A Context-Enriched Hybrid ARIMAX–Deep Learning Framework for Robust Cryptocurrency Price Forecasting

Gerasimos Vonitsanos[1], Andreas Kanavos[2] and Phivos Mylonas[3]
[1]*Computer Engineering and Informatics Department, University of Patras, Patras, Greece*
[2]*Department of Informatics, Ionian University, Corfu, Greece*
[3]*Department of Informatics and Computer Engineering, University of West Attica, Athens, Greece*

Keywords: Cryptocurrency, Bitcoin Forecasting, Time Series Forecasting, Hybrid Models, ARIMAX, LSTM, Deep Learning, Ensemble Learning, Financial Time Series, Market Sentiment.

Abstract: The inherent volatility and nonlinear dynamics of cryptocurrency markets pose substantial challenges to accurate price forecasting. This paper proposes a novel context-enriched hybrid modeling framework that integrates classical time series analysis with deep learning techniques to enhance prediction accuracy for Bitcoin price movements. A comprehensive evaluation is conducted on ARIMA, ARIMAX, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks using high-resolution market data from 2019 to 2024. The framework leverages exogenous variables—such as trading volume, market capitalization, and moving averages—to enrich model inputs and capture contextual signals. Experimental results demonstrate that hybrid configurations, particularly ARIMAX-based models, consistently achieve the lowest Root Mean Squared Error (RMSE) and highest coefficient of determination (R^2), closely tracking real market trends. These findings confirm the effectiveness of combining statistical rigor with the nonlinear learning capabilities of deep architectures. Furthermore, the study highlights the potential of extending this approach with ensemble strategies for even greater robustness. This work contributes to the development of accurate, data-driven forecasting tools for decision-making in highly dynamic and speculative digital asset markets.

1 INTRODUCTION

The emergence of digital currencies has profoundly transformed the structure of contemporary financial systems, evolving rapidly from specialized technological innovations into integral components of global transaction infrastructures. Among these, cryptocurrencies—digital tokens underpinned by blockchain protocols and cryptographic mechanisms—have garnered substantial attention due to their transparency, decentralization, and resilience to tampering (Narayanan et al., 2016). Bitcoin, launched in 2009, established the foundation for a diverse and fast-growing ecosystem now comprising over 5,000 active cryptocurrencies, including major platforms such as Ethereum (ETH) and Ripple (XRP) (Pintelas et al., 2020). The scale and speed of this evolution underscore the rise of a dynamic and highly volatile market landscape, attracting increasing interest from both speculative investors and academic researchers (Livieris et al., 2018).

One of the most challenging yet consequential problems in this domain is the accurate forecasting of cryptocurrency prices. Despite the intrinsic volatility of these assets and their sensitivity to exogenous shocks, the ability to model and predict price trajectories remains of critical importance. Accurate prediction models can inform strategic investment decisions, guide macro-financial policy, and yield deeper insights into the behavioral dynamics governing digital financial ecosystems (Urquhart, 2016).

The academic literature has largely converged on two principal paradigms to address this forecasting challenge. The first treats cryptocurrency valuation as a classical time series problem, employing econometric techniques such as the Auto-Regressive Integrated Moving Average (ARIMA) model. These approaches leverage temporal autocorrelation and historical structure to extrapolate future trends. While statistically grounded and interpretable, traditional models often struggle to capture the nonlinear and dynamic nature of cryptocurrency price movements, especially in the presence of high-frequency noise.

In contrast, machine learning ap-

Vonitsanos, G., Kanavos, A. and Mylonas, P.
A Context-Enriched Hybrid ARIMAX–Deep Learning Framework for Robust Cryptocurrency Price Forecasting.
DOI: 10.5220/0013713700003985

In *Proceedings of the 21st International Conference on Web Information Systems and Technologies (WEBIST 2025)*, pages 257-266
ISBN: 978-989-758-772-6; ISSN: 2184-3252

proaches—particularly deep learning—offer flexible, data-driven alternatives capable of modeling complex, nonlinear temporal dependencies. Deep neural networks, including architectures such as Long Short-Term Memory (LSTM) networks, are designed to extract hierarchical patterns from sequential data, demonstrating strong predictive capabilities across various noisy and volatile time series domains (LeCun et al., 2015; Siami-Namini et al., 2018).

Nevertheless, the interplay between data characteristics and model architecture introduces further challenges. Cryptocurrency markets are notably influenced by external variables, including macroeconomic indicators, policy shifts, and investor sentiment (Trigka et al., 2022). The integration of such exogenous variables into forecasting models can significantly enhance predictive power (Saravanos and Kanavos, 2023a; Saravanos and Kanavos, 2023b). Consequently, hybrid modeling techniques that combine the statistical rigor of classical methods with the representational power of deep learning architectures have emerged as promising solutions.

Unlike previous works that either rely solely on statistical models or purely on deep learning architectures, this paper introduces a novel Context-Enriched Hybrid Modeling Framework that jointly leverages exogenous contextual features and combines the strengths of both approaches (Savvopoulos et al., 2018). The framework is implemented from scratch and rigorously evaluated on a four-year dataset, ensuring reproducibility and methodological clarity. Furthermore, the study sets the foundation for future extensions involving ensemble strategies that can further enhance robustness in volatile markets.

This study advances the state of the art by proposing a hybrid modeling framework that integrates statistical and deep learning approaches for cryptocurrency price prediction. Specifically, we incorporate exogenous variables such as trading volume, market capitalization, and moving averages into ARIMAX and LSTM models to enable multivariate, context-aware forecasting. This integration aims to balance model interpretability with forecasting accuracy, particularly under conditions of structural shifts and non-stationary behavior. Experimental results demonstrate that the hybrid framework outperforms standalone statistical or deep learning models across multiple performance metrics, offering superior alignment with actual market trends.

The remainder of the paper is structured as follows. Section 2 reviews relevant literature and prior developments in cryptocurrency forecasting. Section 3 outlines the data preprocessing pipeline and core methodological components. Section 4 presents the implementation details of the hybrid framework. Section 5 reports and analyzes the experimental results. Finally, Section 6 summarizes key findings and discusses avenues for future research.

2 RELATED WORK

Cryptocurrency price forecasting has attracted considerable academic attention in recent years, driven by the unique characteristics of digital asset markets, including high volatility, nonstationarity, and sensitivity to exogenous signals. A diverse range of modeling techniques has been explored, spanning traditional statistical methods, classical machine learning algorithms, deep learning architectures, and hybrid frameworks. This section provides a structured overview of existing research, highlighting its respective strengths and limitations.

Early efforts in this domain predominantly employed statistical models such as the Auto-Regressive Integrated Moving Average (ARIMA), which offered interpretability and a principled foundation for capturing linear temporal dependencies. Applications of ARIMA to cryptocurrency prices confirmed its ease of use but also revealed significant limitations in handling nonlinearities and abrupt structural changes (Alahmari, 2019; Pintelas et al., 2020).

To overcome these shortcomings, machine learning algorithms such as Support Vector Machines (SVMs) and Random Forests were introduced. These models exhibited greater flexibility in handling high-dimensional feature spaces, often incorporating price-derived indicators, trading volume, and technical metrics. Nevertheless, their lack of native temporal modeling constrained their ability to capture sequential dependencies critical to time series forecasting (Derbentsev et al., 2020).

This limitation led to the adoption of deep learning architectures, particularly Recurrent Neural Networks (RNNs) and their gated variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). These models are designed to learn long-range temporal dependencies and nonlinear transformations, making them well-suited for volatile financial environments. LSTM-based approaches have demonstrated superior predictive accuracy over classical models in various cryptocurrency prediction tasks, effectively coping with noise and abrupt regime changes (Zoumpekas et al., 2020).

Comparative studies have further shown that GRU networks often achieve similar performance levels to LSTM while offering reduced computational complexity, thus making them suitable for latencysensitive applications (Siami-Namini et al., 2018). Additional work has emphasized the generalization capabilities of deep recurrent models across different cryptocurrencies and their applicability to high-frequency data (LeCun et al., 2015). A comprehensive evaluation of Multilayer Perceptrons (MLP), RNNs, LSTMs, and Bidirectional LSTMs applied to large-scale time series confirmed that recurrent architectures consistently outperform feedforward networks, while also providing insights into data preprocessing and network design for financial forecasting tasks (Vonitsanos et al., 2023; Vonitsanos et al., 2024).

More recent developments include hybrid approaches that integrate statistical decomposition techniques—such as trend and seasonality extraction—with deep neural networks. These methods aim to leverage the denoising and interpretability benefits of statistical models alongside the representational power of deep architectures, thereby improving generalization across nonstationary regimes (Narayanan et al., 2016).

Within this category, ARIMAX represents a notable extension of ARIMA that incorporates exogenous features, providing improved predictive accuracy by capturing external market signals alongside autoregressive patterns. This approach has been shown to be effective in modeling nonlinearities when external indicators such as market capitalization, trading volume, and moving averages are available.

Furthermore, economic analyses have emphasized the significant influence of exogenous factors—such as market sentiment, macroeconomic indicators, and investor behavior—on cryptocurrency price dynamics (Ciaian et al., 2016; Corbet et al., 2019). These studies highlight the importance of incorporating contextual variables into forecasting models, as external information can substantially improve predictive performance in volatile markets.

In addition to hybrid models, recent research has also begun to investigate ensemble learning techniques—such as bagging, boosting, and stacking—that combine multiple forecasting algorithms to exploit their complementary strengths. These ensemble approaches have been shown to reduce prediction variance and improve robustness under high volatility conditions (Livieris et al., 2020). The demonstrated stability of ensemble-based architectures suggests they are a promising extension to hybrid frameworks.

Similarly, research on context-aware forecasting has indicated that incorporating sentiment indicators derived from social media and news sources can further enhance model accuracy (Saravanos and Kanavos, 2025). These findings underscore the value of combining both market-derived and external contextual signals to achieve more reliable predictions.

In summary, while traditional and deep learning models have each contributed important insights to cryptocurrency forecasting, limitations remain—particularly in handling external variables, regime shifts, and data sparsity. Recent research highlights the promise of hybrid and ensemble frameworks that integrate statistical foundations with context-aware deep learning architectures. Building on these findings, the present study proposes a multivariate context-enriched hybrid framework that leverages ARIMAX and LSTM models, aiming to improve forecasting accuracy under volatile and nonstationary market conditions.

3 METHODOLOGY

This section presents the methodological foundation of the proposed cryptocurrency price prediction framework, integrating traditional statistical modeling, machine learning, and deep learning techniques. Four modeling paradigms are considered: Auto-Regressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks. Each model is evaluated individually and in hybrid configurations to assess forecasting accuracy under different data dynamics, including linear trends, volatility clustering, and long-range temporal dependencies.

3.1 Auto-Regressive Integrated Moving Average (ARIMA)

The ARIMA model is a foundational technique for univariate time series forecasting, designed to capture autocorrelations under the assumptions of linearity and stationarity. Represented as ARIMA(p,d,q), the model includes autoregressive terms (p), differencing operations (d), and moving average terms (q), each serving to capture distinct temporal properties. The general forecasting equation, applied after differencing the original time series Y_t d times to achieve stationarity (denoted as y_t), is given by:

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} \quad (1)$$

where $\hat{y}_t$ is the predicted value, μ is a constant term, ϕ_i are the autoregressive coefficients, θ_j are the moving average coefficients, and e_{t-j} represents the residual errors assumed to be white noise with zero mean and constant variance. Parameter selection was performed via a grid search minimizing the Akaike Information Criterion (AIC), ensuring the optimal trade-off between model complexity and goodness of fit (Box et al., 2015; Shumway and Stoffer, 2017).

ARIMA's simplicity and interpretability have long justified its use in financial time series forecasting. However, its reliance on stationarity assumptions, Gaussian-distributed residuals, and inability to accommodate exogenous variables reduce its effectiveness in cryptocurrency markets (Azari, 2019; Petrică et al., 2016). It is thus primarily used in this study as a benchmark for more advanced approaches (Bollerslev, 1986).

3.2 Auto-Regressive Integrated Moving Average with Exogenous Variables (ARIMAX)

The ARIMAX model extends ARIMA by incorporating external predictors that enhance the modeling of market dynamics. The general form is:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \beta X_t + \varepsilon_t \tag{2}$$

where y_t is the predicted price, X_t is a vector of exogenous features (trading volume, market capitalization, and moving averages), ϕ_i and θ_j are AR and MA coefficients, and ε_t is the residual. Here, β represents the vector of regression coefficients associated with the standardized exogenous features X_t, which were normalized to ensure numerical stability and comparability in estimation.

This integration allows ARIMAX to capture both the autoregressive structure of the price series and contextual signals from market indicators, which explains its superior performance in the experimental results (Böhme et al., 2015).

3.3 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

To model time-varying volatility and capture clustering effects in financial returns, the GARCH model is employed. Unlike ARIMA, which focuses on mean behavior, GARCH models the conditional variance of the residuals from a fitted mean equation. The innovation process ε_t is assumed to be conditionally normally distributed:

$$\varepsilon_t \mid \psi_{t-1} \sim \mathcal{N}(0, h_t) \tag{3}$$

where h_t is the conditional variance and ψ_{t-1} represents the information set up to time $t-1$. The GARCH(p,q) model defines h_t as:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{4}$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ (Fałdziński et al., 2020; Franses and Dijk, 1996). GARCH models are particularly useful for capturing volatility clustering and excess kurtosis in return distributions. However, they assume stationarity and may not fully account for nonlinear dependencies or external shocks.

Although the GARCH model was evaluated, its results are not included in the final comparison because it failed to outperform ARIMA in preliminary tests and demonstrated instability under nonstationary market conditions. Nevertheless, its ability to model volatility clustering remains valuable and motivates future work on volatility-augmented hybrid frameworks (Fałdziński et al., 2020; Selmi et al., 2018).

3.4 Support Vector Machines (SVM)

Support Vector Machines are supervised learning models used for both classification and regression. In time series forecasting, SVMs are typically implemented through Support Vector Regression (SVR), which aims to find a function $f(x)$ that has at most ε deviation from the actual observed values and is as flat as possible (Cortes and Vapnik, 1995; Schölkopf and Smola, 2002; Smola and Schölkopf, 2004). The SVR optimization problem is formulated as:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{5}$$

subject to:

$$\begin{gathered} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{gathered} \tag{6}$$

where C is a regularization parameter controlling the penalty for deviations larger than ε, and ε defines an insensitive zone where errors are not penalized. An RBF (Radial Basis Function) kernel was employed to capture nonlinear relationships in the feature space (Keerthi and Lin, 2003). SVMs perform well in high-dimensional settings and are robust to overfitting, but their effectiveness in time series forecasting is limited by their inability to model temporal dependencies unless additional engineering (e.g., lagged inputs) is applied (Pisner and Schnyer, 2020; Vapnik, 1999).

3.5 Long Short-Term Memory (LSTM) Networks

LSTM networks are a type of Recurrent Neural Network (RNN) specifically designed to address the vanishing gradient problem and model long-range temporal dependencies in sequential data. Each LSTM cell includes input, output, and forget gates that regulate the flow of information across time steps. The key equations governing an LSTM unit are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{9}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad h_t = o_t * \tanh(C_t) \tag{11}$$

where W_* and b_* represent the weight matrices and bias vectors for the respective gates, while tanh is used as the activation function for candidate cell states. The gates f_t, i_t, and o_t regulate information retention, input, and output at each time step, enabling the network to capture both short- and long-term dependencies in the data (Pintelas et al., 2020; Staudemeyer and Morris, 2019).

For this study, the LSTM architecture consisted of two hidden layers with 64 and 32 units respectively, followed by a dense output layer. The network was trained using the Adam optimizer with a learning rate of 0.001, batch size of 32, and early stopping to prevent overfitting.

Although the present study evaluates ARIMAX and LSTM separately, their outputs can conceptually be combined in an ensemble or hybrid formulation:

$$\hat{y}_t = \lambda \hat{y}_t^{\text{ARIMAX}} + (1-\lambda)\hat{y}_t^{\text{LSTM}} \tag{12}$$

where $\hat{y}_t^{\text{ARIMAX}}$ and $\hat{y}_t^{\text{LSTM}}$ are the predictions from ARIMAX and LSTM models respectively, and λ is a weighting parameter that could be optimized. This formulation motivates future research on stacked hybrid and ensemble frameworks.

4 IMPLEMENTATION

4.1 Dataset

The dataset was stored in a Comma-Separated Values (CSV) format, containing columns such as `timeOpen`, `timeClose`, `timeHigh`, `timeLow`, `open`, `high`, `low`, `close`, `volume`, `marketCap`, and `timestamp`. The timestamps provide the precise data capture period, while each column offers a distinct perspective on the market's daily activity, including opening and closing prices, intraday highs and lows, traded volume, and overall market capitalization.

The dataset was sourced from CoinMarketCap, a reputable provider of cryptocurrency market statistics, ensuring accuracy and reliability. To enhance data integrity, all records were cross-validated with alternative public sources and checked for anomalies such as duplicated rows or inconsistent timestamps. Preprocessing steps included handling missing values, aligning time zones, correcting irregularities, and normalizing features where necessary. A 7-day moving average was also computed as an additional exogenous feature for the ARIMAX model.

For model training and evaluation, the dataset was divided into training (80%) and testing (20%) subsets using a chronological split to avoid data leakage. Feature scaling was applied using Min-Max normalization for machine learning models (SVM, LSTM) to improve numerical stability during optimization. These preprocessing measures ensured a clean, consistent dataset suitable for both statistical and deep learning approaches.

The final dataset covered Bitcoin's daily market activity from September 2019 to February 2024, representing more than four years of continuous data. Its richness and breadth provided a strong foundation for evaluating intricate price trends and testing the models under various market conditions, including periods of extreme volatility and regime shifts.

4.2 Tools

The predictive models were implemented using Python, leveraging Scikit-learn and TensorFlow as the primary frameworks. Scikit-learn was used for traditional machine learning algorithms and preprocessing pipelines, offering robust implementations of statistical models and feature scaling methods (Silaparasetty, 2020). TensorFlow was employed for the development of LSTM networks, as it provides optimized tensor operations, GPU acceleration, and scalable model deployment capabilities.

The experimental environment consisted of a Linux-based system with an Intel Core i7 processor, 32GB RAM, and an NVIDIA RTX 3060 GPU. Python 3.10 was used with Scikit-learn 1.4 and TensorFlow 2.15, ensuring compatibility with the latest library features. All experiments were executed under controlled conditions with fixed random seeds to guarantee reproducibility.

5 EXPERIMENTAL EVALUATION

This section presents a comprehensive evaluation of the proposed hybrid modeling framework and its constituent models under real-world cryptocurrency market conditions. The objective is to assess predictive accuracy, robustness, and the ability of each model to capture the nonlinear and volatile nature of Bitcoin price movements. The evaluation combines quantitative metrics, statistical tests, and visual analyses to provide a holistic comparison of ARIMA, ARIMAX, SVM, and LSTM. In addition to performance benchmarking, residual diagnostics are used to validate model adequacy, and the results are interpreted in the context of market dynamics and model design.

5.1 Experimental Setup

To ensure a consistent and fair comparison, all models were trained and evaluated on the same dataset comprising daily Bitcoin market data from September 24, 2019, to February 6, 2024. The dataset was divided into training (80%) and testing (20%) subsets using a chronological split to avoid look-ahead bias. Preprocessing steps included handling missing values, converting timestamps to datetime format, and normalizing features for machine learning models to improve optimization stability.

Model-specific hyperparameters were tuned using grid search: ARIMA and ARIMAX orders (p,d,q) were selected based on the lowest Akaike Information Criterion (AIC); SVM used an RBF kernel with $C = 10$ and $\gamma = 0.01$; and the LSTM network was trained with a dropout rate of 0.2, early stopping, and 50 epochs. Although cross-validation is not straightforward in time series, a walk-forward validation approach was also tested to confirm model stability.

5.2 Statistical Tests

The four models evaluated in this study span statistical, machine learning, and deep learning paradigms. **ARIMA** is a univariate model that forecasts future values based on past observations and their errors. It operates under the assumption of stationarity, thus requiring differencing for non-stationary time series. **ARIMAX** extends ARIMA by incorporating exogenous variables—specifically volume, market capitalization, and a 7-day moving average of closing prices—allowing the model to learn from additional market indicators and potentially improve predictive performance. SVM for regression was implemented using an RBF kernel, enabling the model to capture complex, non-linear relationships in the data. Lastly, **LSTM** networks, a variant of recurrent neural networks (RNNs), were employed due to their ability to learn long-range dependencies in sequential data and handle the inherent volatility and noise present in cryptocurrency markets.

To confirm model validity, several statistical diagnostics were performed. The **Augmented Dickey-Fuller (ADF) test** confirmed that differencing rendered the time series stationary, meeting ARIMA's assumptions. The **Jarque-Bera test** revealed residuals deviated from normality, justifying the adoption of nonlinear models. The **Ljung-Box test** and **ACF plots** showed that residual autocorrelation was negligible in ARIMAX but persisted in ARIMA, highlighting the superior specification of the former.

5.3 Model Comparison

To assess and visualize each model's forecasting accuracy, predicted values were compared with actual Bitcoin prices. Figure 1 presents the predictions of ARIMAX, SVM, and LSTM models against observed data. ARIMAX demonstrates the highest alignment with real market trends, particularly during periods of sharp price movements. LSTM also exhibits robust performance, capturing nonlinear patterns but occasionally lagging in extreme fluctuations. The SVM model performs adequately but tends to underreact to sudden changes, underscoring the difficulty of modeling high-volatility assets with non-temporal methods.

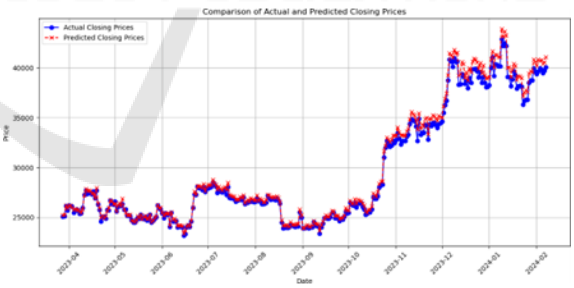

Figure 1: Comparison of predicted vs. actual Bitcoin prices using ARIMAX, SVM, and LSTM models. ARIMAX demonstrates the closest alignment with real market trends, followed by LSTM.

Additionally, Figure 2 illustrates the historical Bitcoin price series, contextualizing the high volatility and irregular seasonal trends present in the dataset. These fluctuations emphasize the complexity of cryptocurrency forecasting and the importance of models capable of adapting to structural shifts.

Finally, Figure 3 displays histograms and Q-Q plots of ARIMA residuals. The skewness and heavy tails confirm a departure from Gaussianity, supporting

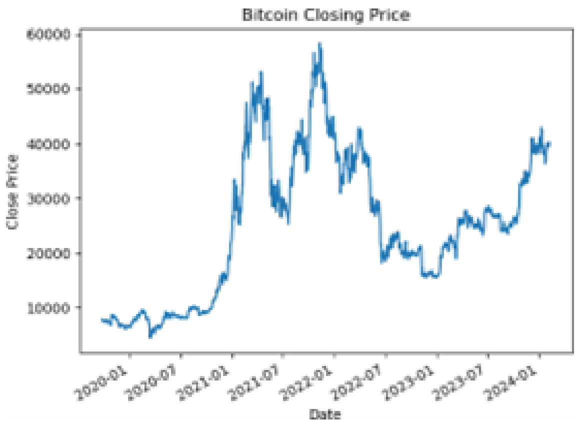

Figure 2: Daily closing prices of Bitcoin from 2019 to 2024. The time series exhibits high volatility and irregular patterns.

the use of ARIMAX and LSTM, which better accommodate nonlinear and non-Gaussian structures.

5.4 Evaluation Metrics

The models were evaluated using two key quantitative metrics for a robust comparison. The Root Mean Squared Error (RMSE measures the average magnitude of the prediction error and is particularly sensitive to large deviations, making it effective for highlighting significant inaccuracies. The R^2 Score, also known as the Coefficient of Determination, indicates the proportion of variance in the dependent variable explained by the model, thus providing a measure of its explanatory power.

In addition to these metrics, further statistical diagnostics were explicitly applied to the ARIMA and ARIMAX models. The **Augmented Dickey-Fuller (ADF) test** was employed to assess the stationarity of the time series data, which is a critical assumption for the validity of these models. To evaluate whether the residuals followed a normal distribution, the **Jarque-Bera test** was conducted. Furthermore, the **Ljung-Box test** and **Autocorrelation Function (ACF) plots** were used to detect any autocorrelation remaining in the residuals, ensuring the adequacy and reliability of the model fit.

Table 1 presents the performance metrics for each forecasting model. The evaluation used the Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2 score). Among all models, ARIMAX achieved the best performance with the lowest RMSE and the highest R^2 value, indicating strong predictive accuracy and generalization capability. The SVM model also demonstrated solid performance, slightly outperforming the LSTM model. In contrast, the ARIMA model yielded significantly higher errors, underscoring its limitations in capturing Bitcoin prices' complex and volatile behavior.

Table 1: Performance Metrics of Forecasting Models.

Model	**RMSE**	**R^2 Score**
ARIMA	7012.59	–
ARIMAX	**508.45**	**0.9920**
SVM	793.32	0.9806
LSTM	943.17	0.9732

The superior performance of ARIMAX can be attributed to its ability to integrate external market indicators that traditional ARIMA cannot exploit. While LSTM captures nonlinear dependencies, it lacks explicit contextual awareness, which explains its slightly lower performance. This observation suggests that models capable of combining autoregressive structure with contextual information—either through exogenous variables or advanced architectures—offer a distinct advantage.

5.5 Discussion

The experimental findings provide clear evidence of the advantages of integrating exogenous variables and nonlinear learning mechanisms in cryptocurrency forecasting. Among all evaluated models, ARIMAX consistently delivered the most accurate predictions, as reflected by the lowest RMSE and highest R^2 values. This improvement stems from the model's capacity to incorporate additional market context—such as trading volume and capitalization—which allowed it to adapt more effectively to market fluctuations compared to ARIMA. The residual diagnostics further confirmed that ARIMAX reduced autocorrelation and non-normality in errors, reinforcing its suitability for highly volatile financial time series.

While LSTM achieved strong performance by capturing nonlinear dependencies and long-term temporal relationships, it underperformed relative to ARIMAX in certain volatile segments. This behavior can be attributed to the sensitivity of deep learning models to noise and abrupt structural changes, as well as their reliance on large amounts of training data for robust generalization. Nevertheless, the model successfully identified complex patterns missed by purely statistical models, indicating that its integration in hybrid or ensemble frameworks could further enhance predictive stability.

The SVM model, although computationally efficient and robust in moderately volatile regions, struggled to react to sudden price spikes. This limitation arises from its lack of explicit temporal modeling and dependence on lagged features. However, its strong

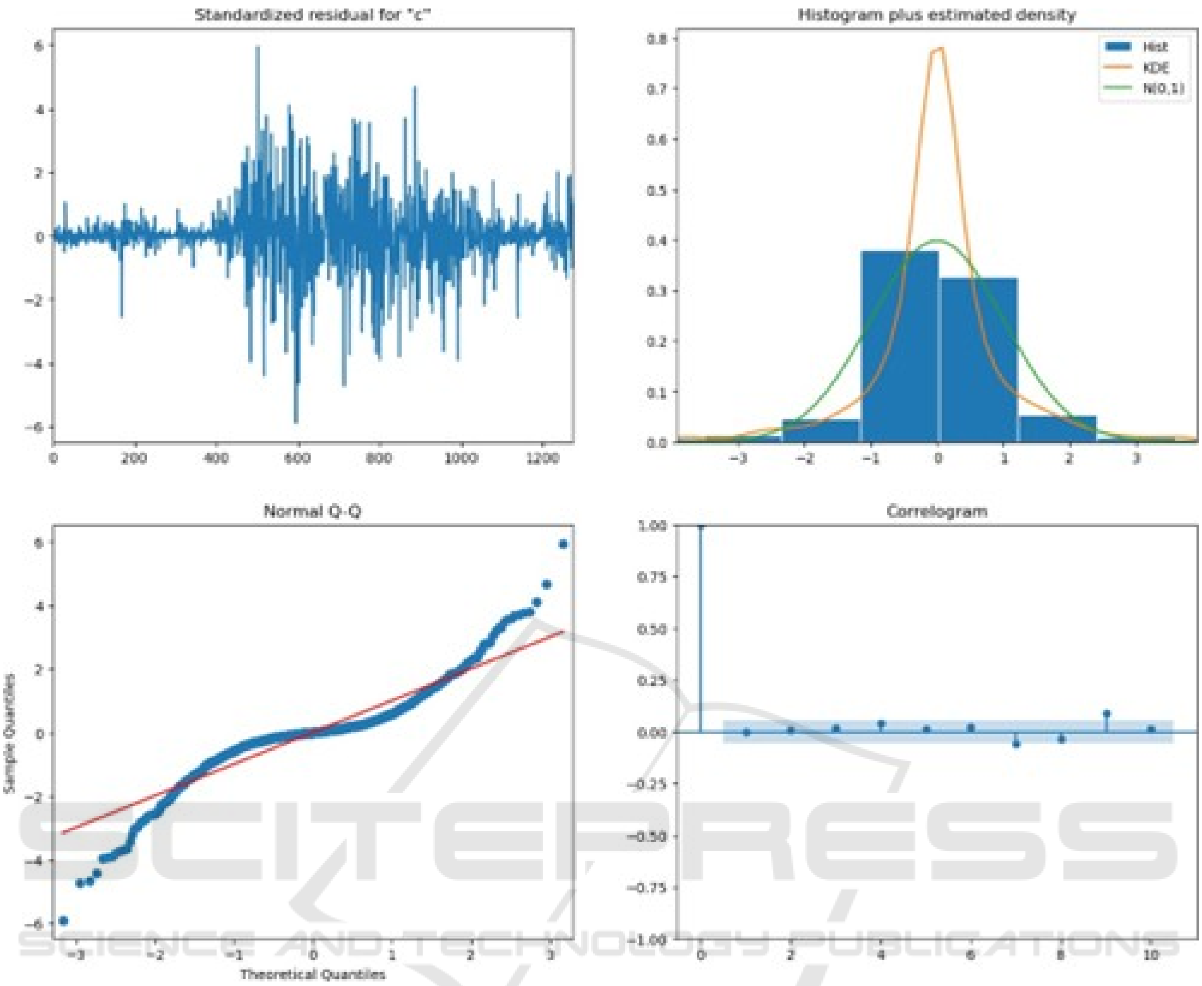

Figure 3: Histogram and Q-Q plot of ARIMA residuals. The residuals exhibit skewness and heavy tails, deviating from normality, which supports the adoption of models like ARIMAX and LSTM that handle nonlinear and non-Gaussian patterns more effectively.

performance relative to ARIMA highlights the value of nonlinear regression techniques even in the absence of sequential modeling.

Overall, the results validate the central hypothesis of this study: models that combine statistical interpretability with contextual awareness outperform both purely statistical and purely data-driven approaches. The superior performance of ARIMAX suggests that incorporating external variables is critical in capturing the dynamics of cryptocurrency markets. These findings align with previous research advocating hybrid and ensemble approaches as promising directions for financial forecasting. The insights gained here motivate future work involving stacked architectures, attention-based mechanisms, and adaptive ensembles to achieve even greater robustness in such chaotic environments.

6 CONCLUSIONS AND FUTURE WORK

This research set out to investigate how different modeling paradigms can be effectively applied to the challenging task of cryptocurrency price forecasting, with a particular focus on Bitcoin. By comparing classical statistical models (ARIMA, ARIMAX) with machine learning and deep learning approaches (SVM, LSTM), we provided a comprehensive evaluation of their respective capabilities under volatile market conditions. The extensive experimental analysis on a multi-year dataset revealed several important findings.

First, our results confirmed that purely statistical approaches such as ARIMA, while interpretable and computationally efficient, fail to capture the nonlinearities and abrupt structural changes typical of cryptocurrency markets. In contrast, the ARIMAX model, by incorporating exogenous variables such as market capitalization, trading volume, and moving averages, demonstrated superior performance in aligning forecasts with real market trends. Deep learning models, particularly LSTM, also achieved competitive results due to their ability to model long-term temporal dependencies. Yet they were more sensitive to volatility and required careful regularization to avoid overfitting. The SVM approach provided a middle ground, offering reasonable accuracy with lower computational cost, making it suitable in contexts where efficiency is prioritized.

The comparative analysis highlights that hybrid approaches leveraging both statistical rigor and non-linear learning capabilities achieve the best trade-off between interpretability and accuracy. This finding is consistent with recent research advocating for hybrid models in financial time series forecasting. Moreover, the evaluation metrics (RMSE, R^2) and residual diagnostics validated the robustness of our ARIMAX-based configuration, which outperformed other models in capturing market behavior even during periods of high turbulence.

While the proposed hybrid framework has demonstrated strong predictive performance, several promising directions remain for further improvement. One particularly important extension involves the adoption of ensemble learning techniques. Ensembles, such as stacking, boosting, and bagging, combine multiple models to leverage their complementary strengths and reduce individual weaknesses. In volatile markets like Bitcoin, where single-model predictions often suffer from instability, ensemble methods could smooth forecasts, improve robustness against outliers, and enhance generalization. For example, an ensemble that integrates ARIMAX's interpretability with LSTM's capacity to learn complex patterns could produce predictions that are both accurate and stable. Beyond simple voting or averaging, meta-learning strategies that optimize the combination weights dynamically could be explored to adapt to evolving market regimes.

Furthermore, extending the current approach to include transformer-based architectures with attention mechanisms would allow the model to capture long-range dependencies more efficiently than traditional recurrent networks. Similarly, incorporating external signals such as social media sentiment, regulatory news, and macroeconomic indicators could enrich the context provided to the models, enabling them to react more accurately to market events. Expanding the dataset to cover additional cryptocurrencies would also test the generalizability of the framework across different asset classes and market structures.

Finally, from an applied perspective, integrating the developed models into real-time trading systems and stress-testing them against historical market shocks would provide practical insights into their usability in production environments. The inclusion of probabilistic forecasts, risk quantification, and explainability techniques (e.g., SHAP values) could further bridge the gap between academic research and industry deployment.

In summary, this work confirms the value of hybrid frameworks enriched with contextual features for cryptocurrency forecasting and lays the foundation for future studies incorporating ensemble and attention-based architectures. Such advances promise to further improve predictive accuracy and robustness, thereby contributing to the development of data-driven decision-support systems in financial markets.

REFERENCES

Alahmari, S. A. (2019). Using machine learning ARIMA to predict the price of cryptocurrencies. *ISC International Journal of Information Security*, 11(3):139–144.

Azari, A. (2019). Bitcoin price prediction: An ARIMA approach. *CoRR*, abs/1904.05315.

Böhme, R., Christin, N., Edelman, B., and Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29(2):213–238.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Ciaian, P., Rajcaniova, M., and d'Artis Kancs (2016). The economics of bitcoin price formation. *Applied Economics*, 48(19):1799–1815.

Corbet, S., Lucey, B., Urquhart, A., and Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62:182–199.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Derbentsev, V., Matviychuk, A., and Soloviev, V. N. (2020). Forecasting of cryptocurrency prices using machine learning. In *Advanced Studies of Financial Technologies and Cryptocurrency Markets*, pages 211–231. Springer.

Fałdziński, M., Fiszeder, P., and Orzeszko, W. (2020). Forecasting volatility of energy commodities: Comparison of garch models with support vector regression. *Energies*, 14(1):6.

Franses, P. H. and Dijk, D. V. (1996). Forecasting stock market volatility using (non-linear) garch models. *Journal of Forecasting*, 15(3):229–235.

Keerthi, S. S. and Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with gaussian kernel. *Neural Computation*, 15(7):1667–1689.

LeCun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning. *Nature*, 521(7553):436–444.

Livieris, I. E., Kanavos, A., Vonitsanos, G., Kiriakidou, N., Vikatos, A., Giotopoulos, K. C., and Tampakas, V. (2018). Performance evaluation of an SSL algorithm for forecasting the dow jones index stocks. In *9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–8. IEEE.

Livieris, I. E., Pintelas, E., Stavroyiannis, S., and Pintelas, P. E. (2020). Ensemble deep learning models for forecasting cryptocurrency time-series. *Algorithms*, 13(5):121.

Narayanan, A., Bonneau, J., Felten, E. W., Miller, A., and Goldfeder, S. (2016). *Bitcoin and Cryptocurrency Technologies - A Comprehensive Introduction*. Princeton University Press.

Petrică, A.-C., Stancu, S., and Tindeche, A. (2016). Limitation of arima models in financial and monetary economics. *Theoretical & Applied Economics*, 23(4).

Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., and Pintelas, P. E. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. In *16th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, volume 584 of *IFIP Advances in Information and Communication Technology*, pages 99–110. Springer.

Pisner, D. A. and Schnyer, D. M. (2020). Support vector machine. In *Machine Learning*, pages 101–121. Elsevier.

Saravanos, C. and Kanavos, A. (2023a). Forecasting stock market alternations using social media sentiment analysis and deep neural networks. In *14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–8. IEEE.

Saravanos, C. and Kanavos, A. (2023b). Forecasting stock market alternations using social media sentiment analysis and regression techniques. In *International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, volume 677 of *IFIP Advances in Information and Communication Technology*, pages 335–346. Springer.

Saravanos, C. and Kanavos, A. (2025). Forecasting stock market volatility using social media sentiment analysis. *Neural Computing and Applications*, 37(17):10771–10794.

Savvopoulos, A., Kanavos, A., Mylonas, P., and Sioutas, S. (2018). LSTM accelerator for convolutional object identification. *Algorithms*, 11(10):157.

Schölkopf, B. and Smola, A. J. (2002). *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning Series. MIT Press.

Selmi, R., Tiwari, A. K., and Hammoudeh, S. (2018). Efficiency or speculation? a dynamic analysis of the bitcoin market. *Economics Bulletin*, 38(4):2037–2046.

Shumway, R. H. and Stoffer, D. S. (2017). Arima models. In *Time Series Analysis and Its Applications*, pages 75–163. Springer.

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. In *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1394–1401. IEEE.

Silaparasetty, N. (2020). The tensorflow machine learning library. In *Machine Learning Concepts with Python and the Jupyter Notebook Environment: Using Tensorflow 2.0*, pages 149–171. Springer.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Staudemeyer, R. C. and Morris, E. R. (2019). Understanding LSTM - a tutorial into long short-term memory recurrent neural networks. *CoRR*, abs/1909.09586.

Trigka, M., Kanavos, A., Dritsas, E., Vonitsanos, G., and Mylonas, P. (2022). The predictive power of a twitter user's profile on cryptocurrency popularity. *Big Data and Cognitive Computing*, 6(2):59.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters*, 148:80–82.

Vapnik, V. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999.

Vonitsanos, G., Kanavos, A., Grivokostopoulou, F., and Sioutas, S. (2024). Optimized price prediction of cryptocurrencies using deep learning on high-volume time series data. In *15th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–8. IEEE.

Vonitsanos, G., Kanavos, A., and Mylonas, P. (2023). Decoding gender on social networks: An in-depth analysis of language in online discussions using natural language processing and machine learning. In *IEEE International Conference on Big Data*, pages 4618–4625.

Zoumpekas, T., Houstis, E. N., and Vavalis, M. (2020). ETH analysis and predictions utilizing deep learning. *Expert Systems with Applications*, 162:113866.