# Performance Evaluation and Analysis of Supervised Machine Learning Algorithms for Bitcoin Cryptocurrency Price Forecast

**Sani Abba** saniabba2004@gmail.com
*Department of Mathematical Sciences,*
*Abubakar Tafawa Balewa University,*
*Bauchi, Nigeria.*

**Souley Boukari** bsouley2001@yahoo.com
*Department of Mathematical Sciences,*
*Abubakar Tafawa Balewa University,*
*Bauchi, Nigeria.*

**Mohammed Ajuji** majuji@gsu.edu.ng
*Department of Computer Science,*
*Gombe State University,*
*Gombe, Nigeria.*

**Amina Nuhu Muhammad** aminanuhumuhammad@gmail.com
*Department of Computer Science,*
*Gombe State University,*
*Gombe, Nigeria.*

**Abstract**

Earlier to the advent of computers and the internet. Transactions such as buying, selling, hiring, and cash transfer are performed physically, hand-to-hand and/or face-to-face using hard-printed currency also known as traditional means. The recent advances in internet and networking technologies have significantly refurbished and improved the methods and limitations of the traditional ways, through cryptocurrency or digital money especially in terms of cost, speed, and access. These technologies which bring people together irrespective of geographical location have fashioned a revolution in trading and transaction processing; online transaction processing and real-time processing. However, like every other pioneering development, this is not without resistance from stakeholders, whom have been using the traditional means for long; its validity and legitimacy have been seriously challenged. In this work, several models leveraged to forecast bitcoin price were Linear Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Elastic Net Regression (EN), Lasso Regression (Lasso) and Ridge regression (RR). The models' accuracies were determined and evaluated using Mean Absolute Error (MAE), Mean Square Error (MSE), and R Square Error (R2). It revealed good performance except for SVM which falls in the negative even after fine-tuning and improved performance. The LR led in performance, then EN, Lasso, and RR. Decision Tree on the other hand present an encouraging and challenging result. Whereas the SVM model presents worst-case prediction accuracy of -22.38%. Therefore, the linear regression model has the best fit for bitcoin price prediction amongst the algorithms tested and evaluated. The study would reduce researchers throughput by presenting firsthand model for price prediction, support vector machine need to be further studied to unravel reasons for its undesirable performance.

**Keywords:** Bitcoin, Cryptocurrency, Machine-Learning, Prediction.

## 1. INTRODUCTION

At the moment government and corporate organizations make use of centralized currencies that are being controlled by the government or other corporate entities (Cocco et al., 2021) such as the central bank or single administrator over the supply of currency by printing and distributing new hard currencies. The decentralized nature of cryptocurrency make it fit to the modern generation (Lucci & Ketel, 2020). Digital money is also known as cryptocurrency such as bitcoin is decentralized (Yılmazer et al., 2021). It has no single administrator. Cryptocurrency utilizes cryptography to secure the processes involved in transactions and the generation of units (Valencia & Gómez-espinosa, 2019). Many other cryptocurrencies are available, for example, Etherum, Litecoin, and so on, but bitcoin is the most popular of them all based on a peer-to-peer network (Livieris et al., 2021). Network nodes are used to verify transactions through cryptography and recorded in a public distributed ledger called a blockchain (Harris & Waggoner, 2019). Which also stores records of several exchanges and timestamp information for bitcoin transactions.

The rate of bitcoin fluctuates (Aljojo et al., 2021) suddenly. Uncertainty has been the major attribute of the future price of bitcoin, which has made it difficult to accurately predict its future price. However, with the recent advances in machine learning techniques; it became very possible to make accurate and reliable predictions. Accurately predicting the price for Bitcoin and other cryptocurrencies is therefore of paramount importance for investors and market players in the cryptocurrency market to make a veracious decision-making process. As a result, yield significant returns for investors and market players in the cryptocurrency sector.

According to (Munim & Shakil, 2019), the cryptocurrency market is amongst the fastest-growing of all the financial markets in the world. Distinct from the traditional markets, such as equities, foreign exchange, and commodities, the cryptocurrency market is considered to have larger volatility and illiquidity. For some time, bitcoin rate prediction has been an active and dynamic research area. Being a volatile and speculative type of money, it is imperative to understand the price behavior of Bitcoin before one could decide whether to invest in it or not. Various methods were used to predict the price of bitcoin including machine learning (Livieris et al., 2021), and deep learning techniques (Ibrahim et al., 2020). Predicting the future has been and will always be on the top list of machine learning algorithms and applications. Bitcoin is selected in this study for the prediction problems because of its popularity and the size or magnitude of its market capitalization. This study is structured as follows: Section 2 explains briefly the theory of methods applied in this study. Section 3 gives details of the datasets obtained and the variables used in this study. Section 4 presents the results obtained and discussion. In section 5, conclusions of the study was drawn.

## 2. LITERATURE REVIEW

Bitcoin simply means a digital currency or a cryptocurrency. The concept of bitcoin was first advocated by Nakamoto (Munim & Shakil, 2019) and (Munim & Shakil, 2019), and it was the first distributed cryptocurrency released as open-source software. It is a digital money that can be used for transactions on the peer-to-peer Bitcoin network devoid of intermediaries (Cohen, 2020). Bitcoin will serve as an alternative means of exchange, as revealed by (Livieris et al., 2021).

The need to predict the price of cryptocurrency especially bitcoin has attracted so much attention in the past few years, especially with the advent of machine learning techniques. There have been a good number of already conducted researches on Bitcoin price prediction using machine learning and time-series analysis (Dutta et al., 2020). Though not all machine learning techniques were explored accordingly. This study will be restricted to only peer-reviewed published journals. A future prediction has been on the top list of machine learning applications. Bitcoin cryptocurrency is selected in this study for the prediction problem because of its popularity and the size or magnitude of its market capitalization. Sentiment Analysis and Machine Learning methods were used in a study by (Valencia & Gómez-espinosa, 2019) to predict the price movement of different cryptocurrencies. The study compared the utilization of neural networks, support vector machines, and random forest while using elements from Twitter and market data as input features. The outcomes showed that it is possible to predict cryptocurrency markets using machine learning and sentiment analysis. Neural Network (NN) outperformed the other models.

The study by (Livieris et al., 2021) revealed that the most popular and promising type of profitable investment today is cryptocurrency. The study by (Munim & Shakil, 2019) uses two methods and explored prognoses of Bitcoin price using the Autoregressive Integrated Moving Average (ARIMA) and Neural Network Auto-regression (NNAR) models to forecast next-day Bitcoin price both with and without re-estimation. It revealed that despite the complexity of (NNAR), their finding revealed that it demonstrates ARIMA's enduring power of volatile Bitcoin price prediction.

In their study (Munim & Shakil, 2019) compared different approaches using the Root Mean Squared Error (RMSE) evaluation method. The results showed that the gated recurring unit (GRU) model with recurrent dropout performs better than popular existing models. They also showed that simple trading strategies, when implemented with their proposed GRU model and with proper learning, can lead to economic advancement.

For example, a study by (Reddy & Sriramya, 2020) analyzed the predictability of the bitcoin market by testing various machine learning models and discovered that recurrent neural network and gradient boosting algorithms performed prediction pretty well. Random classifiers are given the worst output. Deep learning methods were used (Lamothe-fern et al., 2020) for modeling the price of bitcoin, a new model created to make the prediction. The Deep Recurrent Convolutional Neural Network (DRCNN) model obtained the highest levels of precision of about 92.61-95.27%. Which according to them was higher than previous findings.

(Sujatha et al., 2021) in their study analyzed bitcoin trends with three training algorithms using a Bayesian Regularized Neural Network (BRNN). It revealed that the BRNN supersede other algorithms in performing prediction. Time series of bitcoin data was analyzed with blockchain information. The results showed that BNN model learned effectively with the selected features and produce a high performing model (Jang & Lee, 2018).

The study was conducted to detect bitcoin market manipulations based on forecasting anomalies using both machine learning and statistical forecasting tools. It presents a comprehensive study in terms of presenting scientific and realistic solutions for detection of manipulations in Bitcoin market (Akba et al., 2021). (Jay et al., 2020) Used stochastic neural network model to predict the price of cryptocurrencies. It showed that the deterministic versions of neural network were outperformed by the stochastic version of the Neural Network. In some studies found in the literature, traders must be longing and waiting for forecasted results for every hour; which may be engaging and time-consuming. Prediction for the next day is a little better compared to the next hour prediction but having a prediction for a longer period, say a week or a month; even a year could be feasible. Most of the work found in the literature is computationally expensive and requires highly skilled and experienced forecasters to obtain an accurate and reliable result. Compared with some simple and fundamental machine learning prediction models, the theoretical model and structural relationships are not distinct. In this study, a dataset of the period of about ten years will be used, which is the longest interval compared with the existing studies.

### a) Rise of Bitcoin

The concept of bitcoin was first advocated by Nakamoto (Reddy & Sriramya, 2020), and it was the first distributed cryptocurrency released as open-source software. It is digital money that can be used for transactions on the peer-to-peer Bitcoin network devoid of intermediaries (Jiang, 2020). Cryptocurrencies are distributed digital currency as opposed to the formal existing centralized traditional money (Lamothe-fern et al., 2020). Cryptocurrency is a network-based digital exchange medium, where the records are secured using strong cryptographic algorithms such as Secure Hash Algorithm 2 (SHA-2) and Message Digest 5 (MD5). It uses blockchain technology to make the transactions secure, transparent, traceable, and immutable. The world's leading cryptocurrency today is bitcoin (Jaquart et al., 2021) and the blockchain is the technology that supports it.

### b) Price and Volatility

The price of Bitcoin is volatile and fluctuate easily (Dimpfl & Odelli, 2020), historically according to a study (Lamothe-fern et al., 2020) it crashes at the magnitude of more than 15% sharp drop within 3 weeks in less than ten years; from 2011 to 2019 and observe a total of 51 crashes in less than a decade. Its crash duration and the time gap between crashes can be very short while more than half of the crashes saw over a quarter of a price drop (Dutta et al., 2020). The number of crashes increases as the Bitcoin long-timescale bubble grows, and a succession of short timescale crashes can be triggered as the long timescale bubble matures. Since the first release of Bitcoin in 2009, the price of Bitcoin has suffered through a series of roller-coaster ups and downs. Table 1 below summarizes the previously used machine learning methods for constructing bitcoin price prediction models. This study will explore some of the previously used algorithms for bitcoin price prediction and those that are not used for the same prediction.

**TABLE 1:** Summary of machine learning algorithms tested in bitcoin price prediction literature.

| Study | Methods | Drawbacks/Limitations |
|---|---|---|
| (Jang & Lee, 2018) | LR, BNN, SVR | Reasonable parameters selection is very stimulating which may be considered as the reason that it is not very satisfactory. |
| (Valencia & Gómez-espinosa, 2019) | NN, SVM, RF | Slow training largely due to dataset size. A lot of noise in the dataset reduces model performance i.e., target classes might easily get overlapped. A slight modification in the dataset could greatly affect the prediction. (It is highly sensitive). |
| (Munim & Shakil, 2019) | ARIMA, NNAR | Features selected for model applications may be highly biased.<br>Feature scaling is necessary to avoid inaccurate forecasts. |
| (Ibrahim et al., 2020) | VAR, BVAR | It assumes a normal distribution of features/predictors.<br>It is not good sometimes for few variables categories. |
| (Dutta et al., 2020) | GRU, LSTM | A lot of preprocessing is required to reduce over fitting.<br>The curse of dimensionality affects model performance. |
| (Lamothe-fern et al., 2020) | KNN, RR, PR, LR, RF | Performance may be affected by a larger dataset. High dataset dimensionality also affects its performance<br>Feature scaling is necessary. |
| (Jay et al., 2020) | MLP, LSTM | Redundancy is possible in high dimensions because the number of aggregate parameters can grow to very high. |
| (Jaquart et al., 2021) | GRU, LSTM, FNN, LR, GBC, RF | Prone to over fitting especially when a high dimensional dataset is used.<br>Difficult to apply non-linear correlation amongst the forecaster and result. |
| (Chevallier et al., 2021) | RF, RR, ANN, SVM, KNN, AB | Model performance may be bias.<br>Too much noise in the dataset may result in decrease algorithm performance.<br>Low model interpretability. |
| (Livieris et al., 2021) | DRCNN, DNDT, DSVR, SA | It requires strong assumptions on the distribution of the features. Independence assumption is needed in the predictors. |
| (Sujatha et al., 2021) | NN, LM, BR, SCG, BRNN | It requires generally a number (n) cycles to reach the minimum. Basically few cycles is needed. |
| (Akba et al., 2021) | ARIMA Variation | Features selected for model applications may be highly biased which could result in over fitting. |

**GRU=Gated Recurrent Unit, LSTM=Long Short Term Memory, FNN=Feed forward Neural Network, LR=Linear Regression, GBC=Gradient Boost Classifier, RF=Random Forest, RR=Ridge Regression, ANN=Artificial Neural Network, SVM=Support Vector Machine, KNN=K-Nearest Neighbor, AB=AdaBoost. DRCNN=Deep Recurrent Convolutional Neural Network, DNDT=Deep Neural Decision Tree, ARIMA=Autoregressive Integrated Moving Average, NNAR=Neural Network Autoregression, VAR=Vector Autoregression, BVAR=Bayesian Vector Autoregression, LM= Levenberg-Marquard, BR=Bayesian Regularized, SCG=Scaled Conjugate Gradient, BRNN=Bayesian Regularized Neural Network, MLP=Multilayer Perceptron

### c) Taxonomy of Machine Learning Algorithms

It is a division of Artificial Intelligence (AI) techniques that enables computer systems to learn from previous drill and improve their behavior for a given task (Reddy & Sriramya, 2020). Machine learning techniques are twofold: supervised and unsupervised machine learning techniques. The supervised technique utilizes the training data to predict hidden future activities. Unsupervised techniques are leveraged to recognize hidden data models provided without providing training data. Machine learning methods have become an essential tool for future prediction. Machine learning could accurately predict the future. Several machine learning algorithms, especially supervised learning has been used by different researchers are stated below. The taxonomy of machine learning models is presented in figure 1 below.

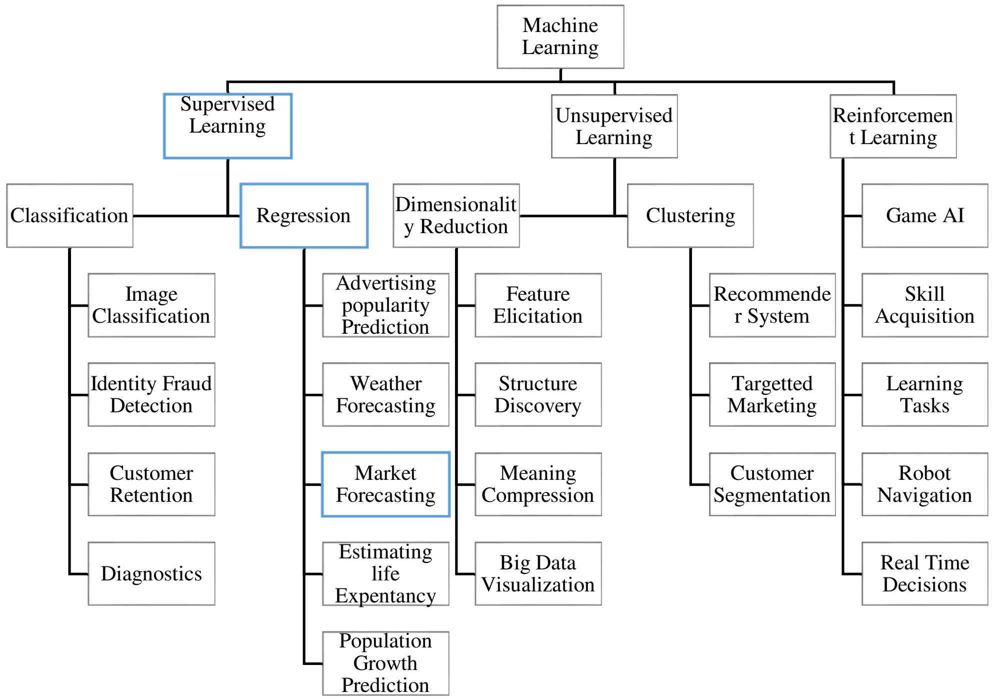


**FIGURE 1:** Taxonomy of Machine Learning Algorithms.

### d) Dataset

It is a collection of related data. Datasets are prerequisite for building a flexible and accurate model. Training and testing are conducted using dataset. Dataset can be generated primarily or secondary and artificial which are generated by a computer. Table 2 is a dataset table that contain the list, sources of data and the range of data used in the literature review. The dataset for bitcoin price history was obtained from the popular machine learning repository kaggle.com, yahoo finance and coin market cap. It contain a value at Open, High, Low, Close, the volume traded in Bitcoin and in USD, the weighted price, and the date. This research focuses on predicting Bitcoin prices in the future by using the price of past years. For evaluation purposes, the bitcoin price dataset is divided into training and validation sets. Methodically, the training set comprised 70% of the dataset while the remaining is used for validation. It is worth noting that this is the longest period dataset covered as regards the previous study; it includes data for the period of the recent pandemic outbreak that almost put everything on hold globally, Coronavirus also known as Covid-19 has affected the normal methods of operations in many aspects including bitcoin price behavior. It has reportedly been attributed to considerable volatility and deviations from the regular bitcoin value as well as operational disturbances (Sujatha et al., 2021).

**TABLE 2:** Summary of datasets used in bitcoin price prediction literature.

| Study | Range | Cryptocurrency | Dataset Source |
|---|---|---|---|
| (Jang & Lee, 2018) | September 2011 - August 2017 | Bitcoin (BTC) | bitcoincharts.com/markets/ |
| (Valencia & Gómez-espinosa, 2019) | 16 February 2018 – 21 April 2018 | Bitcoin (BTC), Etherium (ETH), Ripple (XRP), Litecoin (LTE) | https://cryptocompare.com, Twitter data |
| (Munim & Shakil, 2019) | 1 January, 2012 – 4 October, 2018 | Bitcoin (BTC) | Quandl |
| (Dimpfl & Odelli, 2020) | 7 March 2017 – 7 January 2019 | Bitcoin (BTC) | Bitfinex and Kraken |
| (Ibrahim et al., 2020) | 10 December, 2013 – 30 September, 2016 | Bitcoin (BTC) | https://blockchain.com |

| | | | |
|---|---|---|---|
| (Cohen, 2020) | Early 2012 – March, 2020 | Bitcoin (BTC) | Nill |
| (Dutta et al., 2020) | 1 January, 2010 – 30 June, 2019 | Bitcoin (BTC) | Bitcoincharts.com |
| (Lamothe-fern et al., 2020) | 2011 – 2019 | Bitcoin (BTC) | IMF's IFS, World Bank, Fred Sant Louis, Google Trends, Quandl and Blockchain.info |
| (Jay et al., 2020) | Mid 2017 to end 2019 | Bitcoin (BTC) | bitinfocharts |
| (Jaquart et al., 2021) | 4 March 2019 – 10 December 2019 | Bitcoin (BTC) | Bloomberg, Twitter, and Blockchain.com |
| (Livieris et al., 2021) | 1 January, 2017 – 31 October, 2020 | Bitcoin (BTC), Etherium (ETH), Ripple (XRP) | https://coinmarketcap.com |
| (Akba et al., 2021) | 2010 to 2019 | Bitcoin (BTC) | coinmarketcap |
| (Sujatha et al., 2021) | 2019 | Bitcoin (BTC) | https://santiment.net |

From the above table, it can be deduced that bitcoin is the most popular amongst the available cryptocurrency; it has the highest market capitalization, and therefore, accurate forecasting might increase traders' returns and confidence.

## 3. METHODOLOGY

This study work focused to evaluate and analyze performance of some supervised machine learning models, as a result, deductive research techniques is embraced. Several algorithms are in place that could potentially make an accurate prediction. Nevertheless, randomly picking one of the algorithms without experimenting and assume it is the best for the forecast is not scientific. Consequently, this research is concerned with the most commonly used algorithms, to have the best prediction. All the considered algorithms are fairly run and compared to ensure that each algorithm is evaluated in the same way on the same data (Chevallier et al., 2021). This was achieved by forcing each algorithm to be evaluated on a consistent test harness. The algorithms include:

### a) Linear Regression

It was developed in the field of statistics and borrowed by machine learning algorithms, it is now both a machine learning and statistical algorithm. It is a model for understanding the relationship between input and output numerical variables (Jaquart et al., 2021). Linear regression is twofold, that is; simple linear regression which is a statistical method that enables users to summarize and study relationships between two continuous or quantitative variables.

$$y = B0 + B1 * x \qquad (1)$$

Where $y$ is the dependent variable (predicted result), $x$ is an independent variable, $B1$ is the slope of a line and $B0$ is the $y$-intercept.

Linear regression is a linear model wherein a model assumes a linear relationship between the input variables (x) and the single output variable (y). When there is a single input variable (x), the method is called a simple linear regression, otherwise, the procedure is referred to as multiple linear regression.

### b) Ridge Regression

It is a model tuning method that is used to analyze any data that suffers from multicollinearity. Ridge regression performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values (Chevallier et al., 2021). Ridge regression is an extension of linear regression where the loss function is modified to minimize the complexity of the model measured as the sum squared value of the coefficient values.

The cost function for ridge regression:

$$MIn\left(\left||Y - X(theta)\right||^2 + \lambda\left||theta\right||^2\right) \tag{2}$$

Lambda is the penalty term. $\lambda$ given here is denoted by an alpha parameter in the ridge function. So, by changing the values of alpha, we are controlling the penalty term. The higher the values of alpha, the bigger is the penalty and therefore the magnitude of coefficients is reduced.

- It shrinks the parameters. Therefore, it is used to prevent multicollinearity
- It reduces the model complexity by coefficient shrinkage

In simple terms,

Ridge $R = loss + \lambda||w||\hat{}2$

Here, $\lambda$ is constant,

$||w||2 = \ w12 + w22 + w32$ ... Here, $w$ is a vector of coefficient.

So ridge regression puts limitations on the coefficients$(\mathrm{w})$. The penalty term lambda regularizes the coefficients with the end goal that if the coefficients take expansive qualities the enhancement work is penalized. Along these lines, ridge relapse shrivels the coefficients and it helps the model unpredictability and multicollinearity.

The assumptions of ridge regression are the same as that of linear regression: linearity, constant variance, and independence. However, as ridge regression does not provide confidence limits, the distribution of errors to be normal need not be assumed. Ridge regression can also be simply defined as the variation of linear regression.

**c) LASSO Linear Regression**

LASSO stands for Least Absolute Selection Shrinkage Operator wherein shrinkage is defined as a constraint on parameters (Reddy & Sriramya, 2020). The goal of LASSO regression is to obtain the subset of predictors that minimize prediction error for a quantitative response variable. The algorithm operates by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero (Chevallier et al., 2021).

Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients variables are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical, or both. This lasso regression analysis is a shrinkage and variable selection method and it helps analysts to determine which of the predictors are most important.

**d) Elastic Net Regression**

Elastic Net is a convex combination of Ridge and Lasso. It emerged as a result of critique on the Least Absolute Selection Shrinkage Operator (LASSO), whose variable selection can be too dependent on data and thus unstable (Chevallier et al., 2021). The solution is to combine the penalties of ridge regression and lasso to get the best of both. Elastic Net aims at minimizing the following loss function:

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i-1}^{n}(y_i - x_i\hat{\beta})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j-1}^{m}\hat{\beta}^2{}_j + \alpha\sum_{j-1}^{m}|\hat{\beta}_j|\right), \tag{3}$$

Where $\alpha$ is the mixing parameter between ridge $(\alpha = 0)$ and lasso $(\alpha = 1)$.

### e) K-Nearest Neighbors

The model representation for KNN is the entire dataset. KNN makes a prediction using the training dataset directly. Prediction is made for a new data point by searching through the entire training set for the K most similar instances and summarizing the output variable for those K instances (Reddy & Sriramya, 2020). For this regression study, this might be the mean output variable. KNN can be used for both regression and classification problems.

To determine which of the K instances in the training dataset are most similar to the new input a distance measure is used. The most popular distance measure is Euclidean distance, and it is used here because it deals with real-valued input variables. Euclidean distance is calculated as the squared differences between points a and b across all input attributes $i$.

$$Euclideean\ Distance\ (a, b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{4}$$

There are many other distance measures available but Euclidean distance is used because the input variables are similar in type. For regression problems, the KNN prediction is based on the mean and the median of the K-most similar instances.

### f) Classification and Regression Trees

Classification and Regression Trees (CART) also referred to as Decision Tree Algorithms that can be used for both classification and regression problems. The CART model is represented using a binary tree. Each node represents a single input variable $(x)$ and a split point on that variable (Brownlee, 2019). The leaf nodes of the tree contain an output variable $(y)$ which used to make a prediction. Predictions are made by CART by traversing the binary tree given a new record. The tree is learned using a greedy algorithm on the training data to pick splits in the tree. Stopping criteria define how much a tree learns and pruning can be used to improve generalization on a learned tree.

### g) Support Vector Machine (SVM)

SVM can best be explained in practice by using a hypothetical classifier called Maximal-Margin Classifier. The numerical input variables $(x)$ in the data (the columns) form an n-dimensional space (Ibrahim et al., 2020) and (Cohen, 2020). A line that splits the input variable space is known as a hyperplane. The hyperplane is selected in SVM to best separate the points in the input variable space by their class, either class 0 or class 1. In n-dimension say two, you can visualize this as a line and assume that all our input points can be separated by this line. For example:

$$B0 + (B1 * X1) + (B2 * X2) = 0 \tag{5}$$

Where the coefficients (B1 and B2) determine the slope of the line and the intercept,

(B0) is found by the learning algorithm, and X1 and X2 are the two input variables. You can make classifications using this line. By plugging in input values into the line equation, you can calculate whether a new point is above or below the line.

Real in practice can be messy and cannot be separated accurately with a hyperplane. The constraint of maximizing the margin of the line that separates the classes must be relaxed. This is often called the soft margin classifier. This change allows some points in the training data to violate the separating line. An additional set of coefficients are introduced that gives the margin wiggle room in each dimension. These coefficients are sometimes called slack variables. This increases the complexity of the model as there are more parameters for the model to fit the data to provide this complexity.

It is characterized by the usage of kernels, the sparseness of the solution, and the capacity control gained by acting on the margin, or several support vectors, etc. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space. SVM algorithm operates natively on numeric attributes, as a result, it uses a z-score normalization on numeric attributes. In regression, Support Vector Machines algorithms use epsilon-insensitivity loss function to solve regression problems.

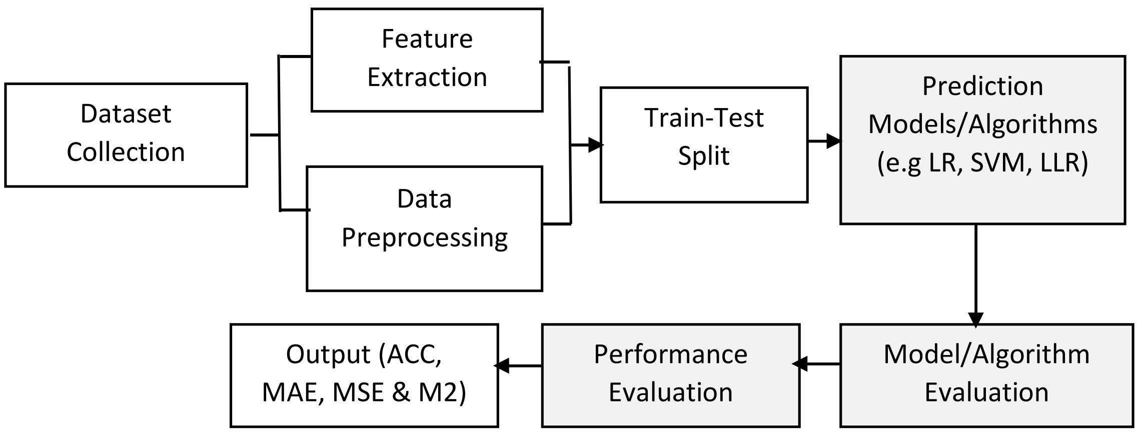


**FIGURE 2:** Framework of the proposed predictive machine learning models.

### 3.1. Data Pre-processing

This step is essential to enhance the data quality to promote the extraction of significant insights from the data. To make sure that too much noise is minimized in the data to avoid over-fitting and/or under-fitting the model. Improve the computational efficiency and accuracy of the model performance. To make the data more relevant for the machine learning forecasting models.

### 3.2. Feature Extraction

Feature extraction will be used to reduce dataset over fitting, improve prediction accuracy and reduce model training time. The dataset features that would be used to train the machine learning models have a great influence on the performance of the algorithm to be achieved. Irrelevant, inappropriate or partially relevant features can negatively influence model performance. Hence, feature selection will be performed on the data to automatically select those features in the dataset that contribute most to the prediction variable or output we are interested in. Having unrelated features in the data can decrease the accuracy of the models, especially linear algorithms like linear and logistic regression (Jang & Lee, 2018).

### 3.3. Performance Evaluation

The generalization of the training datasets is the main goal of building a prediction model using machine learning techniques (Cohen, 2020). Machine learning models should be able to perform pretty well on real data. Training data is used to train whereas testing data is used to test machine learning classifiers. To better understand which of the models outperformed others, we used evaluation metrics and compare the results with what is obtained in the previous studies.

### 3.4. Choice of Evaluation Metrics

*a. Mean Absolute Error*

The Mean Absolute Error or MAE for short is the sum of the absolute differences between predictions and actual values. It gives an idea of how wrong the predictions existed. The measure gives an idea of the magnitude of the error, but no idea of the direction, for example, over prediction or under prediction.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{6}$$

*b. Mean Square Error*

The Mean Square Error (MSE) is also called Root Mean Square Error (RMSE) is much like the mean absolute error in that it provides a gross idea of the magnitude of error. Taking the square root of the mean squared error converts the units back to the original units of the output variable and can be meaningful for description and presentation. The formula below provides a demonstration of calculating mean squared error.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \qquad (7)$$

*c. R Square*

This metric is usually used for regression models. It provides an indication of the goodness of fit of a set of predictions to the actual values. In statistics, this measure is called the coefficient of determination. It is a value between zero (0) and one (1) for no-fit and perfect fit respectively. The mean $R^2$ for a set of predictions is calculated thus.

$$R2 = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \bar{y}_i)^2} \qquad (8)$$

Where the meaning of the parameters in equations (6) to (8) is as follows:

$x_i$ = Actual value
$y_i$ = Predicted value
$n$ = Number of data points/rows
$\bar{y}_i$ = Mean of all actual values

## 4. EMPIRICAL RESULT

Table 3-5 present result and Figures 3-9 showed the level of accuracy, the mean absolute error (MAE), mean square error (MSE) and the r-square error (R2). In all models, the level of accuracy always exceeds 95.69% except for SVM model which has a negative result (-72.1198). For its part, the RMSE and MAPE levels are adequate. The model with the highest accuracy is that of Linear Regression (LR) with 99.99%, followed by the model of Elastic Net, LASSO and Ridge Regression methods with 99.95%. K-Nearest Neighbor (KNN) has the accuracy of 95.69%. Taken together, these results provide a level of accuracy far superior to that of previous studies. Thus, in the work of Ji and co-workers (Lamothe-fern et al., 2020), an accuracy of around 97.34% is revealed; in the case of (Jang & Lee, 2018), it is not close to this accuracy; and in the study conducted by (Valencia & Gómez-espinosa, 2019), it approaches 72%. Therefore, our model have enhanced solution as regards the existing studies.

**TABLE 3:** Results of Accuracy Evaluation on CoinMarketCap Dataset.

| MODEL | ACC.(%) | MAE | MSE | R2 |
|---|---|---|---|---|
| LR | 99.9999 | 0.0000 | 0.0000 | 1.0000 |
| KNN | 95.6889 | 24.8620 | 1045.3151 | 0.9969 |
| SVM | -72.1198 | 121.4453 | 22006.2248 | -0.3706 |
| DT | 99.4975 | 14.6347 | 436.5837 | 0.9995 |
| EN | 99.9560 | 8.7290 | 139.6206 | 0.9999 |
| LASSO | 99.9560 | 8.7290 | 139.6204 | 0.9999 |
| RR | 99.9560 | 8.7290 | 139.6204 | 0.9999 |

**TABLE 4:** Results of Accuracy Evaluation on Yahoo Finance Dataset.

| MODEL | ACC.(%) | MAE | MSE | R2 |
|---|---|---|---|---|
| LR | 99.9999 | 0.0000 | 0.0000 | 1.0000 |
| KNN | 98.7840 | 16.5478 | 662.0565 | 0.9973 |
| SVM | -6.3609 | 86.9643 | 13558.3816 | -0.1230 |
| DT | 99.8639 | 5.9376 | 114.1202 | 0.9999 |
| EN | 99.9714 | 8.5000 | 181.0574 | 0.9998 |
| LASSO | 99.9714 | 8.5001 | 181.0578 | 0.9998 |
| RR | 99.9714 | 8.5001 | 181.0578 | 0.9998 |

**Table 5:** Results of Accuracy Evaluation on Kaggle Dataset.

| MODEL | ACC.(%) | MAE | MSE | R2 |
|---|---|---|---|---|
| LR | 99.9652 | 8.9194 | 213.4400 | 0.9997 |
| KNN | 99.2459 | 14.6247 | 537.4759 | 0.9980 |
| SVM | -22.3791 | 80.7690 | 12722.0353 | -0.1323 |
| DT | 99.8412 | 12.4458 | 473.4499 | 0.9984 |
| EN | 99.9561 | 9.2464 | 227.6392 | 0.9996 |
| LASSO | 99.9561 | 9.2464 | 227.6394 | 0.9996 |
| RR | 99.9561 | 9.2464 | 227.6394 | 0.9996 |

### 4.1 Models Accuracy Comparison

The estimated accuracy of all models developed for this study were compared to check and determine which of the models has the best fit performance accuracy. It is observed at the end of the experiments, as can be seen from the figure below that most of the models performed pretty good except Support Vector Machine (SVM) which falls in the negative even after algorithms fine-tuning and improved performance. The Linear Regression algorithm which is used for both classification and regression prediction problems was leading in performance with approximately 99.97% percent followed closely by three other algorithms EN, Lasso, and RR with 99.96% accuracies. Decision Tree on the other hand has 99.84% accuracy. Whereas the SVM model presents worst-case prediction accuracy of -22.38%. Therefore, the linear regression model has the best fit for bitcoin price prediction in this experiment.

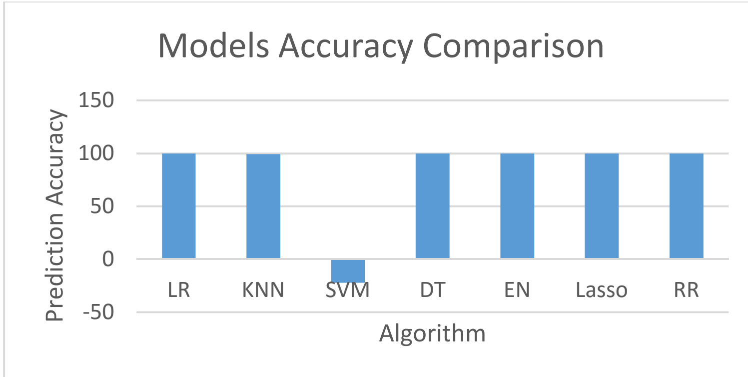


**FIGURE 10:** Algorithms Accuracy Performance.

### 4.2 Models Error Rate Comparison

The Figures 11, 12 and 13 presented the level of errors found in the models prediction performance based on the evaluation metric used for the study. Which include the mean absolute error, root mean square error and r square error. In all the metrics applied, only the support vector machine model which presented the worst case prediction performance scenario therefore it encompasses more errors in its performance compared to the rest of the models whereas linear regression, ridge regression and lasso regression gave good results as shown below.

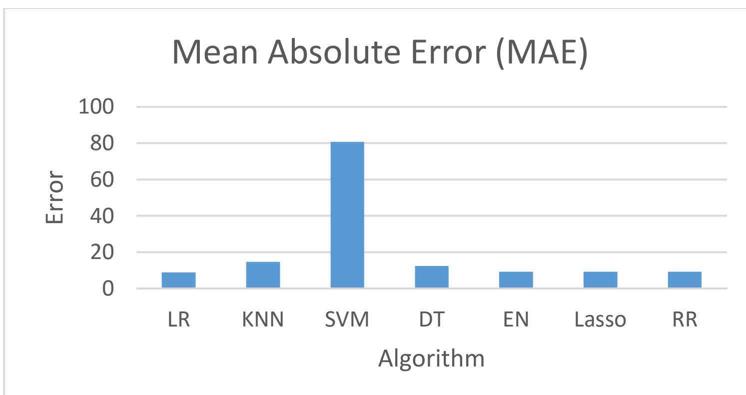


**FIGURE 11:** Accuracy Evaluation MAE.

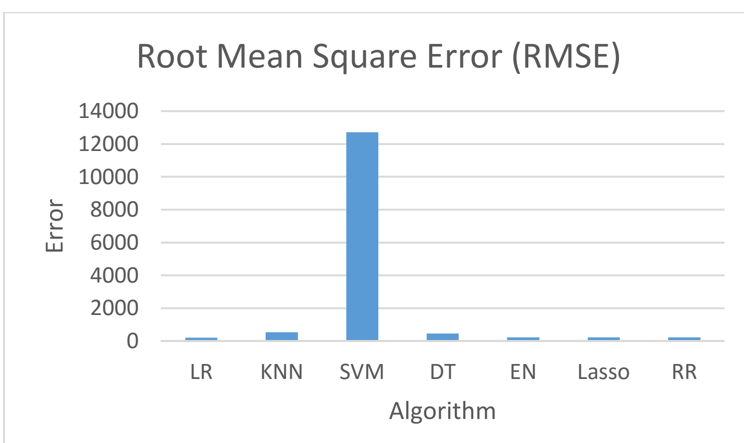


**FIGURE 12:** Accuracy Evaluation MSE.

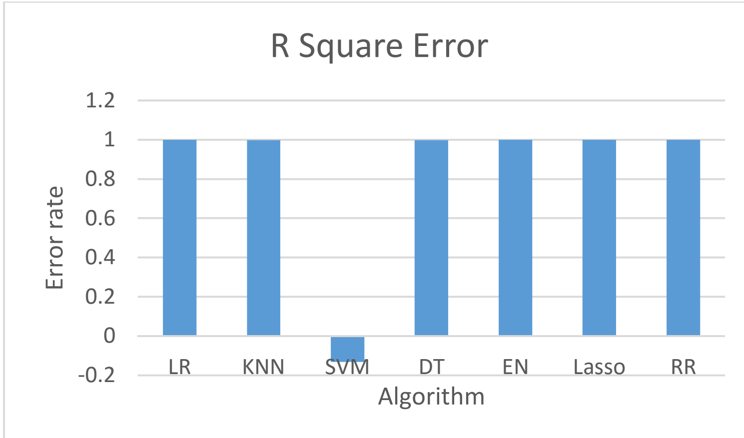


**FIGURE 13:** Accuracy Evaluation R2.

MAE was also used to measure the accuracy of prediction outcome by comparing the forecasted values to the actual value. In other words, it give us the distance between the forecast values to the true or actual values. From figure 11, we can understand that the forecast distance of LR from the actual value is 8.919409 approximately 9. EN 9.246402, Lasso and RR 9.246409. DT and KNN has 12.4458 and 14.62473 respectively. SVM is worst with 80.76897 distance from the true value. The Mean square error (MSE) is the average of the square of the errors. The smaller the value or number the smaller the error. In this case, the error means the difference between the actual values and the predicted ones. From figure 12 above, we can see the level of errors of all the models. The LR model has the least error rate, therefore, it could be classified as the best. EN second while Lasso and RR third. DT and KNN performance was not encouraging but SVM present exceedingly worst-case errors.

Here, in figure 13, R2 present the difference between the actual and predicted values squared so that negative and positive values do not cancel out each other. LR have close to a perfect correlation with approximate of 0.999681. EN, Lasso and RR display an equal but good result of 0.999637. DT and KNN have R2 of 0.998432 and 0.997979 respectively. SVM has a negative root mean squared error of -0.13231 which is not encouraging.

## 5. CONCLUSION

Cryptocurrency is a digital money; it has been in use for a quite sometimes in some parts of the world for the exchange of goods and services. Its popularity and acceptability keep increasing day-by-day by leap and bound; with time it will be generally accepted globally. In this study, we applied several machine learning models to predict bitcoin future price so that people could possibly believe and invest in it. The models revealed different degrees of accuracies scores and were evaluated using popular regression metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), and R Square Error (R2). All models have demonstrated a good performance except for SVM which falls in the negative even after fine-tuning and improved performance. The Linear Regression led in performance, followed closely by EN, Lasso, and RR. Decision Tree on the other hand present an encouraging and challenging result. Whereas the SVM model presented a worst-case prediction accuracy of -22.38%. Therefore, the linear regression model has the best fit for bitcoin price prediction in this experiment. The content derived from this study has the potentials to add value to the cryptocurrency market and public opinion regarding bitcoin and cryptocurrency as a whole. It also add value to the already established literature inform of published research articles in reputable international journals. Lastly, we have a firsthand information about the best performing model in terms of price prediction.

## 6. FURTHER RESEARCH

This work was conducted on a system using a microprocessor which is central processing unit, it could also be conducted on a high speed processor like GPU, AWS and so on to check the performance of the machine these learning classifiers. Support vector machine model in this work has presented a worst case scenario compared to other models, consequently, further studies may be conducted to uncover hidden features and why its prediction accuracy is negative and what is the possible way of improving its performance.

## 7. REFERENCES

Akba, F., Medeni, I. T., Guzel, M. S., & Askerzade, I. (2021). *Manipulator Detection in Cryptocurrency Markets Based on Forecasting Anomalies. IEEE Access*, *9*, 108819–108831. https://doi.org/10.1109/ACCESS.2021.3101528

Aljojo, N., Alshutayri, A., Aldhahri, E., Almandeel, S., & Zainol, A. (2021). A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of timestamp influence on bitcoin value. IEEE Access, 9, 148611–148624. https://doi.org/10.1109/ACCESS.2021.3124629

Balkantime.Press (2021) https://balkantimes.press/en/artificial-intelligence-5-a-taxonomy-of-machine-learning-and-deep-learning-algorithms/ Retrieved August 5, 2021, from https://balkantimes.press/en/artificial-intelligence-5-a-taxonomy-of-machine-learning-and-deep-learning-algorithms/

Brownlee, J. (2019). *Master Machine Learning Algorithms. Discover How they Work and Implement Them from Scratch*. Edition, v1.1 http://MachineLearningMastery.com

Chevallier, J., Guégan, D., & Goutte, S. (2021). *Is It Possible to Forecast the Price of Bitcoin ?* Forecasting J., MDPI pp. 377–420, 2021. doi: 10.3390/forecast3020024.

Cocco, L., Tonelli, R., & Marchesi, M. (2021). Predictions of bitcoin prices through machine learning based frameworks. PeerJ Computer Science, 7, 1–23. https://doi.org/10.7717/PEERJ-CS.413

Cohen, G. (2020). *Forecasting Bitcoin Trends Using Algorithmic Learning Systems Learning Systems,* Entropy J., MDPI vol. i, 2020.

Dimpfl, T., & Odelli, S. (2020). *Bitcoin Price Risk — A Durations Perspective*. J. Risk Financial Manag., 13, 157; doi:10.3390/jrfm13070157.

Dutta, A., Kumar, S., & Basu, M. (2020). *A Gated Recurrent Unit Approach to Bitcoin Price Prediction*. J. Risk Financial Manag., MDPI vol. 13, 23; doi:10.3390/jrfm13020023.

Harris, J. D., & Waggoner, B. (2019). Decentralized and collaborative AI on blockchain. Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019, 2, 368–375. https://doi.org/10.1109/Blockchain.2019.00057

Hu, Z., Zhao, Y., & Khushi, M. (2021). A Survey of Forex and Stock Price Prediction Using Deep Learning. 1–34.

Ibrahim, A., Kashef, R., Li, M., Valencia, E., & Huang, E. (n.d.). *Bitcoin Network Mechanics : Forecasting the BTC Closing Price Using Vector Auto-Regression Models Based on Endogenous and Exogenous Feature Variables*. J. Risk Financial Manag., MDPI, 13, 189; doi:10.3390/jrfm13090189.

Jagannath, N., & Barbulescu, T. (2021). A Self-Adaptive Deep Learning-Based Algorithm for Predictive Analysis of Bitcoin Price. 9. https://doi.org/10.1109/ACCESS.2021.3061002

Jaquart, P., Dann, D., & Weinhardt, C. (2021). ScienceDirect Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, *7*, 45–66. https://doi.org/10.1016/j.jfds.2021.03.001

Jang, H., & Lee, J. (2018). An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*, *6*, 5427–5437. https://doi.org/10.1109/ACCESS.2017.2779181

Jiang, X. (2020). *Bitcoin Price Prediction Based on Deep Learning Methods*. 132–139. https://doi.org/10.4236/jmf.2020.101009

Lamothe-fern, P., Alaminos, D., & Lamothe-l, P. (2020). *Deep Learning Methods for Modeling Bitcoin Price*. 1–13. Mathematics J., MDPI, 8, 1245; doi:10.3390/math8081245

Livieris, I. E., Kiriakidou, N., & Stavroyiannis, S. (2021). *electronics An Advanced CNN-LSTM Model for Cryptocurrency Forecasting*. 1–16. Electronics J., MDPI, 10, 287. https://doi.org/10.3390/electronics10030287

Lucci, N., & Ketel, M. (2020). Application of Blockchain and Smart Contracts on the Internet of Things. International Journal of Computer Science and Security, 14(2), 60–71. https://www.cscjournals.org/manuscript/Journals/IJCSS/Volume14/Issue2/IJCSS-1553.pdf

Munim, Z. H., & Shakil, M. H. (2019). *Next-Day Bitcoin Price Forecast*. Journal of Risk and Financial Management, MDPI, 12, 103; doi:10.3390/jrfm12020103

Reddy, L. S., & Sriramya, P. (2020). *A Research On Bitcoin Price Prediction Using Machine Learning Algorithms*. International Journal of Scientific & Technology Research *9*(04), 1600–1604.

Science, D., & Management, P. (2021). *What is CRISP-DM ?* 1–22. Retrieved on August 04, 2021, from https://www.datascience-pm.com/crisp-dm-2/

Sujatha, R., Mareeswari, V., Chatterjee, J. M. O. Y., & Hassanien, A. E. (2021). *A Bayesian Regularized Neural Network for Analyzing Bitcoin Trends*. *IEEE Access*, 37989–38000. *10.1109/ACCESS.2021.3063243*

Valencia, F., & Gómez-espinosa, A. (2019). *Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning*. Entropy J., MDPI, 21, 589; doi:10.3390/e21060589.

*Yılmazer, S., Aybars, A., & Bozkurt, G. (2021). The Relationship Between BITCOIN and Other Financial Instruments : An Examination With VAR Models 1. 31–47.*

https://pandas.pydata.org/ Accessed on 25th August, 2021.

https://numpy.org/ Accessed on 25th August, 2021.

https://seaborn.pydata.org/ Accessed on 25th August, 2021.

https://matplotlib.org/ Accessed on 25th August, 2021.

https://scikit-learn.org/stable/ Accessed on 25th August, 2021.

https://www.python.org/ Accessed on 25th August, 2021.

https://jupyter.org/ Accessed on 25th August, 2021.