***Original Research Article***

# A Comparative Study of Bitcoin's Price ~~Prdiction~~ Prediction Using Regression Algorithms


## ABSTRACT

*Cryptocurrency is playing a very crucial impact globally, due to its rising popularity and financial acceptance. Unlike the current fiat currencies, bitcoins offer a unique possibility to predict their price. They are an evolution in money. Despite the fact that many individuals are investing in cryptocurrencies, little is known about their dynamic properties and predictability, which puts money at risk. The aim of this paperis to evaluate and compare different regression algorithms in oder to forecast the price of most popular cryptocurrency – Bitcons. We obtained a secondary bitcoin historical data from Kaggle which features an updated daily record of 24 variables over a seven-year period. Since the bitcoin data is so volatile, we implemented an effective pre-processing of data in order to have a better prediction result. The different models applied include – Linear Regression, Ridge Regression, LASSO Regression and Elastic Net Regression model. However, elastic net performed better with an RMSE of 0.0228 without showing signs of overfitting.*



**Keywords**: Cryptocurrency, Linear Regression, Ridge Regression, LASSO Regression and Elastic Net Regression.




## 1.0 INTRODUCTION

In recent years, the cryptocurrency market experienced an all-time boom. With the emergence of more exchanges, cryptocurrencies have become readily available and accessible to the global population and thus further risen in popularity. This, combined with a plethora of auspicious crypto projects pioneered by some of the founders, has led to a surge in number of crypto users and crypto interest alike. Cryptographic currencies represent a growing asset class which has, to a great deal drawn the attention of key players in the forex market and major financial communities. Companies such as Microsoft, Dell and Tesla have already recognized the surging popularity and are accepting virtual currencies (Kethineni and Cao, 2020). Compared to the traditional fiat currencies like The Nigerian naira, The US dollar or The British pound, cryptocurrencies are relatively new. They do not have a real physical form; they are stored in digital wallet and cold storage devices like hard drive. There are different types of cryptocurrencies they include Tether, Solana, Bitcoin, Cardano, Terra, Ethereum, Ripple, and Litecoin. But the major ones dominating the entire market space are Bitcoin, Ethereum, and Litecoin. The list continues to evolve as more capitalists emerge. The same network of computers are used by Bitcoin, Ethereum and Litecoin to store the data of all transactions, therefore it is highly unlikely to have anomalies thus the network is safe (Apostolaki *et al.*, 2017). All cryptocurrencies are decentralized; which means that the market is not influenced by the government or bank and the currency cannot be inflated (Jiménez, 2022). Crypto markets are solely determined by the buyers and sellers of the currency; they allow peer-to-peer direct electronic payment eliminating the need for any middle man such as banks (Nakamoto, 2008). This decentralized market form makes the crypto currency market very volatile; traders can either makea lot of profit or take big losses. The most valuable cryptocurrency in the world is Bitcoin, a kind of electronic money, which was invented by an unknown person or group of persons using the pseudonym Satoshi Nakamoto (Nakamoto *et al.*, 2008), in 2009 they began their network of nodes. The introduction of Bitcoin posed a major question; Are these digital coins considered real money? History has shown, according to Ali *et al*., (2014), that money should satisfy the following criteria (1) A value store. Having monetary value that consumers can use to purchase items from the present to the future. (2) An exchange medium. (3) A unit of account and the capacity to make remittances. Money should theoretically meet all of these requirements, but this is not always the case. All three criteria are arguable when looking at Bitcoin and other cryptocurrencies in their current state. One could argue that it has a store value because of the ability to execute various forms of transactions, however, because of the uncertainty, one cannot ascertain whether Bitcoin can be used in the same way it is currently in the future. Some may argue that bitcoin can be used as a medium of exchange, but others may disagree since to them there exists a limitation to the number of goods that can actually be exchanged. If these three conditions are defined as pre-requisites for any asset to be granted the status of money, it should be recognized in the framework of its utilization (Carrick, 2016). Cocco *et al.*, (2019), the major feature that distinguishes Bitcoin from other traditional currencies is that it allows direct transactions between individuals. Bitcoin is the first transfer and transaction system that employs nodes and doesn't depend on transaction processing and validation by a third party. Because of these distinct traits, the system as a whole is autonomous and decentralized (Brito, et al., 2014). Salman *et al.*, (2021), with escalating geopolitical and economic concerns, worldwide currency values have been declining in the previous years, and stock markets have had a dismal run, with investors losing wealth. This has rekindled interest in digital currency. Due to its consistent performance over the last few years, cryptocurrency, one of the most renowned digital currencies, has found itself in the spotlight, with investors demanding a piece of it and commercial establishments adopting it as a means of payment.

Despite its extreme volatility, Bitcoin is presently the most valuable cryptocurrency on the market, and it has become an appealing option for investors. As a result, the primary goal of this research is to compare four regression algorithms; Multiple Linear regression, Ridge, LASSO and elastic net regression in order to predict the price of bitcoin.

## 2.0 Methodology for Analysis

### 2.1 Multiple Linear Regression

For data modeling and analysis, regression analysis is an essential technique. It is a method of predictive modeling that looks at how dependent (target) and independent variables (predictor) are related. In order to minimize the variations in the distances between the data points, the model fits a curve or line through them. The link between the target (Y) and independent variables (X) is established by linear regression utilizing the best fit straight line. In order to find the line that best fits the observed data, the model minimizes the sum of the squares of the vertical deviations between each data point and the line.

The multiple regression model may be written as:

~~$y_1 = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_n x_{ni} + e_i$ (3.1)~~

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i \quad (3.1)$$

where ~~$Y_i$~~ $y_i$ is the dependent variable on the $i$th observation.
~~$\beta_0, \beta_1, + \ldots + \beta_k$~~ $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients and
~~$x_{1j}, x_{2j}, \ldots, x_{kj}$~~ $x_{1i}, x_{2i}, \ldots, x_{ki}$ are observed values of the independent variables ~~$x_1, x_2, \ldots x_k$~~ $x_1, x_2, \ldots, x_k$, respectively and ~~$e_j$~~ $e_i$ is the error term.
The simple linear regression model was applied in order to fit the best line with minimum cost of the study data, so as to predict the Bitcoins market price given the input futures.

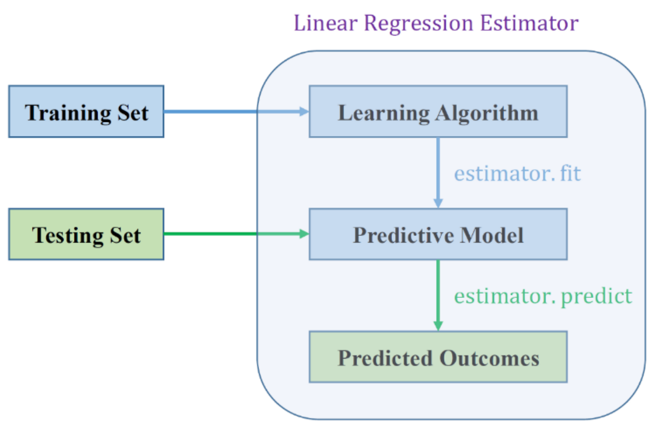



Figure 1: Regression, machine learning block diagram

**2.1.1 Assumptions of Multiple Linear Regression**

Multiple linear regression analysis makes several assumptions

- **Linear relationship:** Each predictor variable and the response variable have a linear relationship. If a relationship is linear or curved, scatterplots can demonstrate this (Orme, 2009).
- **No Multicollinearity:** No predictor variable has a strong correlation with any other. The Variance Inflation Factor (VIF) numbers and correlation table are used to evaluate this hypothesis. (Thompson *et al.*, 2017).
- **Homoscedasticity:** The residuals have constant variance at every point in the linear model. A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of the independent variables.
- **Multivariate Normality:** The residuals of the model are normally distributed. The assumption can be tested by looking at the Q-Q plot for the model. The closer the dots lie to the diagonal line, the closer the residuals are distributed.

### 2.2 Ridge Regression

This technique carries out L2 regularization. Predicted values are far from the real values when the problem of multicollinearity arises, least-squares are unbiased, and variances are high.

#### 2.2.1 Features of Ridge Regression

In Ridge Regression, most of the variables are considered while setting up the model, so as to bring the coefficients of the variables closer to zero so as to remain them from the model. It is resistant to overfitting.

- It has a little variance while being biased.
- When there are too many parameters, it performs better than the Least Squares technique.
- Provides a countermeasure to multidimensionality. The issue in this case is that there are more variables than observations. It provides a remedy for this.
- It works well in problems involving multiple linear regression. The strong correlation between the independent variables in this situation is problematic.
- It is important to find an optimum value for λ. Cross-Validation is used for this.

Model overfitting occurs as a result of the bias decreasing as the variance increases with increasing model complexity. Bias is the deviation between the expected model forecast and the actual value

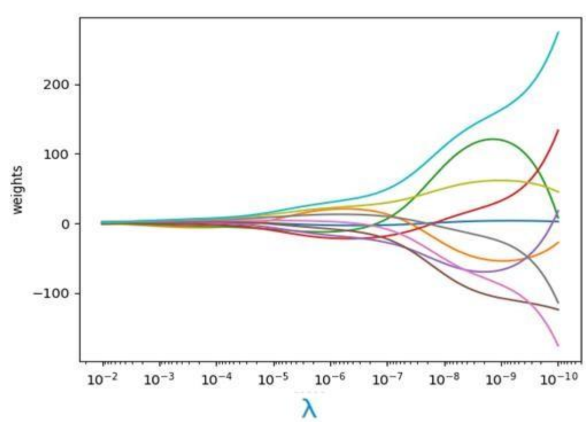



Figure 2: Ridge Coefficient Path

Variance measures how much the target function changes over the data training process. Another explanation is the model's adaptability to the dataset's data points. When the model is exceptionally specialized to the training set, it is said to be highly fitted (Metha *et al.,* 2019). Ridge Regression makes an effort to balance between (i) the complexity of the model and (ii) the prediction function that fits the data the best.

**2.2.2Ridge Regression Model**

$$SSE_{L_2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{P} \beta_j^2 \quad (2.2)$$

$$SSE_{L_2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{P} \beta_j^2 \quad (2.2)$$

Where λis a training parameter and coefficient magnitude. Therefore, it is used to prevent multicollinearity and It reduces the model complexity by coefficient shrinkage.

In Ridge Regression, λ plays a critical role. It allows controlling the relative effects of the two terms. So actually, λ is the penalty term. By changing the alpha value, we control the penalty term. If λ is zero, this gives us the classical regression equation. Consequently, the higher the Alpha values, the greater the penalty. Therefore, the size of the coefficients is reduced.

**2.3 LASSO Regression**

Least Absolute Shrinkage and Selection (LASSO) Regression is an alternative to Ordinary Least square to avoid over fitting in the presence higher number of independent variables. Higher number of coefficients are significant because they draw attention to characteristics that might be reliable outcome predictors. By performing L1 regularization, Lasso regression imposes a penalty equal to the magnitude of the coefficients in absolute terms. Sparse solutions are produced through L1 regularization.

.

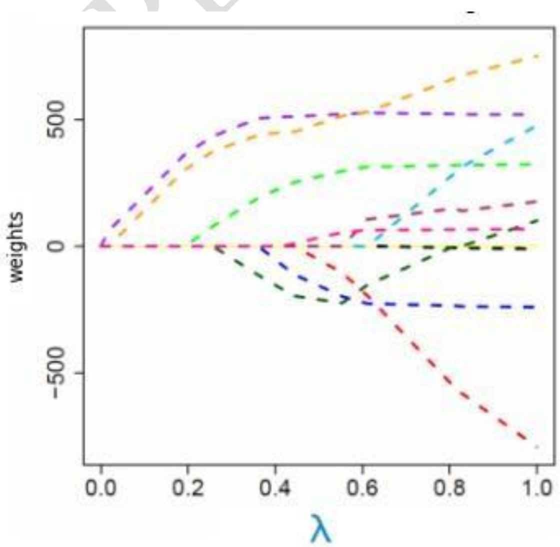



Figure 3: LASSO Coefficient Path

**2.3 LASSO Regression Model**

$$SSE_{L_1} = \sum_{i=1}^{n} \left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad (2.3)$$

$$SSE_{L_1} = \sum_{i=1}^{n} \left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad (2.3)$$

- λ denotes the amount of shrinkage.
- λ = 0 implies all features are considered and it is equivalent to the linear regression where only the residual sum of squares is considered to build a predictive model
- λ = ∞ implies no feature is considered i.e, as λ closes to infinity it eliminates more and more features
- The bias increases with an increase in λ
- Variance increases with a decrease in λ

Simple and sparse solutions are encouraged by the LASSO technique. This can do feature selection by shrinking some coefficients to zero. More coefficients will be set to zero as value rises, allowing for their removal and the examination of just features of substantial size.

## 2.5 Elastic Net Regression

L1 and L2 (Lasso and Ridge) techniques are combined in Elastic Net. It performs a smoothing procedure more effectively as a result. When Lasso was first criticized for having variable selection that could be overly dependent on data and unstable, Elastic Net was born. To achieve the best of both worlds, the approach is to mix the penalties of Ridge regression and Lasso (Casella *et al.*, 2010).

### 2.5.1 Features of Elastic Net Regression

- It combines the L1 and L2 approaches.
- It performs a more efficient regularization process.
- It has two parameters to be set, *λ* and *α*.

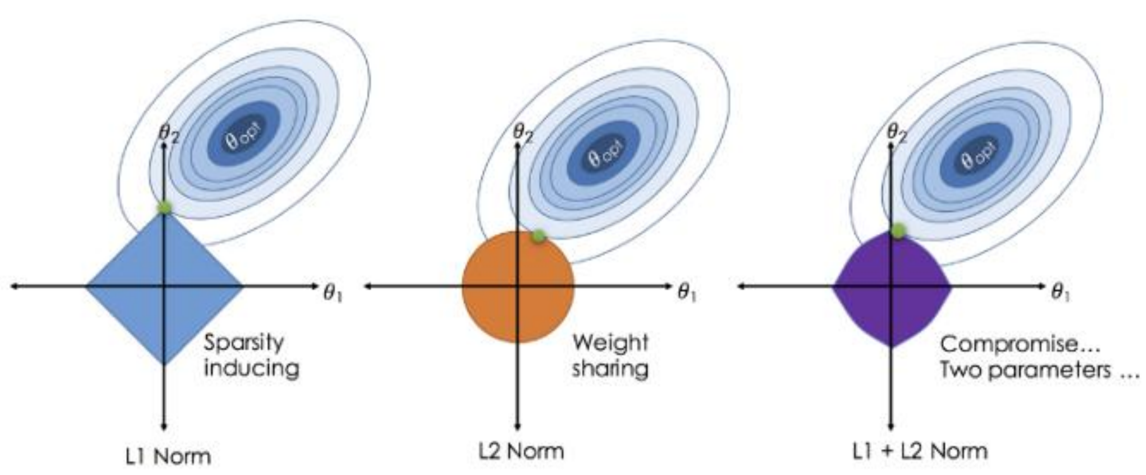


Figure 4: Differences between L1, L2 and L1+ L2 norm

The elastic net method improves on lasso's limitations, i.e., where lasso takes a few samples for high dimensional data, the elastic net procedure provides the inclusion of "n" number of variables until saturation. In a case where the variables are highly correlated groups, lasso tends to choose one variable from such groups and ignore the rest entirely.

### 2.5.2 Elastic Net Regression Model

Elastic Net aims at minimizing the following loss function:

~~$$L_{enet}(\hat{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\boldsymbol{\beta}})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\boldsymbol{\beta}}_j^2 + \alpha\sum_{j=1}^{m}\left|\hat{\boldsymbol{\beta}}_j\right|\right), (2.4)$$~~

$$L_{enet}(\hat{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\boldsymbol{\beta}})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\boldsymbol{\beta}}_j^2 + \alpha\sum_{j=1}^{m}\left|\hat{\boldsymbol{\beta}}_j\right|\right), (2.4)$$

The terms used in the mathematical model are the same as in Ridge and Lasso Regression.

## 3.0 Data Presentation and Analysis

This chapter discusses presentation, analysis and interpretation of data. Data on its own means nothing without proper analysis. Thus, we will subject our already obtained data to Multiple Linear, Ridge, LASSO and Elastic Net regression analysis in a bid to obtain the best model based on its RMSE.

### 3.1 Data Presentation

The data for the analysis consist of24 variables which are; date, market price,total bitcoins, market supply, trade volume, block size, number of orphaned block, number of transaction per block, median confirmation time, hash rate, difficulty, miners revenue, transaction fee, cost per transaction, cost per transactionperent, unique addresses, number of transactions, total transactions, number of transaction excluding 100 popular addresses, number of transactions excluding chains longer than 100, output volume, estimated transaction volume, estimated transaction volume in USD value.

Furthermore, the entire dataset has a total of 2906 samples, 115 samples were reported as missing values. This constituted approximately 4% of the dataset.

### 3.2 Data Analysis

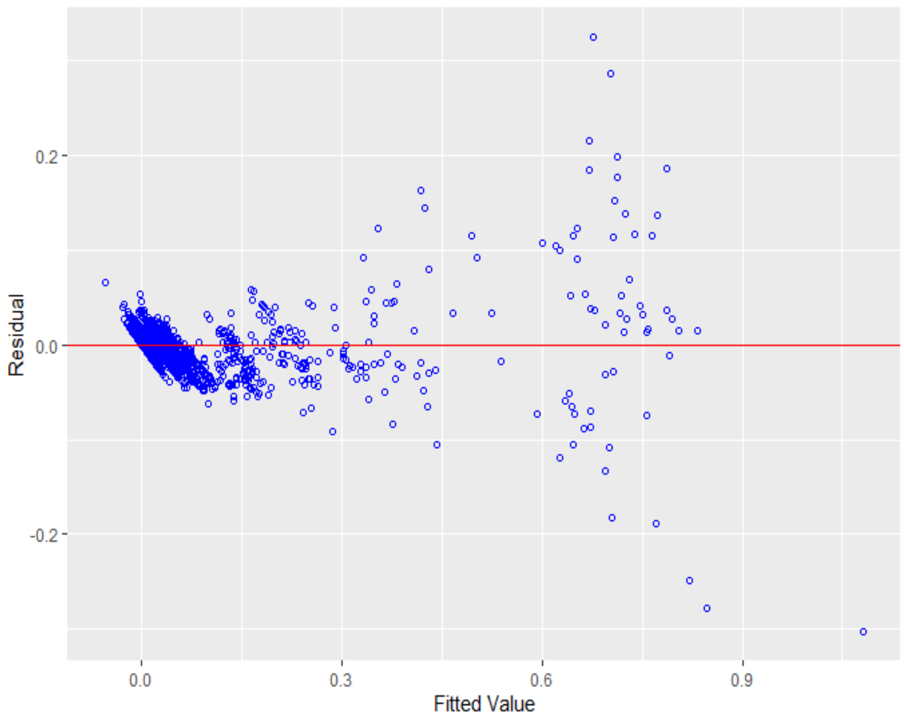
#### 3.2.1 Multiple Linear Regression

The Linear regression model was the first model trained to predict the Bitcoin market price given the input features. The model performed well and produced good accuracy.

#### 3.2.1.1 Testing for Assumptions of Multiple Linear Regression Model

**Testing for Homoscedasticity**

Figure 5: Scatterplot of residuals vs fitted values




**Interpretation:**

The plot of residual vs fitted values shown in figure 5 above depicts a clear violation of this assumption as there's an obvious sign of funneling.

- **Significance Test for Homoscedasticity**

*Breusch Pagan Test for Heteroskedasticity*

Statement of Hypothesis:

Ho: Homoscedasticity is present

$H_1$: Heteroscedasticity is present

**Results of the Breush Pagan Test**

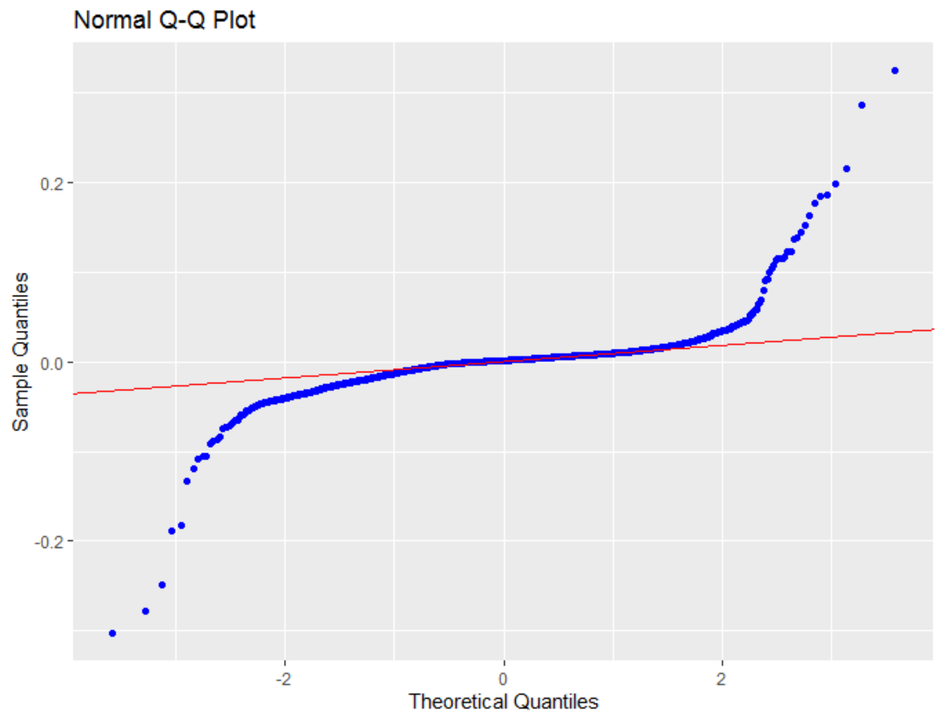
Level of Significance, $\alpha = 0.05$

DF = 1, $Chi^2$ = 23320.9374, p-value => $Chi^2$ = 0.0000

**Decision:** If the p-value <α = 0.05, we reject the null hypothesis and conclude that heteroscedasticity is present in the regression model.
**Conclusion:** Since p value <α = 0.05, reject the null hypothesis, and conclude that heteroscedasticity is present in the regression model. Hence, this is assumption is not met.

**Testing for Multivariate Normality**

Figure 6: Normal Q-Q Plot




**Interpretation:**

The Q-Q plot for the model suggested that assumption of normality of the residuals may have been violated. Since the residuals do not form a straight line.

- **Significance Test for Normality**

Shapiro Wilk Normality Test

**Statement of Hypothesis**:

Ho: The variable is normally distributed

$H_1$: The variable is not normally distributed

**Results of Shapiro Wilk Normality Test**

Level of Significance, α = 0.05

W = 0.62804, p-value < 2.2e-16

**Decision:** If the p-value <α = 0.05, we reject the null hypothesis and conclude that the variable is not normally distributed

**Conclusion:** Since p value <α = 0.05, we reject the null hypothesis, and conclude that the variable is not normally distributed. Hence, this is assumption is not met.

**Test for Linear Relationship**

*The scatter plot for this assumption is contained in appendix A*

From the scatter plot, it is clear that there exist a linear relationship between independent variables. Hence this assumption is met.

**Test for Multicollinearity**

*The correlation chart for this assumption is contained in appendix A*

From the correlation chart, we see that this assumption is largely violated since there is high correlation among independent variables.

As a result of these violations, the estimates gotten from this model won't be reliable.

Table 1 Linear Regression Model Coefficients And Accuracy

| | | | | | |
|---|---|---|---|---|---|
| Intercept | 0.002003 | | | | |
| Coefficiens | p-value: < 2.2e-16 | | | | |
| | C= 0.040681 | I= 0.83207 | L=0.371405 | Q=0.267761 | T= -0.078149 |
| | E= 0.345165 | J= 0.016290 | N=0.058921 | R=0.183696 | U= -0.153976 |
| | G= -0.56365 | K=0.246282 | P=0.220274 | S= -0.181930 | V= -0.029434 |
| Train RMSE | 0.02447473 | | | | |
| Test RMSE | 0.02281077 | | | | |
| R-Squared | 0.9580965 | | | | |

Source: Author's computation from R Software

### 3.2.2 Ridge Regression

The Ridge Regression algorithm trained the model to penalize for overfitting. It introduced an additional L2 regularization term to tradeoff between bias and variance. Moreover, the cross-validation technique was used to validate the r-squares values of the training and test dataset.

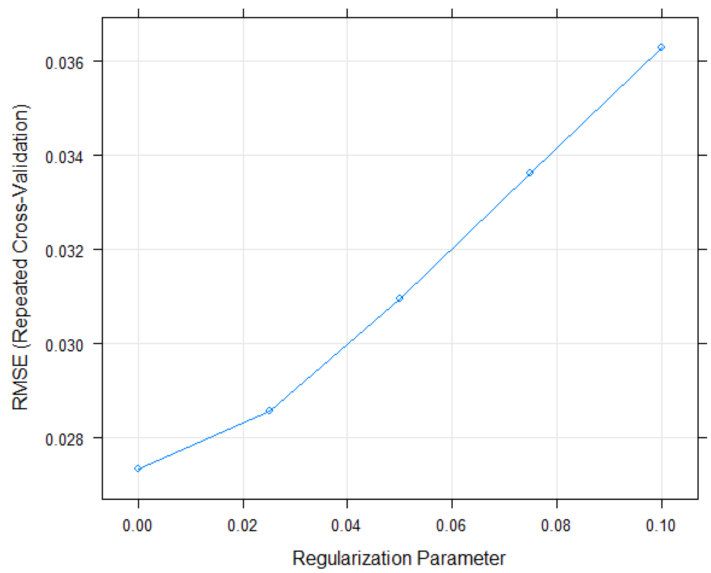



Figure 7Ridge tuning parameter plot of regularization.

**Interpretation**

From figure 7 above, the error increases with the regularization parameter λ. Hence the best value of λ is approximately 0.0001

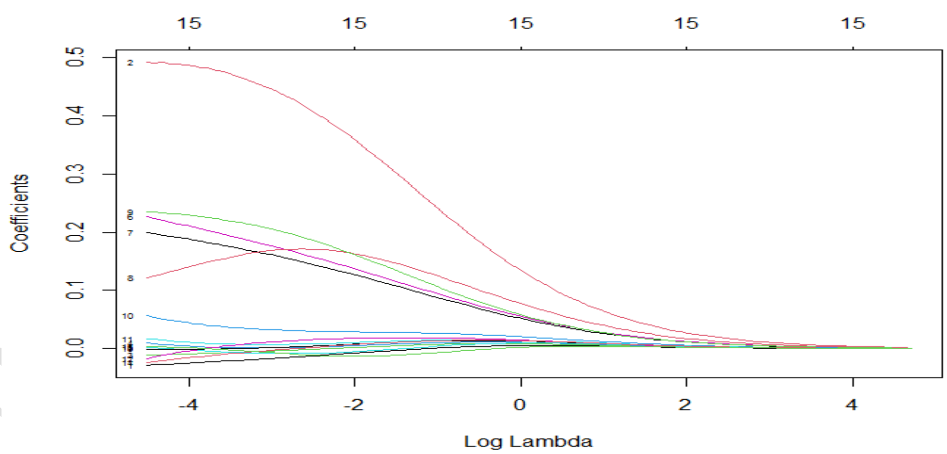



Figure 8 Normalized Model Coefficients vs. Lambda Hyperparameter.

**Interpretation**

From figure 8 above, we notice that as log Lambda increases the coefficients shrinks to zero; But, when it is relaxed, the coeficients begins to grow. Also notice that all 15 independent variables are present in the model. This a graphic illustration of the L2 regularization. It shrinks coefficients towards zero.

Table 2 Ridge Regression Model Metrics and Accuracy

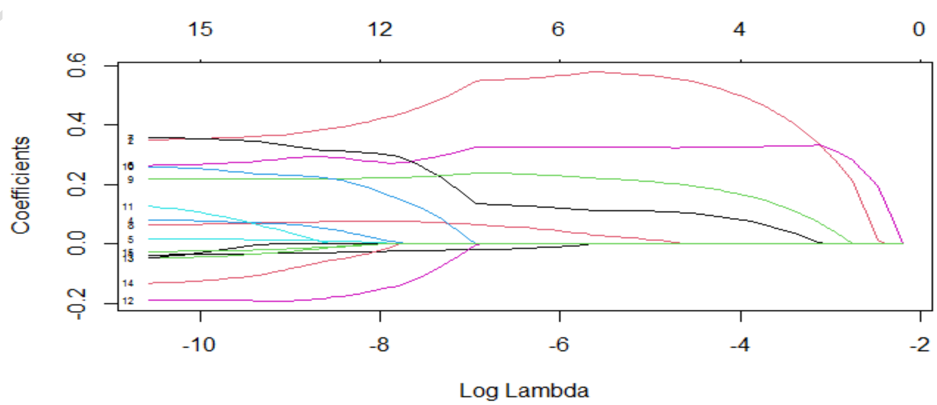
| Model Information | Values | | |
| --- | --- | --- | --- |
| Alpha Range | 0.0001, 0.001, 0.01, 0.1 | | |
| Cross Validation | 10 | | |
| Intercept | -0.0009891647 | | |
| Coefficients | C= -0.0284085068 | K = 0.2269659104 | Q = 0.0560955822 |
| | E = 0.4920755688 | L = 0.1994968887 | R = 0.0168386407 |
| | G= -0.0117126762 | N = 0.1216254065 | S = -0.0172235687 |
| | I = 0.0095508077<br>J= 0.0019298664 | P = 0.2352508838<br>V = 0.0030726452 | T = -0.0007325588<br>U = -0.0248116801 |
| Training RMSE | 0.02720223 | | |
| Test RMSE | 0.02504366 | | |
| R – Squared | 0.9500292 | | |

Source: Author's computation from R Software

### 3.2.3 LASSO Regression

The Lasso Regression algorithm introduces the L1 regularization term that plays a vital role in feature selection. The model performed well with high accuracy. The 5-folds cross validation set is used to evaluate the quality metrics of the training and test datasets.

Figure 9: Standardized Model Coefficients vs. Lambda Hyperparameter



**Interpretation**

LASSO regression model does feature selection; it shrinks coefficients that are of less importance to zero hence eliminating. This is done in order to enhance it's accuracy. As shown in figure 9, the coeffcients are eliminated according to their degree of importance from the least important to the most important. So as log lambda increases, the number of coefficients in the model reduces. Consider the variable importance plot below.

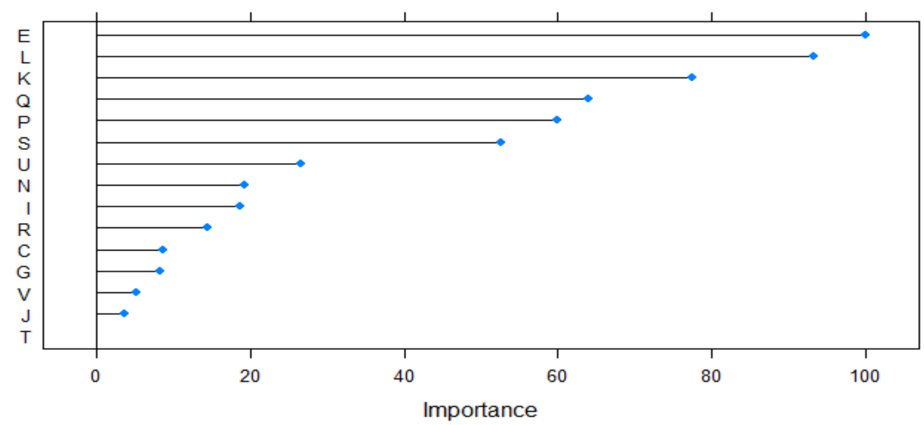



Figure 10.: Variable Importance Plot

Table 3 LASSO Regression Model Metrics and Results

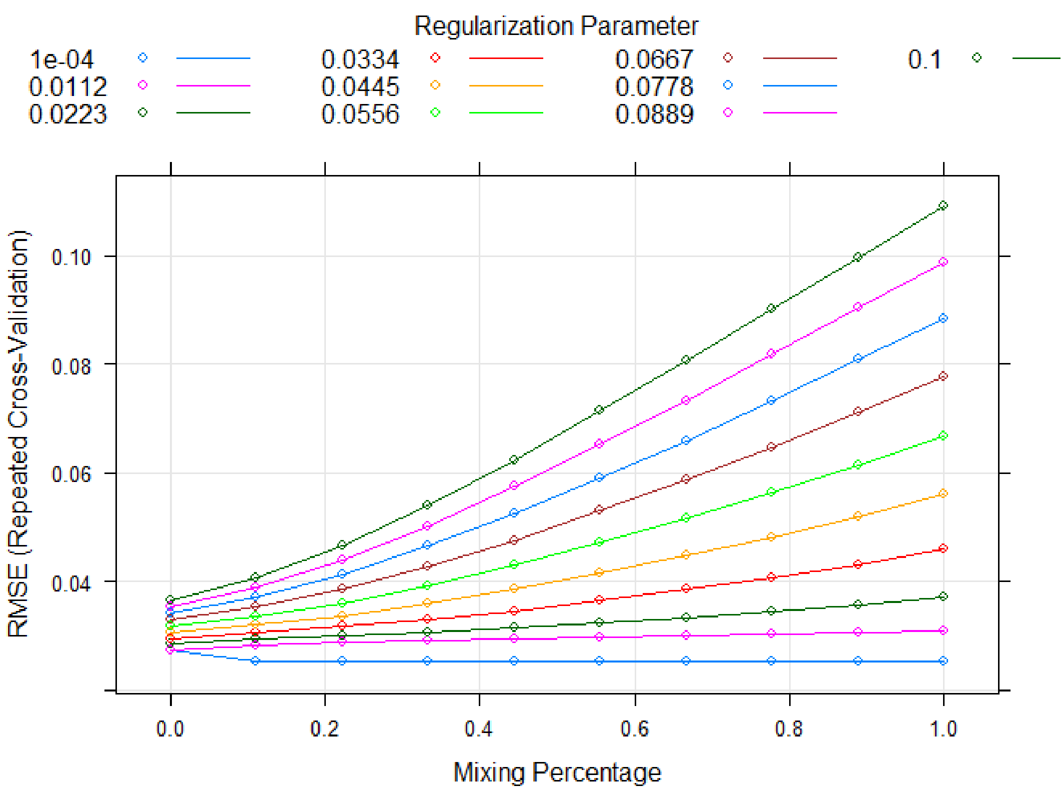
| **Model Information** | **Values** | | |
|---|---|---|---|
| Alpha Range | 0.0001, 0.001, 0.01, 0.1 | | |
| Cross Validation | 10 | | |
| Intercept | 0.0004880324 | | |
| Coefficients | C= - 0.0322869583 | K = 0.2827787982 | Q = 0.2332822381 |
| | E = 0.3647740360 | L = 0.3400630716 | R = 0.0533655394 |
| | G= - 0.0307510865 | N = 0.0709673430 | S = -0.1924586728 |
| | I = 0.0684870305 | P = 0.2184687690 | T = -0.0005746786<br>U = -0.0975323840 |

| | J= 0.0139131578 | V = - 0.0192802854 | |
|---|---|---|---|
| Training RMSE | 0.02463591 | | |
| Test RMSE | 0.02289086 | | |
| R – Squared | 0.9580188 | | |

Source: Author’s computation from R Software

### 3.2.4 Elastic Net Regression

The Elastic Net Regression algorithm trained the model to penalize for overfitting. It combines the L1 and L2 regularization term to tradeoff between bias and variance. Moreover, the cross-validation technique was used to validate the r-squares values of the training and




test dataset.
Figure 11: The 10-fold cross validation RMSE across 10 alpha values (x-axis) and 10 lambda values (line colours)

**Interpretation**
From the plot above it is obvious that λ= 0.0001 has the lowest RMSE with an approximate alpha value of 0.111. Hence our final Elastic Net model is more of a

Ridge model than a LASSO model since α =0.111 is closer to zero (0) than it is to one (1).

Table 4 Elastic Net Regression Model Metrics and Results

| **Model Information** | **Values** | | |
|---|---|---|---|
| Alpha Range | 0.0001, 0.001, 0.01, 0.1 | | |
| Cross Validation | 10 | | |
| Intercept | 0.001132094 | | |
| Coefficients | C=-0.036103685 | K = 0.284283274 | Q = 0.235558750 |
| | E =0.361208724 | L = 0.330862199 | R = 0.107514326 |
| | G=-0.047991196 | N = 0.066386519 | S =-0.178616447 |
| | I = 0.084996095 | P =0.221317484<br>V =-0.023445509 | T =-0.030295349<br>U =-0.124969942 |
| Training RMSE | 0.02453609 | | |
| Test RMSE | 0.02283661 | | |
| R – Squared | 0.9587276 | | |

Source: Author’s computation from R Software

**Please add a conclusion stating that the best method is elastic net regression.**

**REFERENCES**


Ali, R., Barrdear, J., Clews, R., and Southgate, J. (2014). Innovations in payment technologies and the emergence of digital currencies. *Bank of England Quarterly Bulletin*, Q3.

Apostolaki, M., Zohar, A., and Vanbever, L. (2017, May). Hijacking bitcoin: Routing attacks on cryptocurrencies. In *2017 IEEE symposium on security and privacy (SP)* (pp. 375-392). IEEE.

Brito, J., Shadab, H., and Castillo, A. (2014). Bitcoin financial regulation: Securities, derivatives, prediction markets, and gambling. *Colum. Sci. & Tech. L. Rev.*, *16*, 144.

Carrick, J. (2016). Bitcoin as a complement to emerging market currencies. *Emerging Markets Finance and Trade, 52(10), 2321-2334.*

Casella, G., Ghosh, M., Gill, J., and Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian analysis*, *5*(2), 369-411.

Cocco, L., Tonelli, R., and Marchesi, M. (2019). An agent-based artificial market model for studying the bitcoin trading. *IEEE Access*, *7*, 42908-42920.

Jiménez, I. (2022). Marketing Plan-University Coin.

Kethineni, S., and Cao, Y. (2020). The rise in popularity of cryptocurrency and associated criminal activity. *International Criminal Justice Review*, *30*(3), 325-344.

Matta, M., Lunesu, I., and Marchesi, M. (2015, June). Bitcoin Spread Prediction Using Social and Web Search Media. In *UMAP workshops* (pp. 1-10)

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*.

Nakamoto, S. (2008). Re: Bitcoin P2P e-cash paper. *The Cryptography Mailing List*, 1-2.

Orme, J. G., and Combs-Orme, T. (2009). *Multiple regression with discrete dependent variables*. Oxford University Press.

Thompson, C. G., Kim, R. S., Aloe, A. M., and Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, *39*(2), 81-90.