Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 5, No. 4, December, 2024

Cryptocurrency Forecasting Using Deep Learning Models: A Comparative Analysis

Rachid Bourday [1*], Issam Aatouchi [1], Mounir Ait Kerroum [1], Ali Zaaouat [1]

[1] *Computer Science Research Laboratory, Faculty of Sciences, IBN Tofail University, Kenitra, Morocco.*

Received 03 April 2024; Revised 04 November 2024; Accepted 09 November 2024; Published 01 December 2024

Abstract

Bitcoin has recently grown to prominence as a decentralized digital currency, attracting significant interest for its potential transformation of the financial market. Forecasting Bitcoin's price is crucial for investors, traders, and academics, given the currency's inherent volatility, which makes accurately predicting future prices challenging. This article aims to provide a comprehensive and comparative analysis of Deep Learning Forecasting Models in order to predict Bitcoin prices in the short and medium terms: Transformer with XGBoost, Transformer with ANN, Transformer with LSTM, and Transformer with SVR. This study is the first to explore the effectiveness of transformer-based architectures, particularly focusing on feature extraction, in complex financial market predictions. Therefore, we trained these models using historical Bitcoin data from 2016 to 2023 and evaluated their performance on a test dataset. Our experiments demonstrate that the Transformer with the XGBoost model outperforms the baseline models, achieving a Mean Absolute Error (MAE) of 0.011 and a Root Mean Squared Error (RMSE) of 0.018. Our findings suggest that the use of advanced deep learning techniques effectively manages the complexities of the cryptocurrency market, offering significant improvements over traditional methods and guiding investors in the cryptocurrency markets.

Keywords: Forecasting; Deep Learning; Machine learning; Transformer; LSTM; XGBoost.

1. Introduction

Bitcoin is one of the first decentralized digital currencies that both banks and individuals have not yet controlled. Since its creation in 2009, Bitcoin has rapidly gained popularity, especially in 2017, as it became a symbol of the potential social and economic revolution of the future currency model. Its price has experienced multiple fluctuations over the years, drawing keen interest from economic institutions in price prediction. This interest is vital for both current and potential investors, as well as for the government entities, emphasizing a high demand for effective Bitcoin price prediction mechanisms [1].

The ongoing digital transformation is causing profound disruptions across global economies and financial systems. The digitization and encryption processes are fundamentally altering the financial world. A recent report forecasts the digital economy to achieve an impressive 23 trillion dollars by 2025, accounting for about 25% of the global economy and encompassing a wide array of digital assets, both tangible and intangible [2].

Price volatility remains a significant concern in the realm of digital currencies like Bitcoin, which behaves differently compared to traditional stocks [3]. The evolution of Bitcoin price forecasting systems plays a pivotal role in aiding both

* Corresponding author: rachid.bourday@uit.ac.ma

http://dx.doi.org/10.28991/HIJ-2024-05-04-013

human and algorithmic traders to make informed decisions that could potentially enhance profitability in the volatile cryptocurrency market. Various techniques for forecasting Bitcoin's price can be categorized into two main groups: traditional statistical methods and machine learning methods [4].

Early research on cryptocurrency price forecasting often relied on traditional statistical techniques, such as ARIMA (Autoregressive Integrated Moving Average) [5] and GARCH [6, 7]. While these foundational methods primarily detect linear patterns in time series data. Additionally, they typically assume that variables follow a normal distribution—an assumption that is not valid for cryptocurrencies, given their highly volatile and non-normal distribution characteristics [8].

To overcome these limitations, machine learning techniques that can extract nonlinear patterns and efficiently handle large datasets without preconceived assumptions about the data structure have been adopted. Nevertheless, methods such as multilayer perceptron (MLP) neural networks [9] and support vector machines (SVM) [10] occasionally struggle with challenges like overfitting and may not completely capture the complex, hidden patterns within cryptocurrency data. To rectify these issues, numerous deep learning-based forecasting models have been employed to outperform traditional machine learning methods [11-13]. Deep learning has an inherent advantage in financial time series prediction since it does not rely on the assumption of stationarity. The architecture of deep neural networks enhances their strong generalization capabilities. Specifically, Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) Attention Mechanisms, Gradient-based Optimization Techniques, leverage their internal structures to extract temporal correlations from time series data, enabling them to effectively model and predict complex financial sequences.

Therefore, in this research, we present a new hybrid model that effectively combines the strengths of the transformer and the deep learning methods. We utilize the Transformer [14] for its advanced capabilities in feature extraction, especially for its efficiency in handling sequential data. Moreover, our approach includes a strategic training of the model and a careful fine-tuning of the hyperparameters to optimize the performance.

The major contributions in this paper are described as follows:

- ***Novel Methodology***: We introduce a unique hybrid deep learning model that integrates Transformer architectures with traditional machine learning methods such as XGBoost, ANN, LSTM, and SVR. This approach leverages the strengths of each model to enhance predictive accuracy in Bitcoin price forecasting.
- ***Comprehensive Analysis***: This study presents a complete comparative analysis of the proposed models using extensive historical data from 2016 to 2023. Our analysis not only demonstrates the superior performance of our models compared to existing baselines but also sheds light on their applicability to other cryptocurrencies.
- ***Advanced Feature Extraction*****:** We explore the effectiveness of Transformer-based models in extracting complex features from financial time series data. This is one of the first studies to demonstrate the capacity of Transformers to enhance feature extraction capabilities in the context of cryptocurrency market predictions.
- ***Practical Implications***: The findings of this study offer significant insights for investors, providing a robust tool for enhancing trading strategies in the volatile cryptocurrency market. Our results indicate that the use of advanced deep learning techniques can lead to better investment decisions and improved market analysis.

This paper is organized as follows: Section 2 presents related work in the field of cryptocurrency price predictions. Then, a description of the proposed methodology used in section 3. In section 4, the presentation of the experiments and simulation results is discussed. Finally, the article is concluded with a conclusion.

2. Related Works

The following section provides an overview of existing research related to forecasting cryptocurrency prices. It particularly focuses on the utilization of machine learning and deep learning models. The researchers have employed a variety of algorithms to predict cryptocurrency prices, including the LSTM model and other deep learning techniques. Recently, deep learning methods have become increasingly popular for time series prediction, especially in analyzing the volatility of cryptocurrency prices.

One of the approaches uses the LSTM model for predicting cryptocurrency prices. Boongasame & Songram [15] compares the effectiveness of LSTM with traditional linear connection approaches and technical analysis indicators like SMA (Smooth Moving Average), WMA (Weighted Moving Average), and EMA (Exponential Moving Averages). The primary evaluation metric used in this research was the Mean Absolute Percentage Error (MAPE). The findings indicated that the LSTM model, especially when configured with certain time parameter settings, was more effective and outperformed than the other models in accuracy.

Fleischer et al. [16] used the LSTM model to learn and predict future cryptocurrency prices based on historical closing prices. The model's performance was evaluated using the RMSE and was compared to the performance of an ARIMA model. Consequently, results showed that the LSTM-based model was particularly effective in capturing the inherent volatility of cryptocurrency prices and demonstrating more promising results than the ARIMA model [17].

In Tripathy et al.'s [18] study, deep learning models were used to predict Bitcoin prices. The research examined three algorithms: ARIMA, LSTM network, and FB-prophet. These models were evaluated based on their RMSE. Among these, FB-prophet was found to be the most efficient, surpassing both ARIMA and LSTM in terms of accuracy. This study highlights the potentiality of advanced deep learning techniques in financial forecasting, particularly in cryptocurrency price prediction.

On the other side, Wu et al. [19] introduced a novel method using Transformer-based machine learning models for time series forecasting. This approach was used to both univariate and multivariate time series data, including time series embeddings. Moreover, it demonstrated that the Transformer-based models yielded promising results, indicating their potential effectiveness in complex forecasting tasks.

McNally et al. [20] studied the performance of Bitcoin price prediction using LSTM, Recurrent Neural Network (RNN), and Autoregressive Integrated Moving Average (ARIMA) models. The results illustrated that LSTM has reached the highest classification accuracy of 52% and RMSE of 8%, while ARIMA has achieved 50% accuracy and RMSE of 53.74%, demonstrating the limitations of traditional parametric ARIMA models.

Wu et al. [21] developed a hybrid model that combined LSTM with the AR(2) model for predicting Bitcoin prices. This innovative approach demonstrated enhanced predictive capabilities when compared to traditional LSTM models alone. Wu's work represented a significant advancement in the field of cryptocurrency price forecasting, showcasing the potentiality of integrating different modeling techniques for improved accuracy in predictions.

Li & Dai [22] explored an innovative approach to forecast Bitcoin prices by leveraging a hybrid neural network model that combines Convolutional Neural Networks (CNN) and LSTM networks. The results demonstrated that the CNN-LSTM hybrid model outperforms the benchmark models in predicting short-term Bitcoin price fluctuations, which gave a MAE of 209.89, RMSE of 258.31, and MAPE of 2.35.

Ramos-Perez [23] developed a cryptocurrency prediction model leveraging the LSTM and GRU algorithms to forecast the prices of Bitcoin, Ethereum, and Litecoin. Using two types of data samples for each cryptocurrency, their model's performance was evaluated using RMSE and MAE.

Besides, Ramos-Perez et al. [23] developed a neural network architecture of multi-transformer, which was designed to forecast the volatility of the S&P index. This approach involved the adaptation of the traditional Transformer layers to be used in volatility forecasting models. The findings indicated that the Multi-Transformer and Transformer layer-based models were more accurate and provided more appropriate measures for forecasting compared to other algorithms, including those based on feed-forward layers or LSTM models. This highlights the effectiveness of Transformer-based architectures in complex financial market predictions.

Kanaparthi [24] investigated the robustness of LSTM-based RNN in predicting Bitcoin's daily closing prices using data disturbed by Gaussian noise. The results demonstrated that LSTM outperformed the traditional ARIMA(2,1,2) model in resilience to disturbances, even though prediction errors increased with higher noise levels.

Labbaf & Manthouri [25] introduced a novel method for predicting cryptocurrency time series, specifically Bitcoin, Ethereum, and Litecoin. They combined technical indicators with a Performer neural network and BiLSTM (Bidirectional Long Short-Term Memory) to capture effectively temporal dynamics and extract relevant features from cryptocurrency data. This method demonstrated superior performance compared to established models, signifying a substantial advancement in cryptocurrency price prediction.

Jin & Li [26] contributed a novel hybrid prediction model that integrates variational mode decomposition (VMD), GRU neural network, LSTM neural network, and attention mechanisms, significantly enhancing prediction accuracy. This research introduced new methodologies, including a residual re-decomposition prediction method that predicts and retains residuals, thereby improving the accuracy of the final prediction. The empirical results, obtained from daily Bitcoin and Ethereum data, achieved lower error metrics (RMSE, MAE, MAPE) than standalone LSTM and GRU models and outperforms other hybrid models.

Chen [27] focused on improving Bitcoin price prediction accuracy by employing machine learning techniques such as random forest regression and LSTM. The study also aimed to identify key variables that influence Bitcoin's price. It introduced a comprehensive analysis of 47 explanatory variables across eight categories, revealing that Bitcoin’s previous prices, US and Japanese stock market indexes, and the price of Ethereum are significant predictors. Results showed that while Random Forest Regression generally outperforms LSTM in terms of RMSE and MAPE. This study conducted an in-depth analysis using data from 2015 to 2022 showing that the model with only one lag of explanatory variables has the best prediction accuracy, supporting the efficient market hypothesis.

Ladhari & Boubaker [28] explored the effectiveness of hybrid deep learning models, particularly focusing two hybrid deep learning models for predicting Bitcoin prices using high-frequency data. The first model merges LSTM with Attention Mechanisms, and the second combines ANN-LSTM, both with Gradient-Specific Optimization to improve predictive performance. The study uses a dataset of over 50,000 hourly data points from May 2018 to January 2024. Results show that the LSTM-Attention model with Gradient-Specific Optimization achieved higher accuracy than the ANN-LSTM model.

In another research, Frohmann et al. [29] introduced a hybrid model using time series and sentiment analysis with a BERT model to predict Bitcoin prices. The empirical results showed a significant improvement in prediction accuracy, with the model achieving a MAE of 2.67 and a RMSE of 3.28. These metrics indicate a robust enhancement over traditional models. The integration of sentiment analysis, particularly analyzing the weight of sentiments based on the tweet creator's followers, effectively captures market sentiment, which is crucial for accurate cryptocurrency price prediction.

The recent studies have effectively demonstrated deep learning models, specifically Transformers and hybrid models, in forecasting cryptocurrency prices. These algorithms have successfully learned the complex patterns within time series data and overcoming traditional forecasting techniques. Table 1 shows the previous studies of various state of the art models.

Table 1. A relative comparison of state-of-the-art approaches for cryptocurrency price prediction

Reference	Approach	Expected Results
Boongasame & Songram (2023) [15]	LSTM, SMA, WMA and EMA.	LSTM showed a MAPE of 0.0927%.
Fleischer et al. (2022) [16]	LSTM, ARIMA	LSTM showed superior performance than ARIMA using RMSE.
Patel et al. (2020) [17]	LSTM with GRU	LSTM with GRU forecasts the prices with high accuracy compared to exiting models.
Nrusingha (2023) [18]	ARIMA, LSTM, FB-prophet	FB-prophet as the most efficient model.
Wu et al. (2020) [19]	Transformer	Showed promising results in handling complex prediction tasks.
McNally et al. (2018) [20]	LSTM, RNN, ARIMA	LSTM shows higher performance.
Wu (2018) [21]	LSTM with AR(2)	Improved accuracy over traditional LSTM.
Li & Dai (2020) [22]	CNN-LSTM hybrid	Outperformed in prediction accuracy, showcasing the strength of hybrid models.
Ramos-Perez et al. (2021) [23]	Multi-transformer	More accurate for forecasting cryptocurrency prices.
Kanaparthi (2024) [24]	LSTM	LSTM demonstrate robustness in predicting Bitcoin prices, outperforming traditional ARIMA models.
Labbaf & Manthouri (2024) [25]	Indicators-Performer-BiLSTM	The hybrid model boosts accuracy and efficiency, outperforming others in cryptocurrency prediction.
Jin & Li (2023) [26]	VMD-AGRU-RESVMD-LSTM	The hybrid model enhances prediction accuracy and investment strategy efficiency.
Chen (2023) [27]	RFR, LSTM	RFR outperforms LSTM in terms of RMSE and MAPE.
Ladhari & Boubaker (2024) [28]	LSTM-Attention with GSO	Hybrid models show high accuracy, promising for investment and trading strategies.
Frohmann et al. (2023) [29]	Time series with BERT	Hybrid models using linear regression achieve the best performance, MAE 2.67 and RMSE 3.28.

3. Research Methodology

3.1. Problem Statement

Time series forecasting has been considered a challenging issue that predicts future values based on the past. That is widely seen in many real-world applications addressed by researches, including finance, weather forecasting, and power generation. This section provides a formal definition of the problem Bitcoin price forecasting. The dataset consists of records characterized by some features describing the Bitcoin price (e.g., transactions, block size, difficulty see Section 3.2 for the full description of the datasets) and the associated priceUSD. It comes naturally to formalize it all as a regression problem: we want to predict priceUSD. For the mathematical notation, we denote vectors by lower case bold Roman letters like $\mathbf{x}$. While for matrices, we use upper case bold Roman letters like $\mathbf{X}$. A superscript T denotes the transpose of a vector or a matrix; therefore, $\mathbf{x}^T$ will be a row vector. The notation $(x_1, \ldots, x_d)$ denotes a row vector of dimension d. Whereas the corresponding column vector is written as $\mathbf{x} = (x_1, \ldots, x_d)^T$. Let N be the number of observation and M the number of features associated with each observation. Then, formally our dataset is a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$ in which the i^{th} row correspond to the i^{th} observation, i.e., the row vector $\mathbf{x}_i^T$:

$$\mathrm{X} = \begin{pmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_N^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1M} \\ x_{21} & \cdots & x_{2M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{pmatrix} \tag{1}$$

The target value $\mathbf{y} \in \mathbb{R}^N$ will be a vector $\mathbf{y} = (y_1, \ldots, y_N)$ containing priceUSD in our dataset. Particularly, the target price $\mathbf{y}_i$ will correspond to the set of features $\mathbf{x}_i = (x_{i1}, \ldots, x_{iM})$ of the i^{th} observation. The assumption upon which we will rely is that, denoting $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ our dataset, a specific unknown function f exists such that $f: \mathbb{R}^{N \times M} \to \mathbb{R}^N$ and

$f(\mathbf{X}) = \mathbf{y}$. The goal is to find $\hat{f}$, a useful approximation of f, through the minimization of a Loss Function. This measures the distance between $\hat{f}$ and f [30]. The most common loss that we will use in our experiments is the Squared Error Loss calculated in Equation 2:

$$L(\mathrm{y}, \hat{f}(\mathrm{X})) = (\mathrm{y} - \hat{f}(\mathrm{X}))^2 \tag{2}$$

Although the given dataset $\mathcal{D}$ is not deterministic, it comes from a particular distribution $p(\mathbf{x}, y)$, what we will need, here, is to minimize the expected value of $L(\mathrm{y})$. It can be proven that the solution to this problem is the conditional mean calculated by (2):

$$\hat{f}(x) = \int yp(y \mid x)dy = \mathbb{E}[y \mid x] \tag{3}$$

3.2. Models

This section aims to introduce the deep learning models that we will apply to our problem in the Experiments chapter. The Bitcoin prices are modeled by using different deep learning regression-based Transformer-Enhanced Hybrid Models. The models combine the Transformer architecture with various machine learning algorithms to enhance their predictive capabilities for Bitcoin price prediction. The following subsections provide explanations of each model.

- ***Transformer with Artificial Neural Network (ANN)***

The Transformer with ANN [31] model integrates the Transformer Encoder's ability to capture complex temporal dependencies and patterns in Bitcoin price data with the ANN's capability to process encoded features and make predictions. The Transformer Encoder employs self-attention mechanisms to obtain effective long-range dependencies and relationships within the time series data, enabling it to extract meaningful representations of the input. Once the Bitcoin price data is encoded by the Transformer Encoder, it is passed to the ANN component for further processing. The ANN consists multiple layers of interconnected neurons, allowing it to learn intricate non-linear relationships in the encoded features. Through forward propagation, the ANN learns to map the encoded features to the corresponding target Bitcoin prices.

The integration of the Transformer Encoder with an ANN enables the model to extract high-level features from the raw Bitcoin price data and refine them through its learning process. This combination leverages the strength of both architectures, enabling the model to effectively get complex patterns and fluctuations of Bitcoin prices.

- ***Transformer with Long Short-Term Memory (LSTM)***

This model combines the Transformer architecture and LSTM networks [32] to improve Bitcoin price predictions. The Transformer part of the model utilizes self-attention mechanisms to extract complex patterns from the price data, highlighting significant features and temporal relationships, while the LSTM component captures long-term dependencies, allowing the model to learn from historical trends. This hybrid approach enhances the model's predictive accuracy by effectively handling the intricacies and volatility typical of financial time series.

Linking the strengths of both architectures, the Transformer with LSTM model can capture effectively both short-term fluctuations and long-term trends in Bitcoin prices, leading to more accurate predictions.

- ***Transformer with Support Vector Regression (SVR)***

This model merges the Transformer (Encoder) architecture with Support Vector Regression (SVR) [33] to augment the predictive capabilities for Bitcoin price prediction. The Transformer renowned its adeptness in capturing intricate temporal dependencies and patterns in Bitcoin price data through self-attention mechanisms, collaborates with SVR, a robust regression technique known for its effectiveness in capturing non-linear relationships and handling high-dimensional feature spaces.

In this hybrid model, the Bitcoin price data undergoes encoding by the Transformer to extract meaningful representations of the input features. Subsequently, the encoded features are utilized as input to the SVR algorithm, which performs regression analysis to predict future Bitcoin prices based on the extracted features and historical data patterns.

- ***Transformer with Extreme Gradient Boosting (XGBOOST)***

In the following hybrid model, the Bitcoin price data is first encoded by the Transformer Encoder to extract meaningful representations of the input features. Alternatively, the encoded features are utilized as input to the XGBoost algorithm [34], which performs ensemble learning to predict future Bitcoin prices based on the extracted features and historical data patterns. The integration of the Transformer with XGBoost enables the model to effectively capture both linear and non-linear relationships in Bitcoin price data, thereby improving its predictive accuracy.

4. Experiments

This section describes the methodology and various phases of the study; and it is structured to cover all aspects of the experimental process:

4.1. Data Collection

In this study, the data was collected from https://bitinfocharts.com, using a web scraper written in Python. Particularly, Bitcoin's prices were used from 01 January 2016 to 31 December 2023 (for a total of 2922 days), which are presented in Figure 1. During this period, the lowest price was 371.323 USD in 2016-01-16 and the highest price was 67547 USD in 2021-11-09, which is as many times higher than the lowest price. Among various features, we considered the 20 features of the Bitcoin blockchain, listed in Table 2.

Figure 1. Bitcoin's prices from 01 January 2016 to 31 December 2023

Table 2. Bitcoin blockchain features

Feature	Description
Transaction	Refers to the number of transactions completed within a certain period.
Bock Size	The size of each block in the blockchain, measured in bytes or kilobytes.
Sent-addresses	The number of unique addresses that have sent transactions over the network during a given period.
Difficulty	A relative measure of difficulty in finding a new block.
Hashrate	The estimated number of tera hashes per second the Bitcoin network is performing.
MiningProfitability	An indicator of how profitable it is to mine the cryptocurrency.
Send_usd	The total value of transactions sent in USD within a certain period.
av_trs_size	The average size of a transaction on the network.
median_trs_size	Like the average transaction size but uses the median value.
confirmation_time	The average time it takes for a transaction to be confirmed on the network.
market_cap	The total USD value of Bitcoin supply in circulation.
av_trs_value	The average monetary value of a transaction on the network.
median_trs_value	The median value of transactions, providing an alternative to the average that's less skewed by extreme values.
tweets	The number of Twitter posts related to the Bitcoin.
google_trends	A metric indicating the search interest for the cryptocurrency on Google.
active_addresses	The number of unique addresses that have been active in the network over a specific period.
top_100_percent	The percentage of Bitcoin's total supply held by the top 100 addresses.
fee_reward	The fees paid to miners for processing transactions.

4.2. Pre-Processing

Preparing the Bitcoin data for the proposed models, the dataset must be restructured and organized in a way that facilitates the learning process. The data preprocessing involves the following steps:

- ***Importing the Data:*** The Bitcoin_2016_2023.csv file contains historical Bitcoin prices which can read into a pandas DataFrame and converting the 'Date' column to datetime format.
- ***Missing Values:*** The dataset contains some null values; these null values are filled by using interpolation method. After this process, the final prepared dataset is in the form of time series.
- ***Feature Selection:*** It is an essential part of data pre-processing, which is necessary to improve model performance. In our study, a correlation analysis between the rest of the features and the Bitcoin price was conducted. Here, we calculate the Variance Inflation Factor (VIF) [35] as shown Table 3. Through employing VIF, we are able to quantify the extent of multicollinearity and selectively remove features that contribute to high levels of collinearity. Then, we excluded the features with a VIF great than 5, such as Hashrate and Difficulty.
- ***Data Splitting:*** The dataset is split into a training set (90% of the data) and a validation set (10% of the data) for the purpose of training and validating of a deep learning model.
- ***Data Transformation:*** We apply data transformation strategies, such as normalization and standardization, to modify and scale the data. This step is important to align features on a common scale and for addressing issues of skewness [36], aiding in the affective application of deep learning algorithms.

The workflow of price forecasting involves several stages, as illustrated in Figure 2.

Table 3. Bitcoin Blockchain Features VIF

Feature	VIF	Feature	VIF
hashrate	105.471439	mining_profitability	1.492738
difficulty	69.886526	send_usd	1.445723
av_transaction_value	4.765635	median_transaction_value	1.441975
active_addresses	3.706024	av_transaction_size	1.298843
sent_addresses	3.693921	median_transaction_size	1.198973
transactions	2.381497	fee_reward	1.197281
market_cap	2.365479	block_size	1.107445
tweets	2.129348	top_100_percent	1.02119
google_trends	1.980745	confirmation_time	0.445589

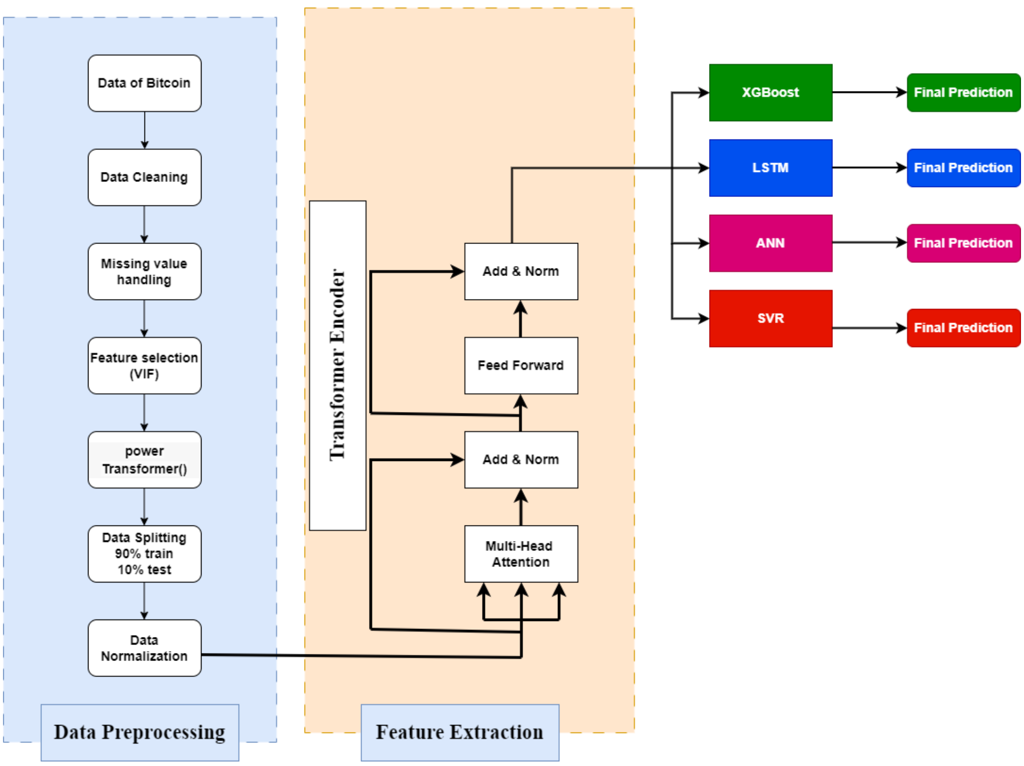

Figure 2. Complete workflow diagram

4.3. The Proposed Models

Figure 3 shows the proposed model. Our research actively explores a comparative analysis of four distinct hybrid models in predicting Bitcoin prices. Each model uniquely merges a Transformer encoder for extracting features with a varied predictive algorithm: XGBoost, LSTM, SVR, and ANN. The purpose is to investigate the impact of the Transformer's capability in recognizing intricate, long-range patterns within time-series data when combined with diverse predictive methods on the precision and effectiveness of Bitcoin price predictions. In this field, as well seeks to pinpoint the most efficient hybrid model for forecasting financial time series and to enrich the understanding of the advanced deep learning applications.

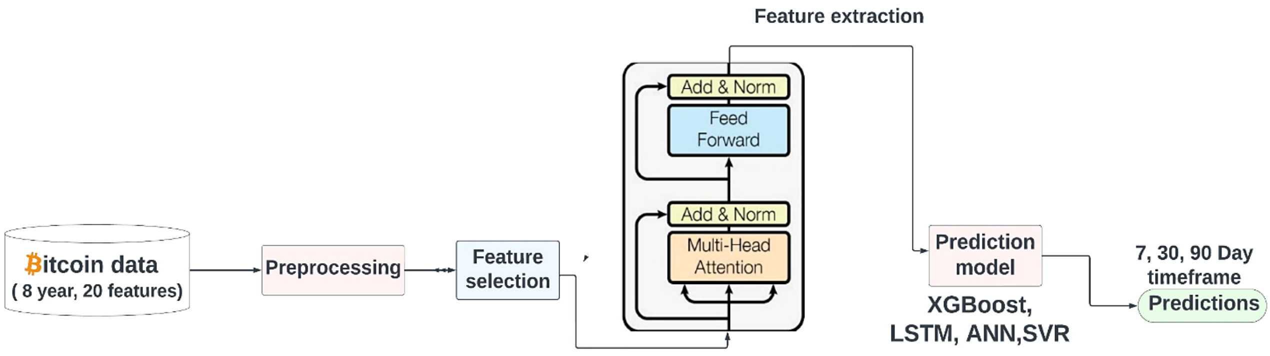

Figure 3. The proposed Model

4.4. Performance Evaluation

Evaluating the performance of the regression models, the following metrics are used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and their Mean Square Error (MSE).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{Predicted} - \text{Actual})^2}{N}} \tag{4}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (\text{Predicted} - \text{Actual})^2 \tag{5}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} (\text{Predicted} - \text{Actual}) \tag{6}$$

5. Results and Discussion

The results in Table 4 compare the performance of four Transformer-based models in a forecasting task, using MAE and RMSE as metrics. The Transformer combined with Xgboost model presents the efficient performance across all metrics, indicating high accuracy and reliability. In contrast, the Transformer with LSTM indicates the least accuracy, having the highest values in both metrics. The Transformer with SVR performs better than the ANN and LSTM models, but not as well as XGboost. This suggests that while each model has its strengths, the Transformer with Xgboost might be the most effective for this specific task. A comparison between our models and existing systems presented in Table 5, proves that the proposed system has achieved much fewer prediction errors.

Table 4. Regression results for proposed approaches and compared models

Metrics	Transformer with ANN	Transformer with SVR	Transformer with LSTM	Transformer with XGBoost
MAE	0.024	0.021	0.067	0.011
RMSE	0.034	0.028	0.10	0.018

Table 6 summarizes the forecast of the regression models for n^{th}-day BTC price. The bar chart in Figure 4 demonstrates the performance of the deep learning models in terms of MSE, MAE and RMSE.

Table 5. Comparison results between the proposed system and the existing models

Reference	Existing Models	Study Currency	Obtaining Results
Kanaparthi (2024) [24]	ARIMA, LSTM	BTC	**ARIMA**: RMSE: 0.1692 **LSTM**: RMSE: 0.1179
Labbaf & Manthouri (2024) [25]	Indicators-Performer-BiLSTM.	BTC, ETH, LTC	**Bitcoin (BTCUSD):** Hourly RMSE: 243, Daily RMSE: 1316 **Ethereum (ETHUSD):** Hourly RMSE: 18.3, Daily RMSE: 100 **Litecoin (LTCUSD):** Hourly RMSE: 1.92, Daily RMSE: 7.5
Jin & Li (2023) [26]	VMD-AGRU-RESVMD-LSTM	BTC, ETH	**Bitcoin (BTC):** RMSE:50.651, MAE: 42.298 **Ethereum (ETH):** RMSE:2.873, MAE: 2.410
Chen (2023) [27]	RFR, LSTM	BTC	**RFR (Period 1):** RMSE: 321.61, MAPE: 3.39% **RFR (Period 2):** RMSE: 2096.24, MAPE: 3.29%
Ladhari & Boubaker (2024) [28]	ANN-LSTM, LSTM-Attention Model	BTC	**ANN-LSTM Model**: RMSE:4,926,484, MAE: 1,451,432, MAPE: 0.086 **LSTM-Attention Model**: RMSE: 1,465,833.911, MAE: 816,256, MAPE: 0.048
Frohmann et al. (2023) [29]	Hybrid approach using sentiment analysis and forecasting models.	BTC	**Linear Regression (LR):** RMSE: 3.28, MAE: 2.67 **LSTM:** RMSE: 35.41, MAE: 35.04 **TCN:** RMSE: 22.18, MAE: 22.72
Our proposed systems	Proposed system (Transformer with ANN)	BTC	MAE: 0.024 RMSE: 0.034
	Proposed system (Transformer with SVR)	BTC	MAE: 0.021 RMSE: 0.028
	Proposed system (Transformer with LSTM)	BTC	MAE: 0.067 RMSE: 0.10
	Proposed system (Transformer With Xgboost)	BTC	MAE: 0.011 RMSE: 0.018

Table 6. Regression result for proposed approaches and compared models for nth day

Metrics	Horizon	Transformer with ANN	Transformer with SVR	Transformer with LSTM	Transformer with Xgboost
RMSE	7	0.044	0.053	0.175	0.053
	30	0.045	0.038	0.139	0.029
	90	0.035	0.032	0.107	0.023
MAE	7	0.035	0.040	0.145	0.026
	30	0.034	0.028	0.107	0.015
	90	0.025	0.022	0.073	0.013
MSE	7	0.001	0.002	0.030	0.0002
	30	0.002	0.001	0.019	0.0008
	90	0.001	0.001	0.011	0.0005

For the 7th-day Bitcoin price forecast, the Transformer with XGBoost model shows remarkable accuracy, having the lowest RMSE, MAE, and MSE among the compared models. This suggests that our proposed system is highly effective for short-term forecasting. The Transformer with SVR also performs well, better than the ANN and LSTM models, but not as efficiently as XGBoost. The higher error rates in ANN and LSTM indicate that they might be less reliable for this specific 7-day forecasting scenario compared to XGBoost and SVR models.

In the 30th day of Bitcoin price forecast, the Transformer with XGBoost model significantly outshines others, maintaining the lowest error rates across RMSE, MAE, and MSE, indicating a strong capability in medium-term forecasting. The SVR model, ranking second, demonstrates better accuracy than the ANN and LSTM models but doesn't match the precision of XGBoost. The higher error rates in ANN and LSTM models suggest that they are less effective for 30-day forecasting. This trend highlights the XGBoost model's robustness and adaptability to different forecasting durations. However, this should be evaluated considering the fluctuations of the models as shown in Figure 5 where the Transformer with XGBoost clearly outperforms all other models.

Lastly, in the 90th day forecast horizon, the Transformer with XGBoost model continues to demonstrate superior performance, maintaining the lowest error rates in RMSE, MAE, and MSE as presented in Figure 5. This consistency highlights its strong predictive capability even in long-term forecasting. The Transformer with SVR also represents good accuracy, better than the ANN and LSTM models. However, the increased accuracy of ANN and LSTM models for the 90-day horizon compared to shorter horizons suggests that they might be more suited for longer-term predictions, despite still being outperformed by the XGBoost and SVR models.

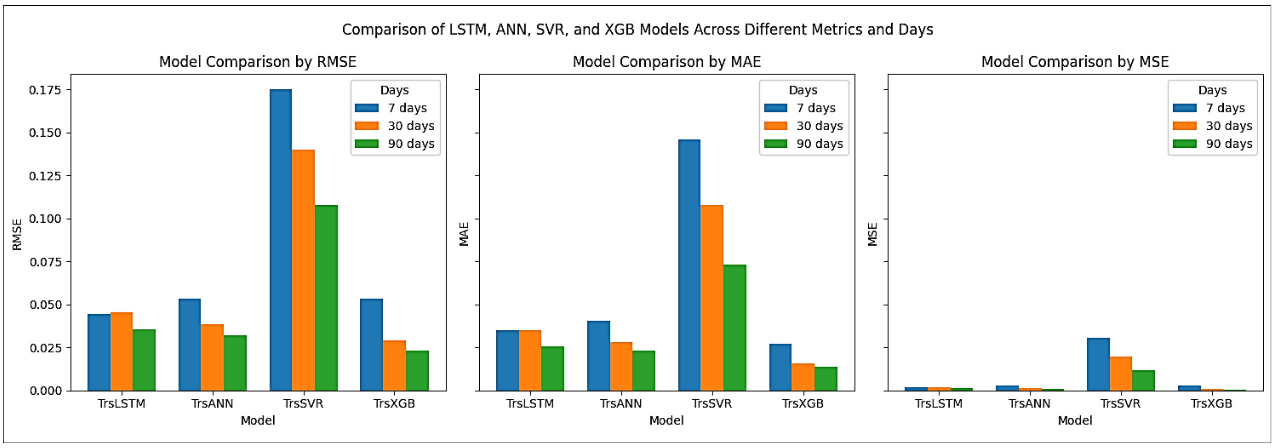

Figure 4. Comparaison of hybrid models across different metric in nth-day

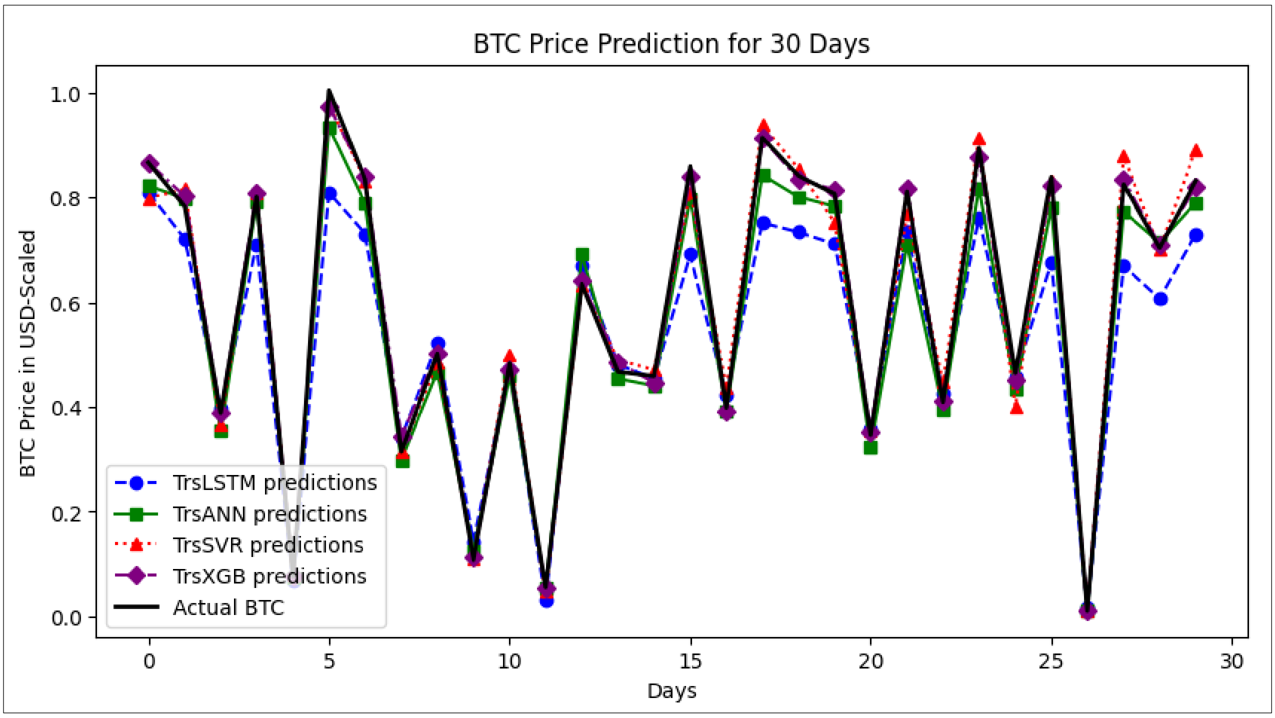

Figure 5. Performance of hybrid deep learning models based on 30-days horizon of forecasting

6. Conclusion

The main study of this work is to analyze Bitcoin's price for time series regression. This research applied to various machine learning models, deep learning, and a combination of these two models to forecast the Bitcoin price in short-term to mid-term. So, we developed and compared various deep learning-based Bitcoin price prediction models using Bitcoin blockchain information and the experimental results, which show that the Transformer with XGBoost outperformed the other models. However, our current models exhibit potential significance in predicting Bitcoin's price but also have notable limitations, particularly in integrating sentiment analysis from dynamic sources such as social media. These models do not currently assess the intensity of market sentiment reflected in text-based data, which is crucial for understanding the volatile cryptocurrency market. This gap presents ongoing opportunities for enhancing forecast accuracy, reducing prediction errors, and improving robustness against the frequently changing market conditions.

Moving forward, our focus will be on expanding the model framework by incorporating a broader array of predictive models, fine-tuning hyperparameters, and enhancing the capabilities of existing hybrid models. The objective is to refine these forecasting tools to ensure they are more reliable and can adapt to market fluctuations effectively. By achieving these improvements, we aim to increase forecast accuracy, deliver trustworthy predictive outcomes, and create models that are responsive to market dynamics, thereby providing valuable insights for market participants. Additionally, only a few critical feature selection methods have been applied to the dataset. Many other feature selection techniques can be explored to improve the model. Future research could forecast other digital currencies, including Ethereum and Ripple. Moreover, it would be important to consider other recent deep learning models, such as Graph Neural Networks (GNNs) [37], to build prediction models.

7. Declarations

7.1. Author Contributions

Conceptualization, R.B. and I.A.; methodology, R.B., I.A., and M.A.; software, R.B. and A.Z.; validation, I.A. and M.A.; formal analysis, R.B. and A.Z.; investigation, R.B. and A.Z.; resources, R.B., I.A., M.A., and A.Z.; data curation, R.B.; writing—original draft preparation, R.B.; writing—review and editing, R.B., I.A., M.A., and A.Z.; visualization, R.B. and I.A.; supervision, I.A. and M.A.; project administration, R.B. All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

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

7.3. Funding

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

7.4. Institutional Review Board Statement

Not applicable.

7.5. Informed Consent Statement

Not applicable.

7.6. Declaration of Competing Interest

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

8. References

[1] Kaushal, R. (2016). Bitcoin: First Decentralized Payment System. International Journal of Engineering and Computer Science, 5(5), 16514–16517. doi:10.18535/ijecs/v5i5.30.

[2] Oliver, J. (2013). Mastering blockchain distributed ledger technology, and smart contracts explained. Packt Publishing, Birmingham, United Kingdom.

[3] Doumenis, Y., Izadi, J., Dhamdhere, P., Katsikas, E., & Koufopoulos, D. (2021). A critical analysis of volatility surprise in bitcoin cryptocurrency and other financial assets. Risks, 9(11), 207. doi:10.3390/risks9110207.

[4] Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting, 37(1), 28–43. doi:10.1016/j.ijforecast.2020.02.008.

[5] Ibrahim, A., Kashef, R., & Corrigan, L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. Computers and Electrical Engineering, 89, 106905. doi:10.1016/j.compeleceng.2020.106905.

[6] Baur, D. G., Dimpfl, T., & Kuck, K. (2018). Bitcoin, gold and the US dollar – A replication and extension. Finance Research Letters, 25, 103–110. doi:10.1016/j.frl.2017.10.012.

[7] Fakhfekh, M., & Jeribi, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models. Research in International Business and Finance, 51. doi:10.1016/j.ribaf.2019.101075.

[8] Khedr, A. M., Arif, I., Pravija Raj, P. V., El-Bannany, M., Alhashmi, S. M., & Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intelligent Systems in Accounting, Finance and Management, 28(1), 3–34. doi:10.1002/isaf.1488.

[9] Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility-forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. Expert Systems with Applications, 109, 1–11. doi:10.1016/j.eswa.2018.05.011.

[10] Hajek, P., Hikkerova, L., & Sahut, J. M. (2023). How well do investor sentiment and ensemble learning predict Bitcoin prices? Research in International Business and Finance, 64, 101836. doi:10.1016/j.ribaf.2022.101836.

[11] Cui, T., Ding, S., Jin, H., & Zhang, Y. (2023). Portfolio constructions in cryptocurrency market: A CVaR-based deep reinforcement learning approach. Economic Modelling, 119. doi:10.1016/j.econmod.2022.106078.

[12] Liu, M., Li, G., Li, J., Zhu, X., & Yao, Y. (2021). Forecasting the price of Bitcoin using deep learning. Finance Research Letters, 40, 101755. doi:10.1016/j.frl.2020.101755.

[13] Ortu, M., Uras, N., Conversano, C., Bartolucci, S., & Destefanis, G. (2022). On technical trading and social media indicators for cryptocurrency price classification through deep learning. Expert Systems with Applications, 198, 116804. doi:10.1016/j.eswa.2022.116804.

[14] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, California, United States. doi:10.48550/arXiv.1706.03762

[15] Boongasame, L., & Songram, P. (2023). Cryptocurrency price forecasting method using long short-term memory with time-varying parameters. Indonesian Journal of Electrical Engineering and Computer Science, 30(1), 435–443. doi:10.11591/ijeecs.v30.i1.pp435-443.

[16] Fleischer, J. P., von Laszewski, G., Theran, C., & Bautista, Y. J. P. (2022). Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory. Algorithms, 15(7), 230. doi:10.3390/a15070230.

[17] Patel, M. M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions. Journal of Information Security and Applications, 55(102583). doi:10.1016/j.jisa.2020.102583.

[18] Tripathy, N., Hota, S., & Mishra, D. (2023). Performance analysis of bitcoin forecasting using deep learning techniques. Indonesian Journal of Electrical Engineering and Computer Science, 31(3), 1515-1522.

[19] Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317.

[20] McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018, 339–343. doi:10.1109/PDP2018.2018.00060.

[21] Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018). A new forecasting framework for bitcoin price with LSTM. In 2018 IEEE international conference on data mining workshops (ICDMW), 168-175. doi:10.1109/ICDMW.2018.00032.

[22] Li, Y., & Dai, W. (2020). Bitcoin price forecasting method based on CNN- LSTM hybrid neural network model. The Journal of Engineering, 2020(13), 344–347. doi:10.1049/joe.2019.1203.

[23] Ramos-Pérez, E., Alonso-González, P. J., & Núñez-Velázquez, J. J. (2021). Multi-transformer: A new neural network-based architecture for forecasting s&p volatility. Mathematics, 9(15), 1794. doi:10.3390/math9151794.

[24] Kanaparthi, V. (2024). Robustness Evaluation of LSTM-based Deep Learning Models for Bitcoin Price Prediction in the Presence of Random Disturbances. International Journal of Innovative Science and Modern Engineering, 12(2), 14–23. doi:10.35940/ijisme.b1313.12020224.

[25] Khaniki, M. A. L., & Manthouri, M. (2024). Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators. arXiv preprint, arXiv:2403.03606. doi:10.48550/arXiv.2403.03606.

[26] Jin, C., & Li, Y. (2023). Cryptocurrency Price Prediction Using Frequency Decomposition and Deep Learning. Fractal and Fractional, 7(10), 708. doi:10.3390/fractalfract7100708.

[27] Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. Journal of Risk and Financial Management, 16(1), 51. doi:10.3390/jrfm16010051.

[28] Ladhari, A., & Boubaker, H. (2024). Deep Learning Models for Bitcoin Prediction Using Hybrid Approaches with Gradient-Specific Optimization. Forecasting, 6(2), 279–295. doi:10.3390/forecast6020016.

[29] Frohmann, M., Karner, M., Khudoyan, S., Wagner, R., & Schedl, M. (2023). Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting. Big Data and Cognitive Computing, 7(3), 137. doi:10.3390/bdcc7030137.

[30] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media, New York, United States. doi:10.1007/978-0-387-21606-5.

[31] Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. Neural Networks, 61, 85–117. doi:10.1016/j.neunet.2014.09.003.

[32] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[33] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. Advances in Neural Information Processing Systems, 9.

[34] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. doi:10.1145/2939672.2939785.

[35] O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673–690. doi:10.1007/s11135-006-9018-6.

[36] Staniswalis, J. G., Severini, T. A., & Moschopoulos, P. G. (1993). On a data based power transformation for reducing skewness. Journal of Statistical Computation and Simulation, 46(1–2), 91–100. doi:10.1080/00949659308811495.

[37] Han, X., Huang, Y., Pan, Z., Li, W., Hu, Y., & Lin, G. (2023). Multi-Task Time Series Forecasting Based on Graph Neural Networks. Entropy, 25(8), 1136. doi:10.3390/e25081136.