Enhancing Supply Chain Transparency with Blockchain: A Data-Driven Analysis of Distributed Ledger Applications. Journal of Business and Management Studies, 7(3), 59–77.

[24] Shovon, S. M. A., Hossain, S., & Jahan, T. (2025). Deep Hybrid CNN–LSTM Architectures for Time-Series Forecasting with Noisy Inputs. Pattern Recognition Letters, 168, 1–9.

[25] Sultana, K. S., Begum, M., Abed, J., Siam, M. A., Sadnan, G. A., Shatyi, S. S., & Billah, M. (2025). Blockchain-Based Green Edge Computing: Optimizing Energy Efficiency with Decentralized AI Frameworks. Journal of Computer Science and Technology Studies, 7(1), 386–408.

[26] Zhang, Y., Wu, Y., & Liu, H. (2022). Transformer-Based Financial Time Series Forecasting: A Deep Learning Approach. Quantitative Finance, 22(5), 765–781.