An Analysis of the Relationship Between Elon Musk's Twitter Posts and Changes in the Value of Dogecoin

Abstract:

Elon Musk frequently makes Twitter posts, many of which have made their way to media headlines for supposedly affecting the value of Dogecoin cryptocurrency. The entrepreneur has a large number of strong supporters who will often make major investments on the basis of his posts. The objective of this study is to analyze whether a statistical correlation can be observed between the sentiment of his posts and changes in the value of Dogecoin. Our data is derived from historical data of the Dogecoin currency and a sentiment analysis of Elon Musk's Twitter posts. We test this with support vector machines and logistic regression models, in which the positivity and negativity of Musk's posts are used to predict whether the value of Dogecoin rose or fell in that time period. We failed to find any statistical relationship between the sentiment of his posts and changes in the value of Dogecoin.

1. Background

The cryptocurrency Dogecoin was created in December 2013 as a satirical response to the rampant speculation in cryptocurrencies at the time. The creators were intending to implement a peer-to-peer cryptocurrency whose technicalities and outward appearance would be more inviting to amateur and mainstream investors than other cryptocurrencies of the time. Dogecoin's popularity was instantaneous; Within the first month it had already experienced its first major rise and major crash, of more than 300% and 80%, respectively.

Since then, it has grown to be one of the leading cryptocurrencies in terms of media popularity. Nearly everyone in the developed world has likely heard of it at one point or another. Countless amateur speculators have invested in it, which has led to incredible volatility. This reality has been further amplified ever since the start of the COVID-19 pandemic. A sudden increase in free time and cash provided by government stimulus invited millions to indulge in the speculation.

An even more interesting note is the interest that Elon Musk has taken in the cryptocurrency and the interest that many speculators have taken in Elon Musk. Many of the amateur investors in Dogecoin have an almost religious view of Musk and have historically deferred to his public opinions in Dogecoin. His Twitter posts have been a beacon for many people, as the entrepreneur has been characterized as wielding considerable soft power over the investors and consequently over the value of the cryptocurrency itself.

2. Research Objective

The objective of this study is to determine whether a correlation can be found between sentiment analysis of Elon Musk's Twitter Posts and a positive or negative change in the value of the Dogecoin cryptocurrency.

3. Methodology

Based on historical data, we developed models using logistic regression and linear discriminatory analysis, then applied cross validation in order to estimate their error and, if necessary, choose the more accurate of the two. These several steps are explained in greater detail below.

Creation of Data

For changes in the value of Dogecoin, we imputed historical data of daily Dogecoin closing value into Microsoft Excel and then applied a simple function to return a 0 if the value decreased, and a 1 if it increased. This gave us our boolean response variable, indicating whether the value grew or shrunk over that day.

The predictors, one positive and one negative, are a sentiment analysis of Elon Musk's tweets. They were found using the natural language processing package VADER in Python. The VADER package has a pre-trained lexicon for sentiment analysis. Musk's tweets from the last year were processed using the VADER library, which returned the positive and negative continuous values for sentiment. In addition to this, we also removed a large number of entirely neutral observations, and transformed those that are often in sediment analysis considered to be more noise than signal

It can be reasonably assumed that, if Musk's Twitter posts do affect values in cryptocurrency, then that effect may not necessarily take place on the same day as the Twitter post itself. In recognition of this, our predictors were transformed into a three day rolling average of the positive and negative values of Musk's posts. The basis for this change was entirely subjective and based on our knowledge of historical changes in market values in similar situations.

Logistic Regression

In order to use continuous variables to predict a binary variable, we employed Logistic Regression. This was done by reforming the linear regression model into the following form:

$$\frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 y)}} = z$$

Where *x* is the positivity value, *y* is the negativity value, β_i for $i = \{1, 2, 3\}$ are the least squares values gathered found in the linear regression, and *z* is the probability of a rise in the value of Dogecoin. Note that this formula was utilized automatically in R.

Linear Discriminant Analysis

Linear Discriminant Analysis was also employed here. This was applied in the hopes that the Bayes Theorem approach would be better suited to modeling this data. Bayes Theorem is the following. This approach essentially applied the Bayes Theorem to the data to fit the observations to classes that would ideally be consistent with their true response (i.e. whether or not Dogecoin rose or fell on that day).

Cross Validation

After the regression was performed via both approaches, they were compared to each other using cross validation. We used a five fold approach (k=5), since we were working with a large number of observations and it is a commonly used value in k-fold cross validation. This figure was also chosen in considering the balance of bias versus variance of our results.

4. Results

You can see the model summaries for logistic regression and linear discriminatory analysis in figures one and two, respectively. In summary, the logistic regression suggests a lack of correlation between the sentiment analysis data we generated. The p-values in the logistic regression are very high and are consequently insignificant. In addition to the lack of correlation found in the model for logistic regression, the model assumptions are also strained (see figure 3). This is because the techniques by which we acquired our predictors left many of the sentimental analysis positive and negative values closer to one or negative one. As mentioned earlier, this was done in order to decrease the noise in the data. P-values were not provided for the linear discriminatory analysis, but the cross validation done suggested a complete lack of collinearity.

The cross-validation, done at k=5, resulted in an estimated classification error rate of 48.8% and 48.5% for LDA and logistic regression, respectively. Obviously, these results suggest a complete lack of collinearity between the response and it's predictors, which is consistent with the high p-value given in the first model. Overall, it serves no purpose choosing a superior model, since they are both incredibly unreliable at predicting an increase or decrease in dogecoin. Note that we did not use the cv function in R to compute these values.

5. Discussion

The data for our study was much more limited than we would have liked. Originally, we hoped to conduct sentiment analysis on all Twitter posts relating to Dogecoin over the same two year period. However, that data, over a sufficiently wide range of time, is either non-existant or not publicly available. We also hoped to use the popularity of the Robinhood trading application as a predictor, however we were also unable to find that data. All this to say, we limited our research proposition to whether the positivity or negativity of Musk's Twitter posts noticeably affect Dogecoin values.

For all the media attention that Musk receives when he tweets about Dogecoin, this study suggests that Bitcoin can frequently rise and fall independent of Musk. In other words, because we failed to find a relationship between the sentiment of Musk's tweet and the value of Dogecoin, we may be able to conclude that there is plenty of change in Dogecoin that Musk is not responsible for, though this is merely conjecture.

Other studies could better observe how Elon Musk influences the value of the cryptocurrency by performing sentiment analysis of all Twitter posts with respect to Elon Musk and Dogecoin, then measure that against changes in value. Future work could also make use of Twitter posts to find correlations between the sentiment of the entire social network with changes in Dogecoin, in an attempt to find a correlation between the sentiment of the social network and Dogecoin, without regard for Musk. Other studies could also analyze correlations between the popularity of do-it-yourself trading applications and changes in value. In essence, a large-scale study of sentiment regarding grassroots investing would be more well equipped to study changes in Dogecoin.

References:

1. Frankfield, Jake. (2020). Dogecoin. https://www.investopedia.com/terms/d/dogecoin.asp

2. Dhruvil, Dave. (2021). Dogecoin Historical Data. kaggle.com/dhruvildave/dogecoin-historical-data

3. Oltenau, Andrea. (2021). All Elon Musk's Tweets. https://www.kaggle.com/andradaolteanu/all-elon-musks-tweets?select=TweetsElonMusk.csv

Appendix

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2313  -1.1825  -0.8294   1.1713   1.4138

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                0.1574     0.1809   0.870    0.384
senDoge$Pos Avg by Day  -1.3093     1.2052  -1.086    0.277
senDoge$Neg Avg by Day  -5.2857     5.0546  -1.046    0.296

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 429.70  on 309  degrees of freedom
Residual deviance: 427.58  on 307  degrees of freedom
  (30 observations deleted due to missingness)
AIC: 433.58

Number of Fisher Scoring iterations: 3
```

Figure 1: model summary of logistic regression

```
lda(response ~ pos + neg, data = senDoge, subset = train)

Prior probabilities of groups:
        0         1
0.4960938 0.5039062

Group means:
        pos        neg
0 0.1056352 0.01154439
1 0.1041054 0.00654055

Coefficients of linear discriminants:
           LD1
pos  -3.079271
neg -33.835528
```

Figure 2: Model summary of LDA

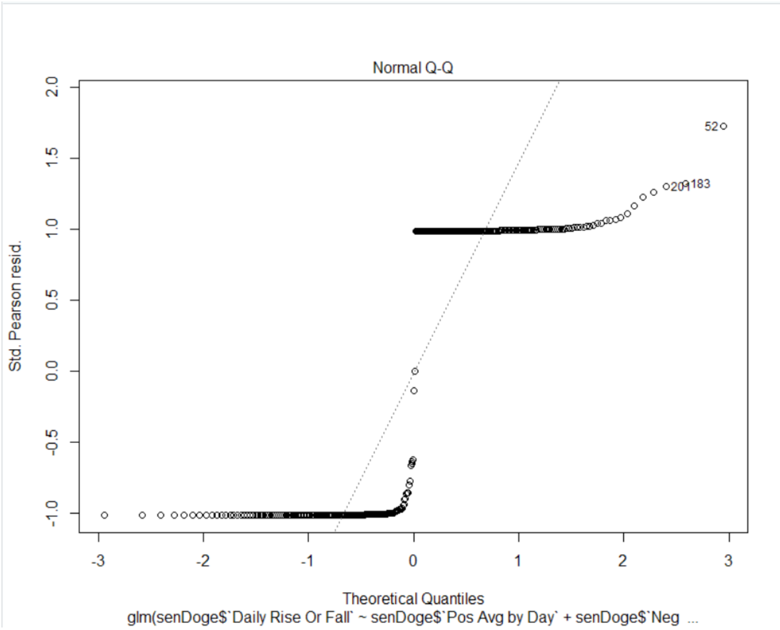

Figure 3: Normal probability plot of residuals and residuals versus fits plot of logistic regression