Cryptocurrency price and volatility predictions with machine learning

Article *in* Journal of Marketing Analytics · August 2023

DOI: 10.1057/s41270-023-00239-1

Samir Poudel
Middle Tennessee State University

Shital Pandey
Embry-Riddle Aeronautical University

Khem Poudel
Middle Tennessee State University

Journal of Marketing Analytics
https://doi.org/10.1057/s41270-023-00239-1

ORIGINAL ARTICLE

Cryptocurrency price and volatility predictions with machine learning

Samir Poudel[1] · Rajendra Paudyal[2] · Burak Cankaya[3] · Naomi Sterlingsdottir[4] · Marissa Murphy[4] · Shital Pandey[4] · Jorge Vargas[5] · Khem Poudel[1]

Revised: 6 February 2023 / Accepted: 10 June 2023

Abstract

In recent years, the digital currency has gained significant popularity owing to its increasing dependence on computers and the Internet. Among various forms of virtual currency, cryptocurrency has emerged as a prominent contender. The advent of digital currency has opened new avenues in the software industry, particularly in finance, data storage, and data collection. This evolution has given rise to exciting opportunities for businesses to explore the potential of digital currency and leverage its benefits. Cryptocurrency (crypto) is very volatile regarding the market value, which carries a host of unknowns that make it difficult to predict and analyze future prices. This paper discusses the use of six types of machine-learning models (Linear Regression, LSTM, Bi-LSTM, GRU, TARCH, and VAR) to predict the Bitcoin and DogeCoin prices; General Least-Squares Regression and Neural Networks algorithms to predict the volatility of a given cryptocurrency and its prices from 2014 to 2023 with daily cryptocurrency volatility data. The results show that high-performance computing techniques such as GRU (Gated Recurrent Unit) neural networks (0.0468 RMSPE) regression models to predict relatively accurate crypto price volatility and past available cryptocurrency price data are proven to be used to verify the prediction results.

Keywords Cryptocurrency · Neural Networks · Price prediction · High performance computing · Machine learning · Linear regression

Introduction

Cryptocurrency is a body of binary data that stores information on transactions agreed upon by everyone participating in the trades in ownership records stored in a secure

Marissa Murphy, Aaron Morgado, Mason Gawler, Min Sun Kim, Shashwat Acharya, Shital Pandey, Rajendra Paudyal, Burak Cankaya and Khem Poudel have contributed equally to this work.

✉ Burak Cankaya
bcankaya@erau.edu

Samir Poudel
sp2ai@mtmail.mtsu.edu

Rajendra Paudyal
rpaudyal@gmu.edu

Naomi Sterlingsdottir
naomi.rz.sg25@gmail.com

Marissa Murphy
marsmurph88@gmail.com

Shital Pandey
PANDEYS2@my.erau.edu

Jorge Vargas
jorge.vargas@mtsu.edu

Khem Poudel
khem.poudel@mtsu.edu

[1] Department of Computer Science, Middle Tennessee State University, TN, Murfreesboro, United States

[2] Department of Computer Science, George Mason University, University Drive, Fairfax, VA 22030, USA

[3] Department of Management and Technology, Embry-Riddle Aeronautical University, 1 Aerospace Blvd, Daytona Beach, FL 32114, USA

[4] Department of Mathematics and Data Science, Embry-Riddle Aeronautical University, 1 Aerospace Blvd, Daytona Beach, FL 32114, USA

[5] Department of Computer Science, Middle Tenneesse State University, 1301 E Main Street, Murfressboro, TA 37132, USA

Published online: 27 August 2023

ledger database called a blockchain (Yaga et al. 2019). Blockchains enable various business solutions by offering them a decentralized transaction and banking experience, which is a need for new-generation organizations and people (Kaabachi et al. 2022). Making investments and spending is highly related to the culture, and we know that starting from Millenials, the investment culture shifts to less risk-averse markets (Singh and Kumar 2021). Due to cryptocurrency being volatile (Hayes 2023), the price of the specific cryptocurrencies fluctuates wildly at any given moment independently from each other, making it difficult to predict its future prices based on its current prices. However, since particular cryptocurrencies are stable, they behave similarly to stocks within the stock market and demand forecasting in the retail industry, possibly allowing linear regression models and non-linear forecasting models to predict future prices (de Almeida and da Veiga 2022). Yet cryptocurrency behaves similarly to stocks, allowing linear regression models and Machine-Learning Models to make predictions about price levels (Pandita 2021). With the ability to predict crypto prices, one can make a prediction for stocks since the popular coin, Bitcoin, affects stock prices. This study utilizes Bitcoin and DogeCoin data to develop predictive models for future cryptocurrency prices by analyzing their volatility and incorporating parameters from past cryptocurrency prices and the level of discussion surrounding their respective cryptocurrencies. By utilizing these parameters, this study aims to provide valuable insights into the future price moves of Bitcoin and DogeCoin, thereby enabling investors to make well-informed decisions when investing in digital assets.

These popular cryptocurrencies have a long history, allowing for different machine-learning models to find the best prediction model for the data being used. Each cryptocurrency will yield different values based on specific parameters, ideally showing strong possibilities of its future value. Bitcoin has the most extended history, making it the most extensive dataset, meaning high-performance computing is necessary due to its extreme volatility over the past decade. Other parameters, such as celebrities, politics, and the global market, vastly influence its volatility, undermine linear regression analysis, and require more complex algorithms. Accurate price predictions require the analysis of price history, encompassing significant historical events, which can be utilized as the decision parameters to train regression models for predicting future cryptocurrency prices. By analyzing the historical price movements of digital assets, including significant events that have impacted their values, predictive models can be developed that enable investors to anticipate future price trends and optimize their investment strategies accordingly. This time-series analysis approach is precious in the highly volatile cryptocurrency market, where even small changes in market sentiment and investor behavior can significantly impact asset prices, whereas DogeCoin, a 7-year-old "meme coin," is a perfect case of how the volatility of a cryptocurrency can substantially influence its value, making it difficult to detect its future price predictions. The hype about DogeCoin makes it one of the most famous cryptocurrencies with its high volatility, which makes DogeCoin have high positive and negative investor risks. With that in mind, creating and running a complex prediction requires parallel computing to observe a large data span and accurately predict.

The remaining portions of this study are divided into five divisions. The research literature reviewed in "Literature review" section is pertinent, emphasizing studies that use analytical models to predict cryptocurrency price volatility. The methods, theoretical discussions, and dataset used for this research are described in "Theoretical discussions, methodology, and data description" section. The results of the suggested framework are presented in "Results" section, along with a summary of the findings. "Discussion and future studies" section offers suggestions for cryptocurrency price prediction researchers and financial professionals. "Conclusion" section concludes by outlining the study's shortcomings and many potential directions for further investigation.

Literature review

The relative performance of various algorithms, such as generalized autoregressive conditional heteroskedasticity (GARCH), is known to be efficient on regressive prediction tasks and can be used to estimate the volatility of cryptocurrencies and global currencies. GARCH-type specifications majorly the standard GARCH (S-GARCH), integrated GARCH (I-GARCH), exponential GARCH (E-GARCH), Glosten-Jagannathan-Runkle (GJR) GARCH (GJR-GARCH), asymmetric Power Autoregressive Conditional Heteroskedastic PARCH (A-PARCH), threshold GARCH (T-GARCH), and component GARCH (C-GARCH) and forecasting performance were performed during the sample period of 2015 to 2019 and evidenced the superiority of I-GARCH model both in-sample and out-sample. In that study, forecasting the volatility of world currencies, C-GARCH modeled the cambio price of the Euro almost perfectly during periods (Naimy et al. 2021). Cryptocurrency is a non-centralized digital currency, so it is important to understand the relationships among cryptocurrencies for policymakers and investors. Multiple researchers examined the volatility of nine popular cryptocurrencies based on market capitalization using a Bayesian Stochastic Volatility and several GARCH models. The stochastic volatility model performs better than the GARCH model for cryptocurrencies (Kim et al. 2021).

Mining difficulty has a significant impact on Bitcoin research finding from the daily time-series dataset from the year 2011 to 2018. Autoregressive-distributed Lag (ARDL) model is used and transaction volume, number of stock, currency exchange rate, and financial development are not the price indicator of Bitcoin (Guizani and Nafti 2019). Model VAR-DCC-GARCH was modeled to examine return and volatility transmission among three different cryptocurrencies. Bitcoin, Ethereum, and Litecoin during pre-Covid-19 and Covid-19 periods. Volatility transmission is not significant among cryptocurrencies during the Covid period. Findings show that during pre-Covid-19, the return spillovers were unidirectional from Ethereum to Bitcoin and from Bitcoin to Litecoin, whereas they were bi-directional between Ethereum and Litecoin. One cryptocurrency's returns can be used to forecast another cryptocurrency's returns during pre-covid. However, research findings suggested that during the covid-19, return spillovers were insignificant between Bitcoin and Ethereum or between Bitcoin and Litecoin. Return spillovers vary across both periods for all three different types of cryptocurrency pairs (Yousaf and Ali 2020). New uncertainty index "UCRY policy" and price uncertainty "UCRY price" constructed around major cryptocurrencies. Price and policy uncertainty in the cryptocurrency market based on the Lexis Nexis database, Research finding provided the decomposition of the UCRY Index with significant events from the year of 2014 to 2020 including Covid-19 crisis, cyberattacks on cryptocurrency, exchanges, and political elections. Such uncertainty does not affect the volatility of the cryptocurrency (Lucey et al. 2022).

Based on stock market predictions, data mining and high-performance computing techniques are common to derive fast and precise results. In one study in 2014, models were trained using data from past market performance for stocks (Ghildiyal 2014). Similarly, in a 2015 study, artificial neural networks paired with metaheuristics were analyzed as an approach to predict stock prices (Olah 2015). Metaheuristic approaches can range from a simple local search to a more complex machine-learning algorithm to find data. Such techniques are helpful to understand the prediction and causal inference in contexts where information is incomplete or flawed (Moriyama and Kuwano 2022). Artificial neural networks have been utilized to predict stock price changes over time. Consequently, results are usually not deterministic. However, parallel metaheuristics can be used to run multiple searches simultaneously, thus, improving the overall solution by bringing in more helpful information. Because cryptocurrency behaves roughly in the same parallel computing manner, other machine-learning methods can be used, such as General Least-Square Regression, LSTM, Bi-LSTM, GRU, TARCH, and VAR algorithms.

As newer machine-learning methods are being created and used, Neural Networks (NN) are a single-layered subset of deep learning that simulates neurons in the brain. This allows the computer to recognize patterns and solve complex problems that traditional machine learning cannot due to processing power and time. Five-year study since the inception of Bitcoin using Binomial logistic regressions (BLR) and random forest(RF) shows that Prices and different blockchain features with 10 min of data give a better sensitivity and specificity ratio than the ten seconds of data (Madan and Saluja 2015). Automated trading on the Bitcoin was done by comparing simple technical analysis methods with more complex machine-learning models, exponential moving average (EMA), support vector machine (SVM), and volume-weighted SVM (VW-SVM) (Zbikowski 2016). NNs have a simple architecture that includes (1) an input layer, (2) at least one hidden layer, and (3) an output layer (Cankaya 2023). In this paper, the specific NNs that will be used are Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bi-directional RNN (Chatterjee 2019). LSTM utilizes a cell, input gate, output gate, and a forget gate to facilitate data going in and out of the LSTM cell and is efficient on predictive gradient models (Alaparthi and Mishra 2021). The GRU model is also a Recurrent NN that is similar to LSTM with a few key differences. The GRU model uses a reset and update gate instead of an input, output, and forget gate. The reset gate combines input data with previous memory and utilizes the update gate to determine how much previous memory to keep for the next task (Dey et al. 2017).

LSTM neural network differs from traditional neural networks because it allows information to persist (Olah 2015). Because of this recurrent neural network's effectiveness, they are essential for capturing long-term temporal dependencies. However, one drawback to this neural network is that this recurrence has been chastised for containing a large number of components whose functions are not as clear (Yildirim 2020).

Schuster and Paliwal (1997) developed bi-LSTM to train a network using past and future input data sequences. Two linked layers are used to process the input data. Bi-directional LSTM uses a finite sequence to forecast or tag the sequence of each element depending on the context of components in the past and future. To produce this result, two LSTMs run in parallel, one from left to right and the other from right to left. Composite output is the forecast of a given target signal. This strategy has shown to be very effective.

TensorFlow Blokdyk (2018) is a Python library used for training and developing machine-learning models. This library was utilized in the General Least-Squares model and is currently being implemented in neural network applications as part of the next steps of this project. While it is used primarily for developing neural networks, it can be used for various tasks, including linear regression. TensorFlow performs its computations using multidimensional data arrays, also called tensors. A tensor describes a multi-linear relationship between sets of algebraic objects concerning a vector space. Tensorflow will also allow the use of NNs since they require tenors to allow neurons in hidden layers to assign weights and values to determine the output layer value.

High-performance computing (HPC) involves the use of parallelization, where data are processed across a multitude of threads. In lament terms, instead of processing data points one by one in a for loop for a processor, the job is divided and given to multiple processors to complete the job simultaneously to yield faster results. Using HPC is common when processing massive datasets with a supercomputer since a typical laptop or desktop does not have enough computational power, such as cores, GPU, and CPU, to compile complex programs for any given massive dataset. In this paper, POSIX Threads (pthreads) and Open Multiprocessing (OpenMP) (Ghildiyal 2014) are used to process the tidy data sets. Pthreading is used to create multiple threads to allow the program to process multiple data points within the program. OpenMP is an API used to execute the region of the program that will be running in parallel, typically supported in high-level programming languages such as C/C++. However, this can be done in other scientific-practical programming languages such as Python, MATLAB, and Fortran. OpenMP allows the option to know whether threads are sharing a memory or if the threads are private. With OpenMP, threads can fork, join, and wait, allowing precise control of how and when code regions are parallelized. Using the Fork and Join Model, threads run in parallel and join at specific points within the program and can "fork" again to run in additional stages of parallelization. However, the number of threads can be chosen to optimize computational speed. The number of threads needs to be determined otherwise; using excessive lines slows the parallelization processes, which is the opposite objective of using HPC.

To sum up, to the best of our knowledge, no existing study on predicting cryptocurrency prices and volatility has concentrated on finding and assessing the temporal effects of variables, which would help financial practitioners concentrate on particular components and variables. Financial service providers would trade more transparently and precisely as well as quickly and reliably make trading decisions while examining the varying effects of cryptocurrency volatility prediction variables over time (Alessandretti et al. 2018). Therefore, by assisting fin-tech companies and literature in forecasting cryptocurrency volatility, our study adds to the body of knowledge in this area by (a) using AI/ML-based feature selection algorithms to build an aggressively sparse model, (b) adopting multiple complex forecasting algorithms to build highly accurate ML-based regression (i.e., high RMSPE), and (c) the research data contributes to the literature by being extensive on the life of DogeCoin (7 years) and 9 years for Bitcoin, and 9 years of Google Search trends for these terms. (d) using Bitcoin and DogeCoin data to determine the dynamic price of these most popular cryptocurrencies.

Theoretical discussions, methodology, and data description

We suggest sparse models and a five-step CRISP-DM technique as data analytics methodology. The main unknown in predicting cryptocurrencies is predicting the price volatility over time for that coin. To predict cryptocurrency price volatility, a scale prepared by a list of numbers ranging from 0 to 100 was generated for a specific date range that represented the number of online searches for a particular cryptocurrency, such as Bitcoin or DogeCoin, on each day. Using this method, an algorithm was created to adjust the least-squares model to match the actual data based on the volatility for each day. This method to relate the online search patterns with cryptocurrency price volatility was then manually compared to see that search volatility patterns fit with cryptocurrency price volatility patterns.

The study follows the industry standard, the cross-industry standard process for data mining (CRISP-DM) methodology. In this methodology, cryptocurrency price data from multiple sources (cryptocurrency and Google search data) go through initial preparation. They are fed to data preprocessing and preparation, where collection, transition, cleaning, filtering, missing data imputation, a transformation of the variables, and data clustering. Then the influential variables for the prediction are selected. The data are randomized by splitting it into ten pieces. Then tenfold cross-validation is applied by testing each piece against the remaining nine training pieces. Tenfold cross-validation ensures that the training and testing are done randomly, and the result of tenfolds are compared. The median fold was found to be random, and machine-learning modeling was done on the random fold dataset. We initially highlighted the price of Bitcoin for 10 years (August 2013–March 2023). We used 9 years of data for prediction (March 2014–March 2023) and randomly cross-validated the predictions in Table 1. Still, we focused on predicting the most volatile timeline for the coin market during and after the Covid-19 crisis for the sake of being able to visualize the positive and negative volatility peaks in data (September-2021–March 2023). The results are evaluated, and finally, the results are visualized. The process requires expertise and knowledge both in machine-learning models and in the financial state of cryptocurrencies. Depending on the results, a feedback mechanism exists to repeat and test the process. Figure 1 shows the demonstration of the process. Further details of the CRISP-DM process can be referred from Schröer et al. (2021).

Here in Table 1, we visualize a sample of the data. The change column shows the percentage of volatility from the previous day and is used as a target variable for prediction. Predicting cryptocurrency volatility has become increasingly necessary due to the intricate dynamics behind its evolution and its relevance in the financial sector, particularly in the financial system. The main factors driving volatility in the financial market are economic conditions, news, social media, international politics, and government regulations. Governments, investors, individuals, and regulators are interested in cryptocurrency because of its prominence and potential to change the current global financial system, adventurous nature, anonymity, exception from conventional government regulations, and transparency by varying motivation for all stakeholders. In addition, a lot of researchers—both academic and professional—have devoted years of their extensive empirical and theoretical work to modeling and forecasting the volatility of financial markets and assets.

Linear regression is commonly used for forecasting problems on matrices such as large datasets within data science. Most datasets are created and stored in the form of a m by n matrix and can be transformed to extract valuable information from the given dataset. To begin measuring the volatility of these specific types of cryptocurrencies, it is important to note each currency will yield different volatility percentages at a time. However, the actual and calculated volatility difference will allow us to predict the price influx. Linear regression can and must be used to create precise but not accurate predictions. In addition to Linear Regression, other computationally advanced methods, such as Neural Networks typically used in Deep Learning, will be used to develop more precise and reliable models to determine cryptocurrency prices accurately.

$$y = x\beta,$$

where x is any mxn design matrix and β is any vector

Similar to the equation of a line,

$$y = mx + c,$$

where $y = mx$ passes though the origin and C is the y intercept with the xy axes.

Therefore, in comparison to the equation of a line, the matrix is a transformation about a specific origin and β intercept. It is a matrix multiplied by a vector to be transformed, giving the values of any possible vectors produced by the system.

In this model, four variables are fitted, creating a vector. The residual vector e (Mean Square Error) is the actual value minus the predicted value. Then the values for β, where the excerpt term is solved where

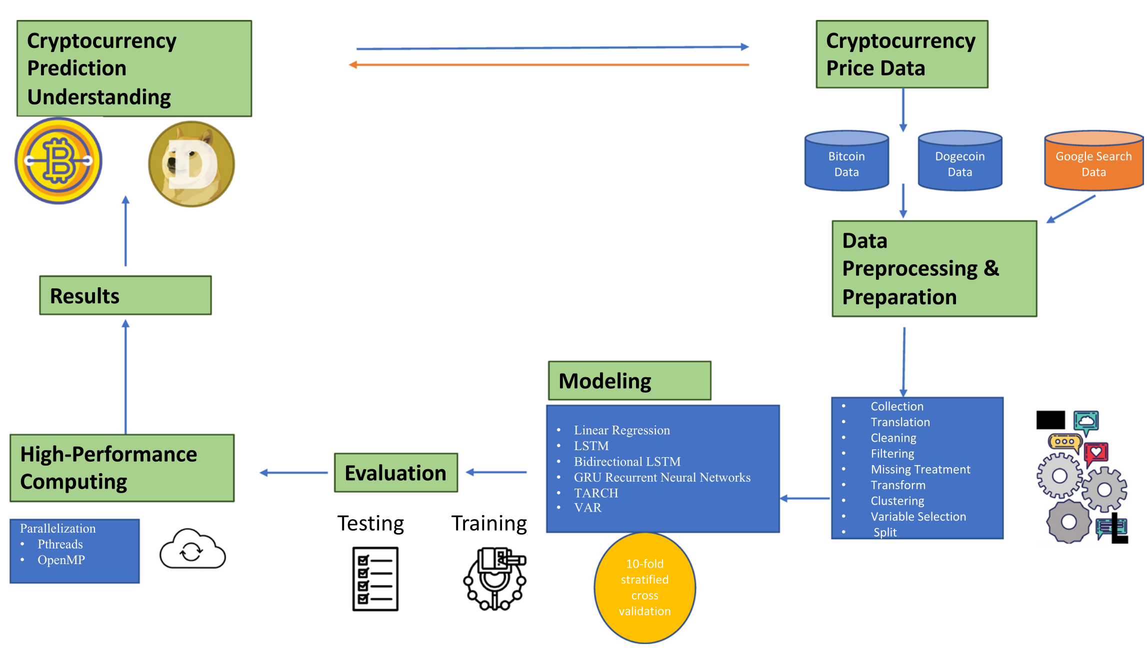


Fig. 1 The methodology of the machine-learning process in this study follows the industry standard CRISP-DM. The figure shows the flow of methods to be utilized to predict cryptocurrency prices. Volatility equations for each coin analyze the difference between actual and predicted volatility to predict price influx. Because of the datasets being used, parallelization techniques such as p-threading with OpenMP can accelerate the execution of the massive dataset through these prediction models

Table 1 Bitcoin market data for a particular asset spanning from August 8th to August 14th, 2022

Date	Price	Open	High	Low	Vol. (K)	Change %
Aug 18, 2022	24,302.8	24,442.1	24,997.3	24,172.2	177.25	– 0.57
Aug 13, 2022	24,442.5	24,398.9	24,882.9	24,318.7	170.59	0.18
Aug 12, 2022	24,398.7	23,935.3	24,440.8	24,616.4	194.96	1.94
Aug 11, 2022	23,935.3	23,963.3	24,873.5	23,864.0	285.36	– 0.12
Aug 10, 2022	23,962.9	23,150.3	24,209.9	22,714.7	243.61	3.53
Aug 09, 2022	23,146.7	23,818.1	24,912.0	22,886.5	169.62	– 2.81
Aug 08, 2022	23,816.3	23,175.3	24,234.1	23,160.6	197.94	2.77

The data includes the date, the opening price, the highest and lowest prices reached, the closing price, the volume traded, and the percentage change in price from the previous day. The data may be useful for analyzing the recent performance of Bitcoin and making informed investment decisions

$e = y - x\beta,$ where y is the volatility.

Therefore, the Mean-squared error of the prediction of β is the following:

$$\frac{1}{n}e^T e = \frac{1}{n}(y - x\beta)^T(y - x\beta),$$

where n is the number of values (Meyers 2009).

$$= \frac{1}{n}(y^T y - y^T x\beta - \beta^T x^T y + \beta^T x^T x\beta).$$

This is how ordinary least-squares regression solves the system to provide a prediction of possible future values with low residuals. The mean-squared error is minimized to create more precise and accurate prediction values. If there is more than one parameter (input value), then partial differentiation is required for each parameter that determines the predicted value.

$$\begin{aligned}\frac{d}{dx}(y - x\beta) &= 0\\ (x^T x)\beta &= x^T y\end{aligned}.$$

Then solve for β which is a vector for the following:

$$\hat{\beta} = (x^T x)^{-1}(x^T y).$$

Therefore, the dot product of x and β, where they are any vector, produces the next predictable point based on previous points by the given system. The following is the matrix notation of the basics of linear regression:

$$x\beta = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix}\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} = \begin{pmatrix} y_1 - (\beta_0 + \beta_1 x_1) \\ y_2 - (\beta_0 + \beta_1 x_2) \\ . \\ . \\ y_n - (\beta_0 + \beta_1 x_n)\end{pmatrix}.$$

Hence, the dot product of a matrix with itself produces the sum of squares:

$$(y - x\beta)^T(y - x\beta) = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2.$$

The predicted values will be denoted as y hat and can be written in term of β,

$$\hat{y} = x\hat{\beta} = x(x^T x)^{-1}x^T y.$$

General least-squares regression aims at finding the best coefficients in a linear model for a given dataset. It uses the model:

$$y = \beta_0 + \beta_1 x,$$

where y is the independent variable, β_0 is the intercept, β_1 is the first coefficient, and x is the mean of the independent variables to minimize the sum of squares for the estimated residuals where least squares want to minimize any predicted values deviations from the dataset's actual values (Dismuke and Lindrooth 2006). For a linear model, this would mean minimizing the residuals, which is the objective function:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

For a polynomial, this would mean minimizing

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m)^2.$$

The closer $r^2$is to 1, the better model fits. The parameter is given by the formula:

$$r^2 = \frac{(S_t - S_r)}{S_t},$$

where S_t is the squared sum of the residuals from the data, and, . S_r is the squared sum of the residuals from the estimate.

GARCH is a statistical model that analyzes and forecasts volatility in time-series data. It accounts for the non-constant variance, or heteroscedasticity, modeling the variance as a function of the past variance and past squared residuals can identify common patterns in financial and economic data. The GARCH model allows for the estimation of conditional volatility and can capture the persistence, clustering, and leverage effects commonly observed in financial markets. It is widely used in financial risk management, asset pricing, portfolio optimization, and marketing analytics problems such as risk segmentation (Cho and Korkas 2022).

The GARCH model is specified as follows:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t = z_t\sigma_t,$$

where zt are the innovations and developments, yt is the earnings, μt is the conditional variance, σt is the volatility process of the conditional variance. where y_t is the earnings μ_t is the conditional variance σ_t, is the valaitility process of the conditional variance, z_t, are the innovations and developments ϵ_t is the error term, ω is a constant, α is the coefficient on the lagged squared error term ϵ^2_{t-1}, and β is the coefficient on the lagged conditional variance term σ^2_{t-1}. The maximum order is one, which means that $p = q = 1$.

The E-GARCH model is a variation of the GARCH model that includes an additional term to to detect the asymmetric reaction of volatility to positive and negative shocks. The formula for an E-GARCH model is

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma^2_{t-1}) + \gamma \frac{\epsilon_{t-1}}{\sqrt{\sigma^2_{t-1}}} + \alpha\left[|\epsilon_{t-1}|\sqrt{\sigma^2_{t-1}} - \sqrt{2\pi}\right].$$

In this equation, σ_t^2 is the conditional variance of the time series at time t, ω is a constant, β is the coefficient on the lagged conditional variance term σ^2_{t-1}, γ is the coefficient on the lagged standardized error term $\epsilon_{t-1}/\sqrt{\sigma^2_{t-1}}$, α is the coefficient on the term $|\epsilon_{t-1}|\sqrt{\sigma^2_{t-1}} - \sqrt{2\pi}$, and $\ln(\cdot)$ is the natural logarithm function.

These formulas can be implemented in statistical software such as R, MATLAB, or Python to estimate and forecast volatility in financial time-series data, including cryptocurrency and world currency markets.

A popular time-series model for forecasting future values of a time series based on past values is the ARMA (Autoregressive Moving Average) model. The model consists of two parts: the autoregressive (AR) part and the moving average (MA) part. The current value of the time series is dependent on its past values in the AR component of the model, whereas the current value is dependent on the past errors in the MA part.

The ARMA model can be written as follows:

$$(\mathbf{y}_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t),$$

where y_t is the observed value of the time series at time t, c is the constant, ϕ_i and θ_j are the corresponding coefficients for the moving average and autoregressive terms. and ϵ_t is the error term at time t.

To fit an ARMA model to a time series, one typically uses maximum likelihood estimation to estimate the model coefficients ϕ_i and θ_j. Once the coefficients are estimated, the model can be applied to forecast future time series values.

The vector autoregressive (VAR) model is also an autoregressive bi-directional time-series model, a subtype of ARMA models. The VAR model is an ML method that helps us to understand how multiple variables influence each other over time. It uses the past values of each variable and the past values of the other variables to predict future values. The model helps us to analyze the relationships and cause-and-effect between variables and can be useful for predicting future trends, understanding how one variable affects another, and determining how much of the variability in each variable is explained by the others. It is a widely used tool in many fields where time-series data are important.

$$(\mathbf{y}(t) = c + A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + e(t),$$

where $y(t)$ is a n-dimensional vector of endogenous variables at time t, $A_1, A_2, \ldots, A_p$ are coefficient matrices, c is a constant vector, p is the order of the VAR model, and $e(t)$ is a vector of error terms (Wang 2022).

Dynamic Conditional Correlation Generalized Autoregressive Conditional Heteroscedasticity (DCC-GARCH) Model:

$$R(t) = V(t)^{1/2} * e(t),$$

where $R(t)$ is a k-dimensional vector of returns at time t, $V(t)$ is the $n \times n$ conditional variance-covariance matrix at time t, $e(t)$ is a n-dimensional vector of standardized residuals at time t, and $^{1/2}$ denotes the square root.

$$V(t) = D(t) * R(t) * D(t),$$

where $D(t)$ is a diagonal matrix of conditional sd which means standard deviations.

VAR-DCC-GARCH Model:

$$\begin{aligned} y(t) &= c + A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + e(t) \\ e(t) &= L(t)^{1/2} * z(t) \\ z(t) &= u(t)/(u(t)' * u(t))^{1/2} \end{aligned},$$

where $y(t)$, $A_1, A_2, \ldots, A_p, c, p$, and $e(t)$ are as defined in the VAR model, $L(t)$ is the diagonal matrix of conditional variances, and $z(t)$ is the vector of standardized residuals. The vector $u(t)$ is the vector of possible errors in the DCC-GARCH model.

ARDL (Autoregressive Distributed Lag) is a time-series econometric technique that is used for modeling the relationship between two or more variables that are non-stationary in nature. It is a popular method for analyzing cointegration between variables that are integrated of different orders, such as I(0) and I(1). ARDL allows for estimating both short-run and long-run relationships between the variables.

The basic formula for the ARDL model is

$$\Delta Y_t = \alpha + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta X_{t-1} + \beta_3 X_{t-1} + \epsilon_t,$$

where ΔY_t is the first difference of the dependent variable at time t, ΔX_{t-1} is the first difference of the independent variable at time $t-1$, X_{t-1} is the level of the independent variable at time $t-1$, and ϵ_t is the error term (Nonejad 2022).

ARDL has been employed in numerous research studies to forecast cryptocurrency prices and volatility. The fundamental concept of using ARDL for cryptocurrency price prediction is to develop a model that reflects the relationship between the cryptocurrency price and its essential factors, including trading volume, market capitalization, network activity, and media attention. This approach aims to uncover the complex associations between these variables, thus, enabling the prediction of future price trends and fluctuations.

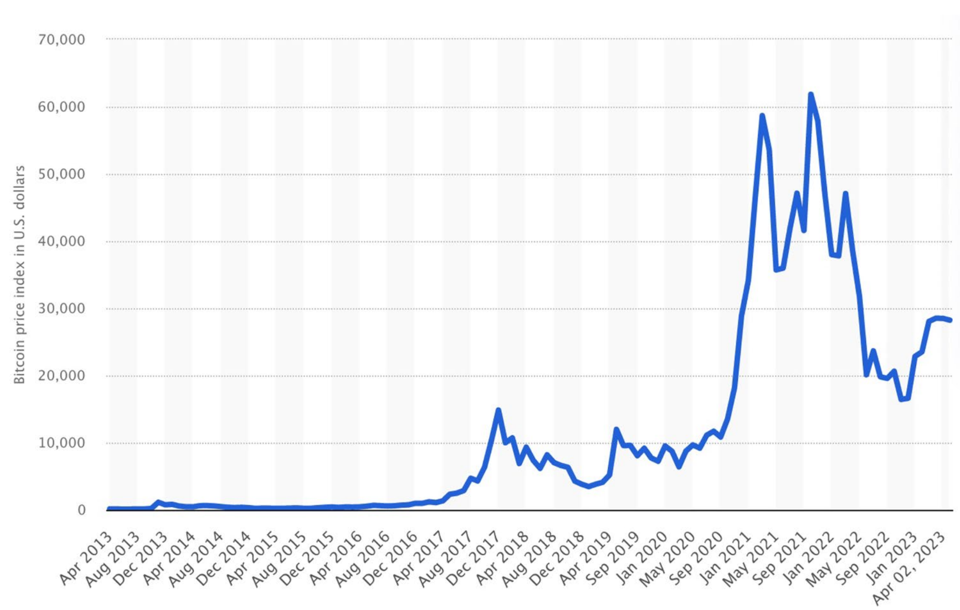
Figure 2 shows the BPI of Bitcoin from 2013 to 2023. BPI is a measure of the price of Bitcoin, a popular cryptocurrency, and is based on the weighted average price of several leading exchanges. The graph shows that the price of Bitcoin experienced a significant increase in 2017, reaching an all-time high of nearly 20,000 US Dollars. However, the price then experienced a sharp decline and remained relatively stable for several years before increasing again in late 2020 and early 2021, reaching a new all-time high of over 60,000 US Dollars. The graph illustrates the volatility of the cryptocurrency market and the potential for rapid fluctuations in the price of Bitcoin.

The least-square model is beneficial due to the fact that residuals can be treated as a continuous quantity where the derivatives can be found. The significant disadvantage of a least-square regression model is the vulnerability to the drastic change in the model when outliers come into play. The least-squares regression model was not enough to have a good model, so a volatility parameter was created to adjust the model correctly. This was done by taking the volatility data and finding its change from one point to another. This change was then manipulated and applied to the already predicted prices. This was perfected by trial and error by adjusting the volatility coefficient until the model was fitted correctly.

Figure 3 shows the change in volatility concerning time. As the figure shows, a large influx in volatility is at the point where Bitcoin took off. Using this, we can adjust the least-squares model off of the change in volatility to predict the price better. With minor changes in volatility, this should not affect the price as much or at all. The only thing the volatility parameter considers is the drastic changes. The slight differences are filtered out in the code while the significant changes are saved. These values are manipulated by dividing by a factor chosen by trial and error so that the model fits the best. This factor found is how much the volatility affects the coin's price. Using the least-squares regression model and attaching a volatility coefficient, an excellent model can be found to predict the price of Bitcoin and DogeCoin accurately.

The polyfit function fits least-squares nth-order polynomials to data. The function: $p = polyfit(x, y, n)$ would be implemented, where x is the independent data, y is the dependent data, n is the order of the polynomial to fit, and p is an array of the polynomial's coefficients. To further compute a value,

Fig. 2 Bitcoin Price Index (BPI) from 2013 to 2023



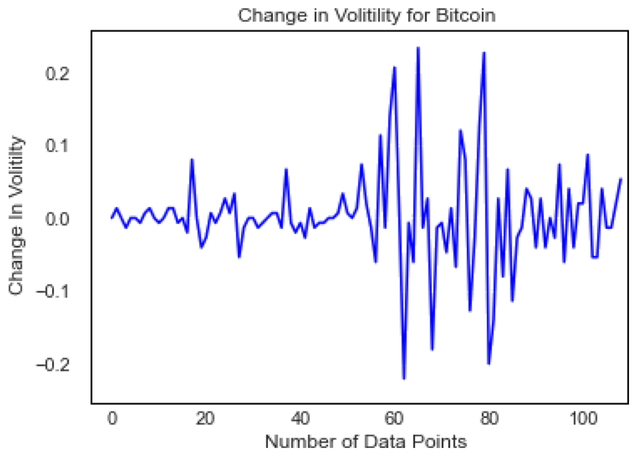


Fig. 3 Change in volatility for Bitcoin. This figure shows the change in volatility for Bitcoin in relation to time

the polyval function can be used utilizing polynomial values as follows: $y = polyval(p, x)$.

Auto Regression is mainly used for economi and signal processing related datasets (Zhang et al. 2017). It is a time-series model which uses observations from previous time steps as inputs to a predetermined regression equation to predict the value at the next time step. As a result, its application can result in accurate forecasts on a range of time-series problems, such as stock or cryptocurrency price predictions (Eric and Brown 2019). The model shows how the output depends, at least partly, on randomness alone. In economic concepts, such as stock price prediction (Stock price 2021), random walk theory describes a path of random steps that data follow (Kim et al. 2016). In other words, a specific outcome or conclusion is known to come from the data. Still, the method, direction, or order in which this outcome or conclusion is reached is achieved by random, unpredictable movements.

Results

As an initial model tune-up initial average for volatility is normalized across the year, where we can assume that any volatility above the average could inflict change on the price; the higher the volatility, the more significant the difference in the price, and the lower the volatility shows the effect on the price. Using the General Least-Squares Regression method without considering the volatility, accurate plots for Bitcoin and DogeCoin are obtained, testing successfully for the initial theory. It was also observed that future prices were predicted based on the size of the dataset and the number of days used in that specific dataset. Figures 4 and 5 both show the search volatility scaled for Bitcoin and DogeCoin, respectively, from 11-01-2019 until 11-01-2021. This date was chosen because it was far back enough to visualize the least-squares model to fit the data. Also, this period includes most of the profound changes in cryptocurrency prices that were seen. The Internet search trends in Figs. 4 and 5 and actual cryptocurrency price volatility can be done by comparing the same search timeline with Figs. 15 and 16.

Figures 4 and 5 show the rise of the search on the internet for these coins while their price also rises in the same timeline. One of the essential qualities of a good analysis model is visualizing and explaining the intuition of the prediction.

Based on these results Table 2, it can be concluded that the deep-learning meach models (LSTM, Bi-directional LSTM, and GRU) generally outperform the linear regression model and VAR model for predicting Bitcoin volatility. Among the deep-learning models, the GRU model appears to be the most accurate, although the difference in performance between the three models is relatively small. Additionally, the Bi-LSTM model also provides a reasonable level of accuracy, suggesting that statistical models may also be effective for predicting Bitcoin volatility.

It is worth noting that the best model for predicting Bitcoin volatility may depend on the specific characteristics of the data and the problem being solved. Therefore, further research may be needed to confirm this study's findings and identify the most effective model for different scenarios. In the current Bitcoin price volatility scenario, the GRU Neural Networks algorithm is the champion model with the lowest RMSPE(0.0418). Additionally, the GRU model has one of the lowest MAE(1030), RMSE(1415), and MSE(2003410) performance metrics. The most important metric for price volatility is RMSPE, and neural networks algorithms show their superiority in predicting this time-series forecasting task.

Figures 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20 will tell how efficiently these algorithms make predictions (Petrescu and Gironda 2019). Figures 6 and 7 show us Bitcoin and DogeCoin (Frankenfield 2021) model for the ordinary least squares. As you can see, the model gives a reasonable value estimate. This model is suitable for showing the change in the direction of the price but does not perform great in predicting the actual price. This model is great for long-term outcomes, but for seeing the inflections and price changes in a short amount of time, this model does not perform well.

Figure 8 shows the volatility and the actual price of Bitcoin. As seen in Fig. 8, volatility's influence affects the price significantly. This creates a relationship as the volatility or the interest of Bitcoin increases or decreases; the price is more likely to increase or decrease as well. As the change of volatility becomes more remarkable, the difference in the price will see the same shift. Since the two are so dependent, the model will improve if the volatility correlates with the current model. A volatility coefficient had to be created

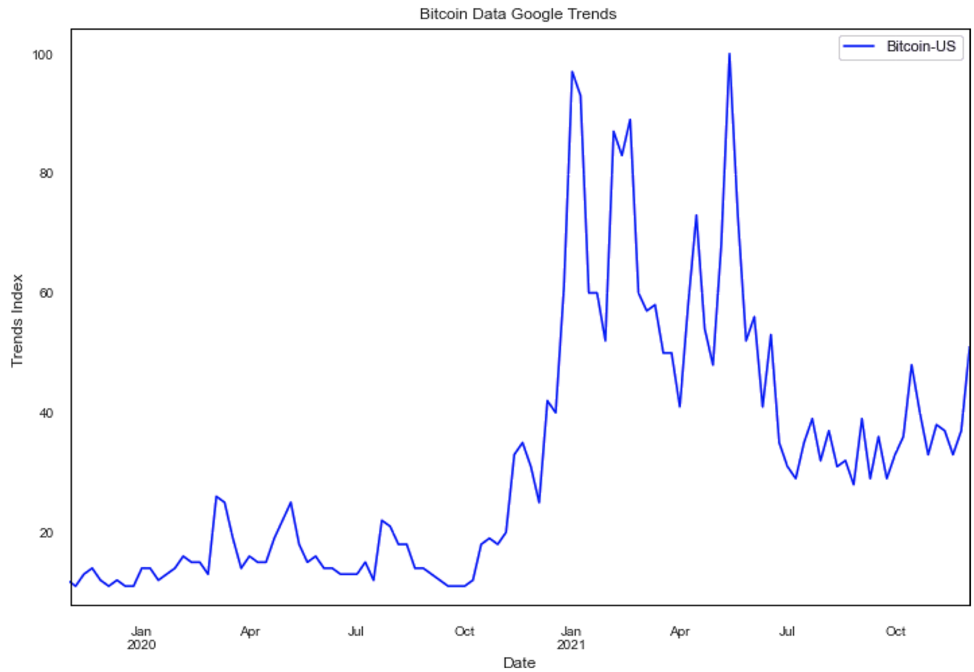


Fig. 4 Bitcoin volatility. This graph shows the volatility of Bitcoin over the past year, with the scale going from 0 to 100. 0 meaning that bitcoin was not being looked up or talked about, and 100 meaning that bitcoin was being searched and talked about a lot on the internet

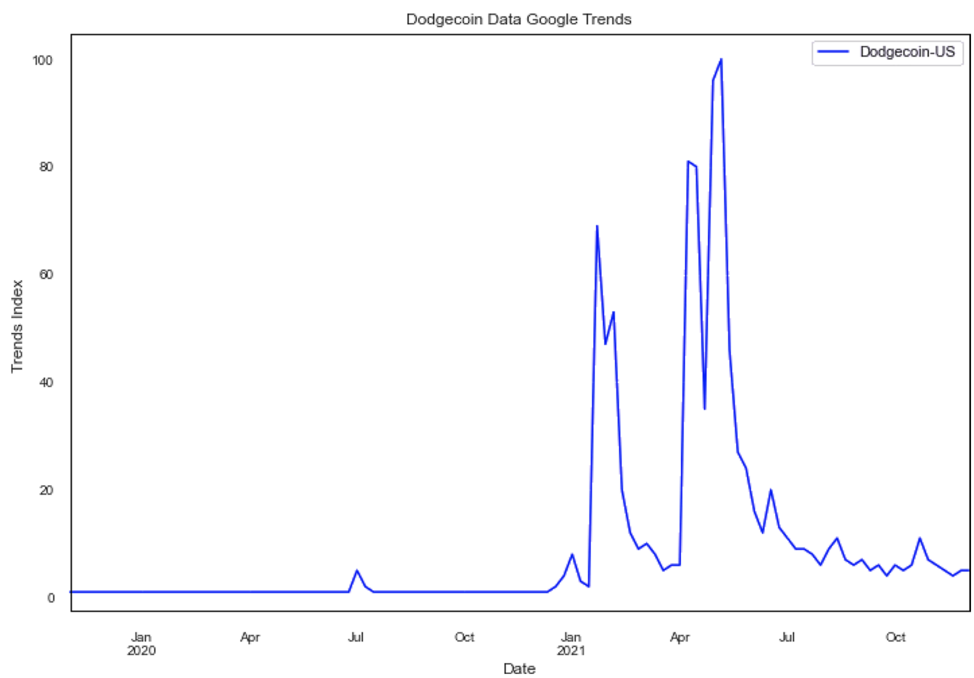


Fig. 5 DogeCoin volatility. This graph shows the volatility of DogeCoin over the past year, with the scale going from 0 to 100. 0 meaning that the coin was not being looked up or talked about, and 100 meaning that the coin was being searched and talked about on the internet

to associate the volatility data and incorporate that into the model, as seen in Fig. 4.

Figure 9 shows the updated model with the volatility parameter included. As seen in the Figure, this model matches the actual bitcoin prices more precisely than the original least-squares model, which can be seen in Fig. 10.

The following, Figs. 11 and 12 show the Regression algorithm results for both cryptocurrencies. The goal of Linear Regression is to show the general trend of the prices. As this paper is being written, the Regression showed a negative trend that was proven true when Bitcoin plummeted to as low as \$42, 333. The Linear regression algorithm becomes overfitted with more terms added to the series. Therefore, it is not a precise model other than being used to analyze the general trend of the coin.

Figures 11 and 12 show that the machine-learning model can precisely predict future Bitcoin and DogeCoin prices based on its dataset. More precise machine-learning models

Table 2 The table presents the performance evaluation metrics for different models used in the paper

	MSE	RMSE	MAE	RMSPE
Linear regression	14,914,797	3861	2963	0.1033
LSTM	2,845,040	1686	1191	0.0535
Bi-LSTM	3,425,114	1850	1403	0.0506
GRU	2,003,410	1415	1030	0.0418
TRACH	384,400	620	603	0.2009
VAR	613,715	783.4	253.4	3.2254

These metrics provide insights into the accuracy and predictive power on Bitcoin price volatility prediction. The models considered are Linear Regression, Long short-term memory (LSTM), Bi-directional long short-term memory (Bi-LSTM), Gated recurrent unit (GRU), Threshold autoregressive conditional heteroskedasticity (TARCH), and Vector autoregression (VAR)

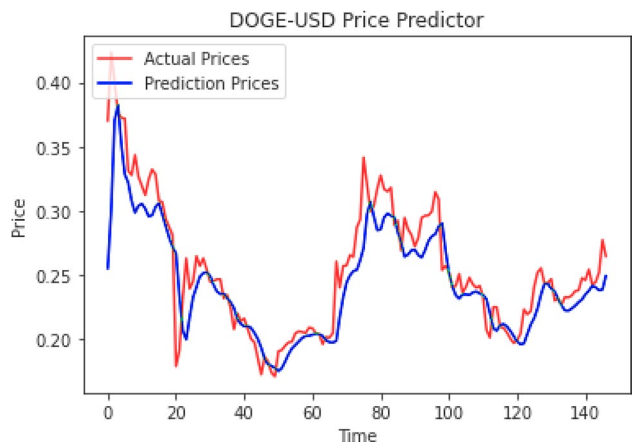


Fig. 6 DogeCoin price prediction—Least-Square Regression was used to fit the best-fit prediction curve for the DogeCoin dataset, where the red curve is the actual close price and the green curve is the prediction curve. (Color figure online)

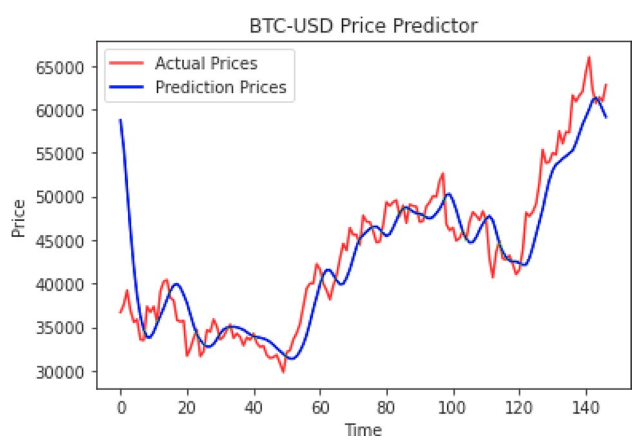


Fig. 7 Bitcoin price prediction—Least-Square Regression was used to fit the best-fit prediction curve for the Bitcoin dataset, where the red curve is the actual close price and the green curve is the prediction curve. (Color figure online)

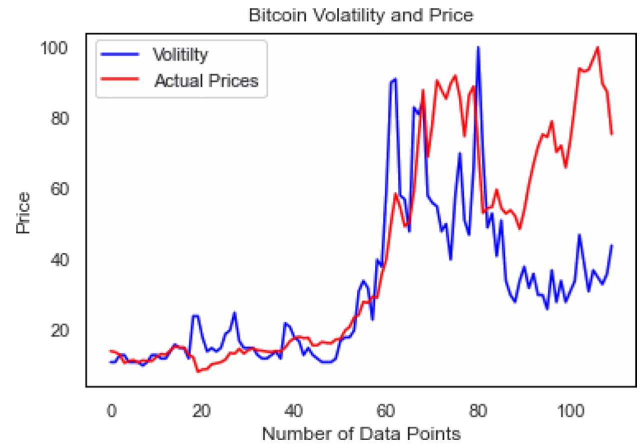


Fig. 8 Bitcoin Volatility and Actual Price. This graph illustrates the relationship that the volatility has on the actual price of Bitcoin

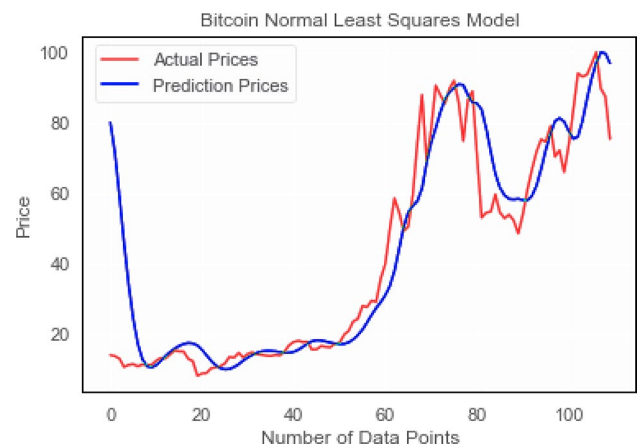


Fig. 9 Bitcoin Price Prediction. This graph shows the bitcoin price prediction according to the least-square regression model. This graph does not include volatility parameters

may include using Neural Networks such as Recurrent Neural Networks and/or Time-Series models, commonly used in data visualization and mining.

In Fig. 14, a specific neural network such as Reset and Update Gate is similar to the Recurrent NN model. The GRU model combines input data with previous memory and utilizes an update gate to determine how much last memory to use for the next task. Although the model is accurate, hyperparameters need to be tuned to better fit the data since this specific deep-learning model is to be far more accurate than traditional deep-learning neural networks.

This model shows the accuracy of GRU morel in an extended amount of time. Hyperparameter optimization is one of the most essential tasks for neural networks algorithms. We have chosen 800 Epochs and the batch size is 50. We believe the prediction accuracy of the GRU model

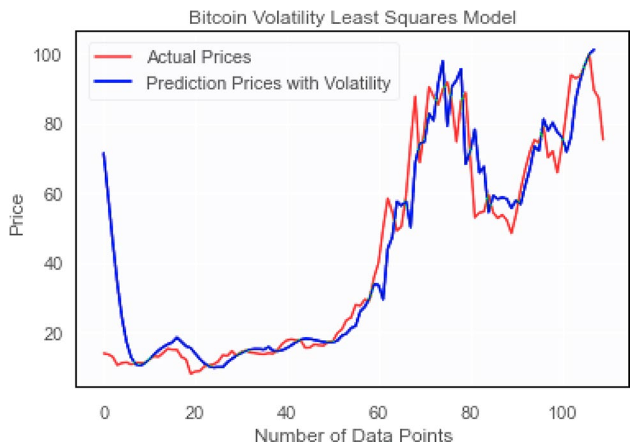


Fig. 10 Bitcoin Volatility with Regression and Actual Price. This graph shows the updated model with the volatility parameter Incorporated with the least-squares regression model

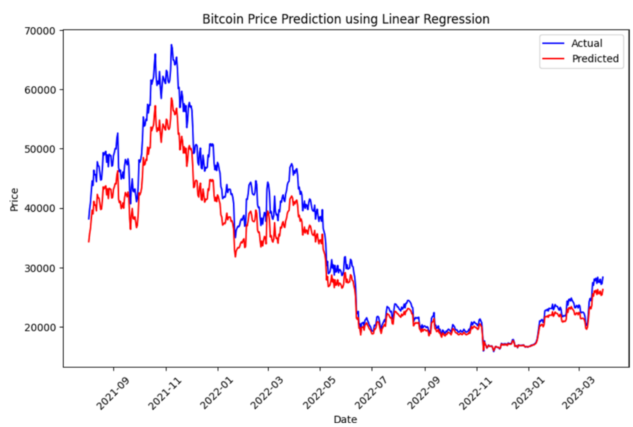


Fig. 11 Bitcoin price prediction using basic machine-learning regression techniques by splitting the data into a training and testing set. Gold color is the actual price and pink is the predicted price using the regression model. (Color figure online)

can be improved by doing further optimization of the hyperparameters. The results for the current model are sufficient to prove the prediction of volatility. Even minimum knowledge about cryptocurrency volatility prediction means significant financial values, the champion model GRU depicts the importance of the study through these visualizations.

In Fig. 15, we can see how the results of a neural network would affect price prediction without considering any other volatility parameters. This figure shows how precise LSTM neural network is. For example, with the recent Russian-Ukrainian War, Bitcoin prices slumped when Russian forces invaded Ukraine in 2022, and the algorithm successfully predicted that price decrease.

There has been a slight increase in coin prices since 2023, and Fig. 16 is an indicator to predict the increase in DogeCoin in 2023 successfully.

Bitcoin price prediction using bi-directional LSTM (BLSTM) and Long short-term memory (LSTM) are two popular deep-learning approaches. Both techniques can learn from Bitcoin's historical price data and use this information to make future price predictions. However, there are some differences between the two methods.

A particular kind of recurrent neural network called an LSTM is made specifically to process sequence data and It works by storing information about past inputs and utilizing that information to anticipate what will be inputs in the future (Lima 2022). The LSTM model learns patterns from historical price data to predict future prices for bitcoin by using that data as input. LSTM algorithm applied to DogeCoin data is presented at Fig. 16.

A Bi-LSTM is a type of LSTM that can learn from both past and future inputs. It works by running two LSTM networks in opposite directions. One for past input and one for future input. This allows the model to capture data trends, potentially leading to more accurate predictions. The Bi-LSTM algorithm applied to BitCoin data is presented at Fig. 17.

Both Bi-LSTM and LSTM have proven to be effective when it comes to Bitcoin price prediction. However, some studies have found Bi-LSTM to be more accurate. This is because Bi-LSTM can capture both past and future trends in data, while LSTM can only capture past trends.

Generalized autoregressive conditional heteroskedasticity (GARCH) models are frequently used to forecast volatility because they were created with the intention of modeling a time series' conditional variance, which is a key characteristic of financial data such as Bitcoin prices.

Generalized autoregressive conditional heteroskedasticity (GARCH) models are widely used for volatility forecasting because they are specifically designed to model the conditional variance of time series, a key feature of financial data such as the Bitcoin price increases.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are powerful machine-learning models for sequential data, but they may not be the best choice for predicting volatility in financial time series. This is because it was not specifically designed to model the heteroskedasticity (i.e., differential volatility) often observed in financial data.

Moreover, the GARCH model is widely used and tested in finance and has a strong rationale, making it a popular choice for volatility forecasting. It is also computationally efficient and easily estimated using standard statistical software. However, there is no one-size-fits-all approach to modeling financial time series, and the model choice ultimately depends on the specific problem and data at hand. LSTM, GRU, and TARCH models are practical in when predicting

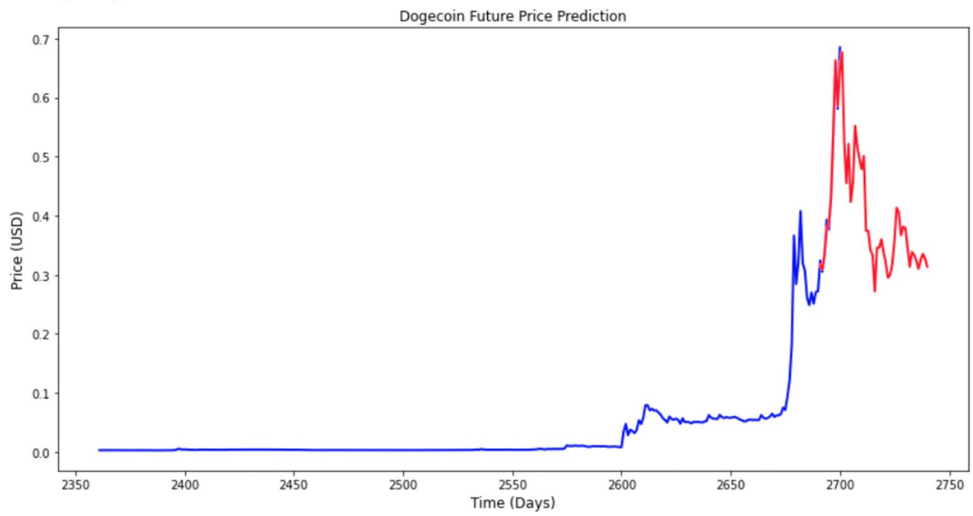


Fig. 12 DogeCoin price prediction using basic machine-learning regression techniques by splitting the data into a training and testing set. Blue color is the actual price and pink is the predicted price using the regression model. (Color figure online)

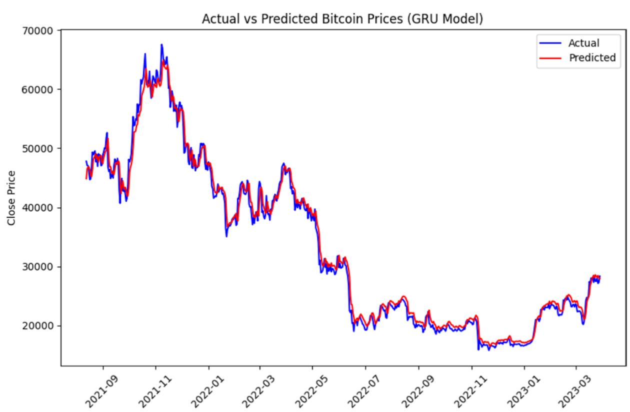


Fig. 13 Bitcoin Price Prediction using GRU. This graph shows bitcoin price prediction using Update and Reset Gate. The model shows the success of GRU model in predicting more extended timelines. As shown in the graph, the model does not accurately predict Bitcoin. However, hyperparameters can be tuned to increase accuracy. Epochs = 800, Batch Size = 50

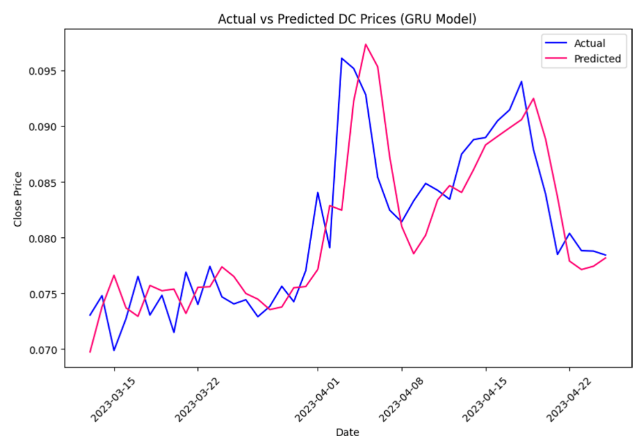


Fig. 14 DogeCoin Price Prediction using GRU. This graph shows the price prediction of DogeCoin. The blue line on the graph shows the actual price of DogeCoin over time, while the pink line shows the predicted price. As shown in the graph, GRU model accurately predicts Bitcoin even when there are high fluctuations in a concise amount of time. (Color figure online)

price trends or directions and predicting volatility. Linear Regression is also practical by being simple and explainable. Although it does not have the best performance numbers, we can understand the patterns better and make financial inferences to predict cryptocurrency volatility.

The image in Fig. 18 depicts Bitcoin's actual instability on the y-axis and distinct time intervals on the x-axis. The term "realized instability" refers to the actual instability of Bitcoin over time, as opposed to the suggested instability caused by Bitcoin choice costs.

The size of each bar demonstrates Bitcoin's actual instability over the time period in question. The time intervals range from 7 days to sixty Understanding historical volatility patterns in this graph can help investors manage risk and develop effective trading strategies. Understanding historical volatility patterns in this figure can help investors to manage risk and develop effective trading strategies.

According to this graph, Bitcoin's realized instability changes primarily over time intervals, with shorter intervals typically exhibiting higher instability. Particularly, it appears from this chart that Bitcoin's realized instability is highest in one- and seven-day intervals and moderately low in longer intervals like 30- and 60-day intervals.

Overall, the chart provides insights into Bitcoin's historical volatility patterns over different time periods, which can help investors to understand the risk profile of their Bitcoin investment and develop effective trading strategies.

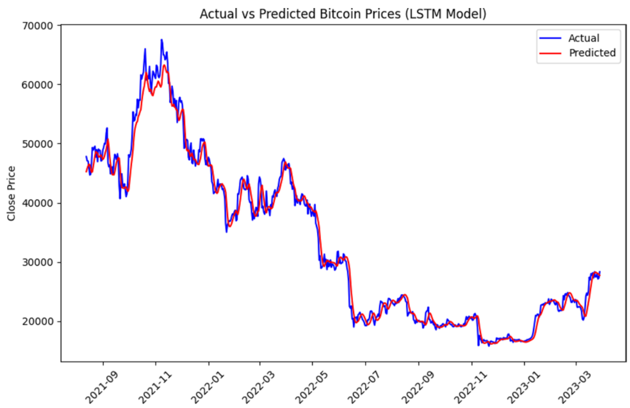


Fig. 15 This graph shows the price prediction of Bitcoin according to the LSTM neural network. This graph does not include volatility parameters

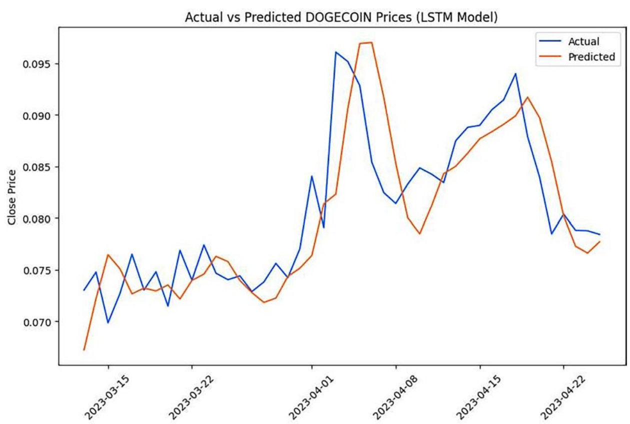


Fig. 16 This graph shows the price prediction of DogeCoin according to the LSTM neural network

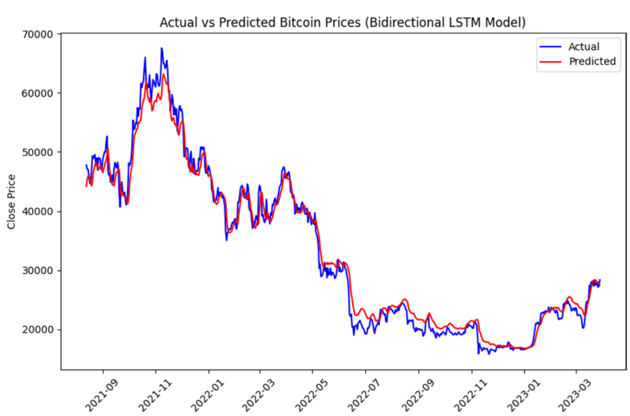


Fig. 17 Bitcoin Price Prediction with Bi-directional LSTM. This graph shows the price prediction of bitcoin according to the Bi-directional LSTM neural network

This plot in Fig. 19 shows the price prediction results using the GARCH model for a DogeCoin over a certain period of time.

Based on this plot, it appears that the GARCH model for price prediction achieved an accuracy rate of approximately 94.67%. While this is a high level of accuracy, it is worth noting that the model may still have room for improvement, mainly if the goal is to accurately predict prices over more extended periods or during periods of high volatility in the market.

However, it is worth noting that other models, such as the GRU model, perform better for price prediction tasks. In fact, in some cases, such as with Bitcoin, the GARCH model might perform much better than in this example.

Overall, this plot suggests that the GARCH model may not be the most accurate model for price prediction for this particular asset and time period. However, more research is needed to determine which models work best for different assets and market conditions so we applied constrained version of GARCH which is TARCH.

Figure 20 shows the actual and predicted volatility of Bitcoin using the TARCH model. The x-axis represents time in days from approx. 2021-08 to 2022-08. The Y-axis represents Bitcoin's daily volatility.

The blue line represents the actual volatility, and the red line represents the predicted volatility. Forecast volatility is based on the TARCH model, a time-series model that captures the non-linear relationship between historical and current volatility. The figure shows that the TARCH model is able to capture the overall trend of Bitcoin's volatility, with the predicted volatility line closely following the actual volatility line. However, there are periods in which the predicted volatility is significantly higher or lower than the actual volatility, suggesting that the TARCH model may have some limitations in accurately predicting volatility under certain market conditions. Overall, this figure demonstrates the potential usefulness of this TARCH model in predicting Bitcoin volatility.

Discussion and future studies

Each mode has advantages and disadvantages for its computational speed, predictive power, explainability, noise sensitivity, and overfitting. One disadvantage of using General Least-Squares Regression is this method is very sensitive to outliers. If the dataset consists of points outside the range of the remainder of the data, the least-squares regression will try to perform a "best fit" that in some way incorporates this point(s), which will change the true accuracy of the regression model. On the contrary, polynomial regression is also advantageous as it can fit a broad range of a function under it.

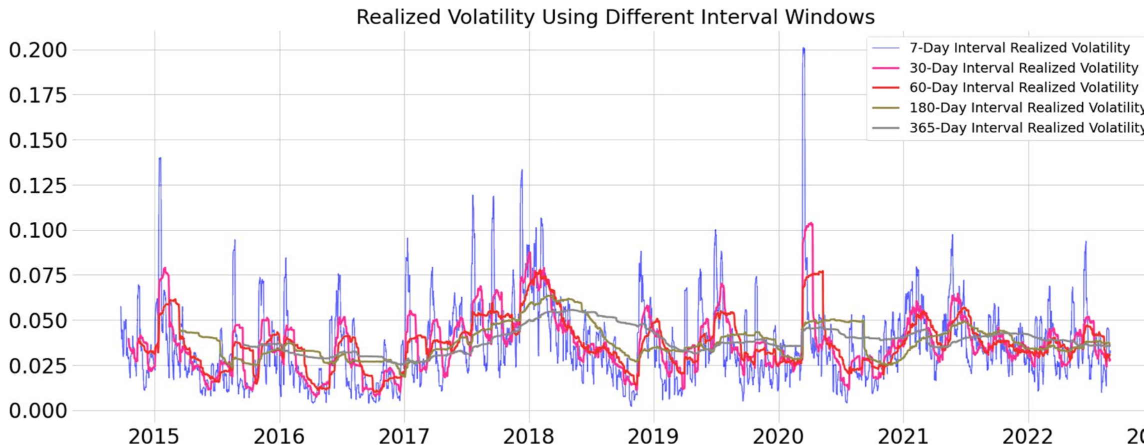


Fig. 18 Line chart showing Bitcoin's realized volatility over different time intervals ranging from 7 to 60 days. Shorter time intervals have higher volatility than the 7-day interval having the highest volatility

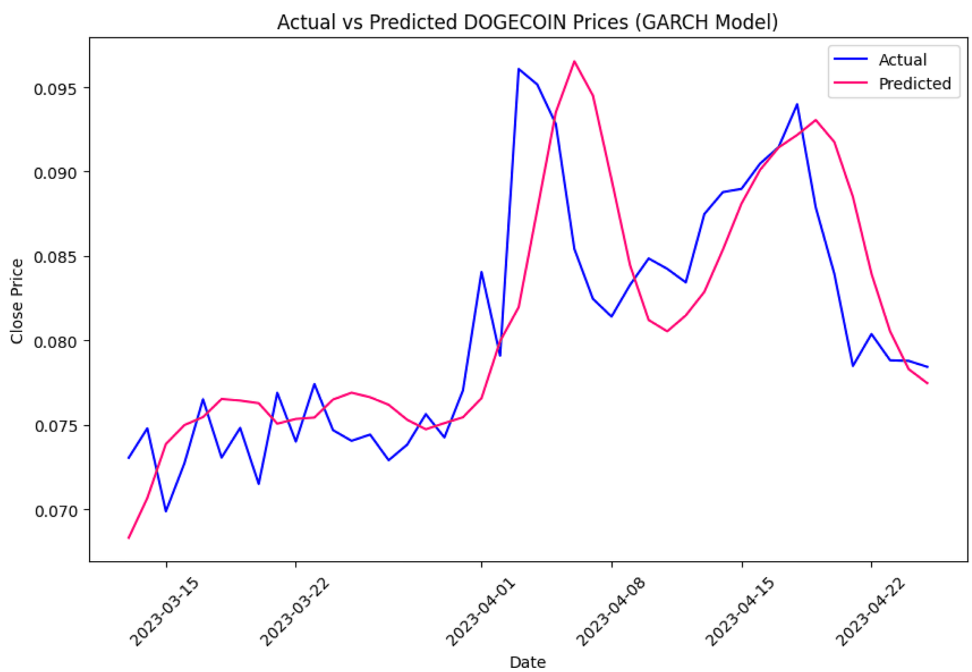


Fig. 19 DogeCoin Price Prediction using GARCH. This graph shows the price prediction of DogeCoin.The blue line on the graph shows the actual price of DogeCoin over time, while the orange line shows the predicted price. (Color figure online)

Linear regression is easier to work with when adding additional terms to the linear regression expression. Although regression models are not the best method for price predictions, Recurrent Neural Networks, specifically Long Short-Term Memory, would play a significant role in determining more precise cryptocurrency price prediction. The TARCH model captured the overall trend of Bitcoin's volatility, with the predicted volatility line closely following the actual volatility line.

However, there were periods in which the predicted volatility was significantly higher or lower than the actual volatility, indicating that the TARCH model may have some limitations in accurately predicting volatility under certain market conditions. Despite these limitations, our findings suggest that the TARCH model can be valuable for investors and analysts in predicting Bitcoin volatility. The results of our analysis also demonstrate the potential usefulness of the GRU, LSTM, Bi-LSTM, and TARCH models in predicting the volatility of Bitcoin. Further research could explore ways to improve the model's accuracy, perhaps by incorporating additional variables or adjusting the model parameters. Overall, our study contributes to the growing body of literature on Bitcoin volatility prediction and provides insights that could inform investment decisions in this emerging market.

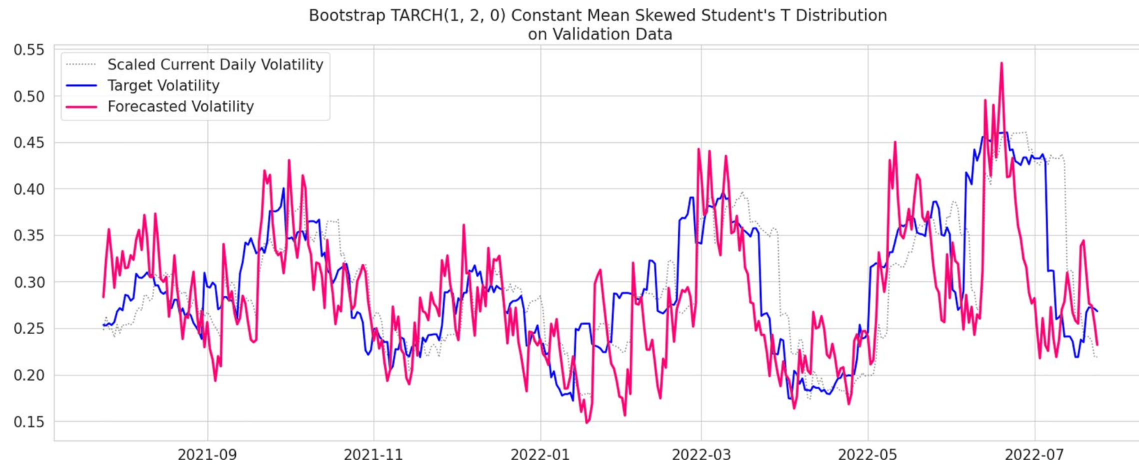


Fig. 20 Actual and predicted volatility of Bitcoin using the TARCH model. The blue line represents the actual volatility of Bitcoin, while the orange line represents the predicted volatility based on the TARCH model. The figure shows that the TARCH model can capture the general trends in the volatility of Bitcoin, but also has limitations in accuratelypredicting volatility in certain market conditions. This suggests the potential usefulness of the TARCH model in forecasting Bitcoin volatility. (Color figure online)

Future studies

Considering Cryptocurrencies' extensive history and dataset, it would also require the model to be trained and more precise in predicting prices based on multiple factors, such as having stock market prices, currencies, and global events. These factors include the word "Bitcoin" trending on various social media platforms, the number of Bitcoin miners, the S&P 500 Index, and several powerful influencers that can increase or decrease the coin's value. The specific social media platforms that can impact the value of Bitcoin can include a variety of channels where discussions and information about cryptocurrency are shared. These can consist of popular platforms like Twitter, Reddit, Facebook, and other online forums and message boards where people discuss and share information about Bitcoin and other cryptocurrencies. As social media can be a powerful tool for shaping public opinion, it can significantly impact the price and value of cryptocurrencies in future studies with better prediction accuracy. It is proven that some social media usage happens with apathetic motivation, but intrinsic and extrinsic motivations coexist with social media usage (Hansen and Levin 2016). It is known that search intensity correlates with stock returns, and we believe that social media activity can be used to predict coin prices (Tajdini 2022). There is also evidence of the trends in social media and celebrities' impact on cryptocurrency prices (Zhang and Huang 2018). A limitation of the study is considering Google search volatility patterns for the "Bitcoin" and "DogeCoin" terms being manually compared with cryptocurrency actual price patterns. In future studies, cryptocurrency price datasets, social media, and Google search datasets should be automatically merged to improve the performance of the predictions. The satisfactory initial results from cryptocurrency prices are accepted as sufficient for this study. Neural Networks would provide more opportunities for creating a more robust model with diverse data that can use other cryptocurrencies and social media datasets and offer price predictions for various cryptocurrencies other than Bitcoin and DogeCoin.

In future studies, the following steps would be needed to determine other parameters that contribute to the volatility of these cryptocurrencies since various social media platforms influence the values of these coins and tokens in future studies. Future studies should conduct additional research to identify other parameters influencing cryptocurrency volatility, such as regulatory factors or technological advances. New models should be developed, which better predict cryptocurrency behavior and volatility over time, considering a wider range of factors. It will also beneficiary to work with cryptocurrency and financial experts to gain further insight into the factors that contribute to the value and volatility of these assets. Continuously monitoring and updating models to reflect new data and changes in the cryptocurrency landscape would also be needed in future studies.

The data's quality and quantity are essential to produce robust analytic models (Petrescu and Krishen 2020a). It is imperative to continue to train the model with additional training/testing datasets and include machine learning to create accurate predictions of future prices. After the models are finetuned, they can be used and applied to other types of cryptocurrency giants such as Ethereum, Ripple, and Cardano since these types of cryptocurrencies play a critical part in the growth of more efficient online transactions, payments, and NFTs (non-fungible tokens). However, as stated earlier, using more robust models such as non-linear regression and Neural networks would project more precise values due to deep learning.

Conclusion

In this study, the volatility of the cryptocurrencies is predicted. The study contributes to the literature by proving the predictability of Bitcoin and DogeCoin volatilities even on the most impactful price peaks. Practically, predicting these volatility peaks has a beneficial impact on the cryptocurrency markets. Methodologically, comparing legacy time-series forecasting models such as Linear Regression, LSTM to novel and complex neural networks based algorithm such as GRU, GARCH, and TARCH on high volatility financial data is a novice methodological contribution to the literature of economic forecasting. The methodology introduced in the study is valuable in proving the viability of application in other markets and other marketing analytics problems. The study's model and results show the study's potential and cause a paradigm shift in the algorithmic trading field.

The study initiated with the introduction to the analysis and summarized the practical research problem of predicting the price of Bitcoin and DogeCoin cryptocurrencies. "Literature review" section presents the literature and the gap in the literature. In "Theoretical discussions, methodology, and data description" section, the methods and the dataset are evaluated and proven to be accepted in the literature. In "Results" section, and the study results are given by numerical and visual explanations. Results for Linear Regression based on Linear Regression, LSTM, Bi-LSTM, GRU, and TARCH models are compared, and the supremacy of the GRU model is demonstrated. In "Discussion and future studies" section, the results and findings of the study are discussed. Finally, in "Conclusion" section, the study's conclusion is given by the study's limitations and ends with future studies that can be done to expand the literature.

Social media is not only a platform where people share their ideas and life but also create hypes and ethical dilemmas and affect our lives in many ways, including impacting the market prices (Petrescu and Krishen 2020b). There is a strong correlation between the cryptocurrency price and when people talk about the currencies. By using Google Trends, it can be seen when people tweet, chat, and search about Bitcoin and other currencies the most or the least. This can be seen with DogeCoin's price spiking after Elon Musk, a well-known and liked entrepreneur, tweeted about DogeCoin, causing it to almost reach one dollar despite initially having a value of 20 cents (Ante 2023). Improving transparency in the data helps with forecasting bias, and using open-sourced social media data helps the forecast quality (Hoyle et al. 2020). Using the importance of the actual price of the coin and the number of people who talk about a specific cryptocurrency can help determine a coefficient in its volatility when using linear regression and other machine-learning models, which would allow for more explainable predictive models (Valluri et al. 2021) and precise price prediction (Sagar 2019). Accuracy is more difficult to determine without accounting for any factors that determine cryptocurrency since it is new to the market compared to stocks, shares, and other types of assets (Ganti 2022). The method used in this study is the GARCH Model where. Among all variants of the GARCH family created, TARCH(1,2) with Bootstrap forecasting method achieved the lowest RMSPE and RMSE on the Validation Set (Ozdemir 2022).

Marketing analytics is also a research field that includes trends in many different areas, including search trends, marketplace purchasing trends, and product marketing trends. These trends have volatility in their nature, and applying similar machine-learning techniques would decrease the adverse risks due to the ambiguity of the future for the marketplaces. The methods used in this study can also be applied to marketing analytics trends and conduct analyses such as customer segmentation. Neural Networks showed their power in customer segmentation proven to be efficient to the extent that they can be used to segment voters and be used in political campaigns (Ward 2018). With the same power, it could be used for customer segmentation. The patterns and root causes of volatility that define cryptocurrency prices found in this study may also help explain patterns in search trends, marketplace purchasing trends, and product marketing trends. The biggest obstacle to cryptocurrencies is gaining the trust of the public. Since government institutions do not back them up, a tweet can change cryptocurrencies' public image and value. The founder and owners of each coin market their currencies similarly to a conventional product. The price trends of cryptocurrencies are directly attached to the marketing campaigns for these products. In this study, we have investigated the predictability of the volatility in some of the most popular cryptocurrencies, such as Bitcoin and DogeCoin, by their trends in open-sourced cryptocurrency market data and Google Search Trends data and proven the predictability of the cryptocurrency prices.

Although the proposed model successfully predicted some of the most common cryptocurrency price volatility, suture studies in the field can be done by advancing the data sources to various social media platforms. The impact of global financial events can be tested against cryptocurrency prices. In addition, numerous cryptocurrencies are using different technologies and business models. Their business models can evaluate these different cryptocurrencies and predict their price volatility. Market trends in the other areas of marketing analytics, such as rare, valuable materials, Brent oil, cambio markets, and stock markets, and their impact on cryptocurrencies can be evaluated in future studies.

References

Alaparthi, S., and M. Mishra. 2021. Bert: A sentiment analysis odyssey. *Journal of Marketing Analytics* 9 (2): 118–126.

Alessandretti, L., A. ElBahrawy, L.M. Aiello, and A. Baronchelli. 2018. Anticipating cryptocurrency prices using machine learning. *Complexity* 2018: 1–16.

Ante, L. 2023. How Elon Musk's twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change* 186: 122112.

Blokdyk, G. 2018. Tensorflow. *5starcooks*.

Cankaya, B., K. Topuz, and A. Glassman. 2023. Business inferences and risk modeling with machine learning; the case of aviation incidents. Proceedings of the 56th Hawaii International Conference on System Sciences. https://scholarspace.manoa.hawaii.edu/items/f8b554da-4482-4c9c-9309-e9e80e72bc0b.

Chatterjee, C.C. 2019. Implementation of RNN, LSTM, and GRU. https://towardsdatascience.com/implementation-of-rnn-lstm-and-gru-a4250bf6c090. Accessed 14 Jan 2023.

Cho, H., and K.K. Korkas. 2022. High-dimensional GARCH process segmentation with an application to value-at-risk. *Econometrics and Statistics* 23: 187–203.

de Almeida, W.M., and C.P. da Veiga. 2022. Does demand forecasting matter to retailing? *Journal of Marketing Analytics* 1–14.

Dey, R., and F.M. Salem. 2017, August. Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.

Dismuke, C., and R. Lindrooth. 2006. Ordinary least squares. *Methods and Designs for Outcomes Research* 93 (1): 93–104.

Eric, D., and D.S. Brown. 2019. Forecasting Time Series Data using Autoregression. Eric D. Brown, D.Sc. Accessed 14 Jan 2023.

Frankenfield, J. 2021. What is dogecoin? In *Investopedia*.

Ganti, A. 2022. What is implied volatility (IV), Investopedia Accessed 14 Jan 2023.

Ghildiyal, P. 2014. Parallel computation: Best practices while testing the speedup. http://pawangh.blogspot.com/. Accessed 14 Jan 2023.

Guizani, S., and I.K. Nafti. 2019. The determinants of bitcoin price volatility: An investigation with ardl model. *Procedia Computer Science* 164: 233–238. https://doi.org/10.1016/j.procs.2019.12.177.

Hansen, J.M., and M.A. Levin. 2016. The effect of apathetic motivation on employees' intentions to use social media for businesses. *Journal of Business Research* 69 (12): 6058–6066.

Hayes A. 2023. Volatility: Meaning In Finance and How it Works with Stocks Investopedia. https://www.investopedia.com/terms/v/volatility.asp. Accessed 14 Jan 2023.

Hoyle, J.A., R. Dingus, and J.H. Wilson. 2020. An exploration of sales forecasting: Sales manager and salesperson perspectives. *Journal of Marketing Analytics* 8 (3): 127–136.

Kaabachi, S., S. Ben Mrad, and T. Barreto. 2022. Reshaping the bank experience for gen z in France. *Journal of Marketing Analytics* 10 (3): 219–231.

Kim, J.-M., C. Jun, and J. Lee. 2021. Forecasting the volatility of the cryptocurrency market by GARCH and stochastic volatility. *Mathematics* 9 (14): 1614. https://doi.org/10.3390/math9141614.

Kim, S.-J., M. Naruse, M. Aono, H. Hori, and T. Akimoto. 2016. Random walk with chaotically driven bias. *Scientific Reports* 6 (1): 1–9.

Lima Ana Lucia. 2022. Bitcoin price prediction using recurrent neural networks and LSTM. Analytics Vidya

Lucey, B.M., S.A. Vigne, L. Yarovaya, and Y. Wang. 2022. The cryptocurrency uncertainty index. *Finance Research Letters* 45: 102147. https://doi.org/10.1016/j.frl.2021.102147.

Madan, I., S. Saluja, and A. Zhao. 2015. Automated bitcoin trading via machine learning algorithms.

Moriyama, T., and M. Kuwano. 2022. Causal inference for contemporaneous effects and its application to tourism product sales data. *Journal of Marketing Analytics* 10 (3): 250–260.

Meyers, R. A. (Ed.). 2009. Encyclopedia of complexity and systems science (Vol. 9). New York: Springer.

Naimy, V., O. Haddad, G. Fernandez-Aviles, and R. ElKhoury. 2021. The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies. *PLoS ONE* 16: e0245904.

Nonejad, N. 2022. Predicting equity premium out-of-sample by conditioning on newspaper-based uncertainty measures: A comparative study. *International Review of Financial Analysis* 83: 102251.

Olah, C. (2015). Understanding lstm networks.

Ozdemir, O. 2022. ARCH-GARCH tutorial with rugarch package. http://users.metu.edu.tr/ozancan/ARCHGARCHTutorial.html. Accessed 29 March 2023.

Pandita. S. 2021. Bitcoin price prediction using linear regression. Hacketdawn. https://medium.com/hackerdawn/bitcoin-price-prediction-using-linear-regression-94e0e5a63c42. Accessed 14 Jan 2023.

Petrescu, M., and J. Gironda. 2019. *Interpris: Intuitive qualitative data analysis*. New York: Springer.

Petrescu, M., and A.S. Krishen. 2020a. *The importance of high-quality data and analytics during the pandemic*. New York: Springer.

Petrescu, M., and A.S. Krishen. 2020b. *The dilemma of social media algorithms and analytics*. New York: Springer.

Sagar, A. 2019. Cryptocurrency price prediction using deep learning, towards data science

Schröer, C., F. Kruse, and J.M. Gómez. 2021. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science* 181: 526–534.

Schuster, M., and K.K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45 (11): 2673–2681.

Singh, A., and A. Kumar. 2021. Designing the marketspace for millennials: Fun, functionality or risk? *Journal of Marketing Analytics* 9 (4): 311–327.

Stock price prediction using artificial neural network. 2021. *International Journal of Innovative Research in Science, Engineering, and Technology*.

Tajdini, S. 2022. The effects of internet search intensity for products on companies' stock returns: a competitive intelligence perspective. *Journal of Marketing Analytics* 1–14.

Valluri, C., S. Raju, and V.H. Patil. 2021. Customer determinants of used auto loan churn: Comparing predictive performance using machine learning techniques. *Journal of Marketing Analytics* 1–18.

Wang, Y. 2022. Volatility spillovers across NFTs news attention and financial markets. *International Review of Financial Analysis* 83: 102313.

Ward, K. 2018. Social networks, the 2016 US presidential election, and Kantian ethics: Applying the categorical imperative to cambridge analytica's behavioral microtargeting. *Journal of Media Ethics* 33 (3): 133–148.

Yaga, D., P. Mell, N. Roby, and K. Scarfone. 2019. Blockchain technology overview. arXiv:1906.11078

Yildirim S. 2020. Cryptocurrency Prediction with LSTM How to predict the trend of currency rates. Towards Data Science. https://towardsdatascience.com/cryptocurrency-prediction-with-lstm-4cc369c43d1b. Accessed 14 Jan 2023.

Yousaf, I., and S. Ali. 2020. The Covid-19 outbreak and high frequency information transmission between major cryptocurrencies: Evidence from the var-dcc-garch approach. *Borsa Istanbul Review* 20: 1–10. https://doi.org/10.1016/j.bir.2020.10.003.

Zbikowski, K. 2016. Application of machine learning algorithms for bitcoin automated trading. *Machine Intelligence and Big Data in Industry*: 161–168.

Zhang, J., and L. Huang. 2018. Loss or gain? the impact of Chinese local celebrity endorser scandal on the global market value of the endorsed brands. *Journal of Marketing Analytics* 6 (1): 27–39.

Samir Poudel is a graduate research student at the Department of Computer Science at Middle Tennessee State University. Completed his bachelor's in computer engineering from Pokhara University, Nepal. He has a research interest in Artificial intelligence and Machine learning.

Rajendra Paudyal is a Ph.D. student at the Department of Computer Science at George Mason University, Fairfax Virginia. He has Masters of Science in Information and Communication Engineering from Tribhuvan University, Nepal. He has more than 10 years of Telecommunication industry experience and his research interest lies in the area of Computer Networks, Network Security, Wireless Communication, Artificial Intelligence, and Machine learning.

Burak Cankaya has a doctoral degree from Lamar University in Industrial Engineering, an M.S. degree from the University of Houston in Technology Project Management, and a B.S. degree from Karadeniz Technical University in Mechanical Engineering. Before joining ERAU as an Assistant Professor, Dr. Cankaya was a Sr. Operations Research Analyst at NetJets, an industry-leading Berkshire-Hathaway-owned Private Jet company. He has previous industry and research experience in the Transportation and Energy industries. His passion is Analytics, Operations Research, and Data Science. His presentation in INFORMS BUAN 2022 Conference received the 2nd best poster presentation award. He received grants from Google, DOT, and ERAU. ERAU COB selected as the 2022 College of Business Faculty of the Year award winner and 2020 COB Faculty of the Year in Research Award. He built the majority of the B.S. in Business Analytics degree, and he believes that conducting research in analytics and data science and teaching the tech skills to lead the 4th industrial revolution is his best contribution to society. In his spare time, Dr. Cankaya enjoys hiking, camping, and wildlife photography.

Naomi Sterlingsdottir is a recent graduate of Embry-Riddle Aeronautical University (ERAU) where she received her degree in computational mathematics with a focus in data science. She is now a graduate student attending Florida Institute of Technology pursuing a master's degree in applied mathematics as she works full-time for Northrop Grumman as a systems engineer. During her undergraduate studies, she was involved in a variety of research projects ranging from medicine, finance, and material science. She was also a receipt of the "Outstanding Achievement in Mathematics Scholarship" during her third year at ERAU. Naomi also presented her many of her findings in numerous conferences including SIAM, NCUR and FUCR over the course of 2 ½ years. For her graduate studies she aims to have her research focus on electromagnetism and material science as her senior undergraduate researcher project involved the use machine learning to quantify STEM micrographs for the Pacific Northwest National Laboratory.

Marissa Murphy has completed his Bachelor's from Embry Riddle Aeronautical University under the Department of Mathematics and Data Science. Marissa's research interest lies in the area of Artificial Intelligence and Machine Learning.

Shital Pandey is a proficient data scientist pursuing an M.S. in Data Science at Embry-Riddle Aeronautical University. Shital's academic journey includes an M.S. in Computer Science from Lamar University and a B.E. in Electronics and Communication Engineering from Tribhuvan University, Nepal. His passion for research and innovation is evident in projects like the prediction of Alzheimer's disease progression using General Adversarial Networks and Vision Transformers, and the development of explainable AI systems for breast cancer classification. Shital's dedication to advancing the field of data science is further evidenced by their research interests, which encompass areas such as cloud computing, natural language processing, and deep learning algorithms.

Jorge Vargas earned his Ph.D. in Electrical Engineering from Florida International University in 2005. Dr. Vargas has over 15 years of experience in academia and has worked with multimillion-dollar funded projects in the fields of RF/Microwave Engineering, Microelectronics,Radar Sensors for Connected and Autonomous Vehicles, Sensor Fusion, and Machine Learning.Prior to joining MTSU in January 2021, he worked as an Associate Professor in Electrical Engineering at Florida Polytechnic University, where he taught several undergraduate and graduate courses on electrical circuits, electronics, logic design, radio frequency, microprocessor design,and sensors. Also, he had the opportunity to design and develop the Electrical and Computer Engineering curricula for the undergraduate and graduate programs, supported collaborative educational grants for external funding opportunities, provided outreach projects to the community, and served as an academic program coordinator in the Electrical Engineering department. His industry background at IBM involves product and process development of RF/Microwave devices.

Khem Poudel (M'15) earned his B.E. and MS degrees in electronics and communication engineering and information and communication engineering, respectively, from Tribhuwan University in Kathmandu, Nepal, in 2010 and 2013, and his MS degrees in computer science and electrical and computer engineering, both from Middle Tennessee State University in Murfreesboro, Tennessee, in the United States. He earned his Ph.D. in computational science from Middle Tennessee State University in Murfreesboro, Tennessee, in the United States. In 2022, he started working as a faculty member at Middle Tennessee State University. His areas of research interest include Big Data, Deep Learning, and Bio-Medical Signal Processing.