Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Advanced sentiment analysis of social media for short-term cryptocurrency price prediction.

Krzysztof Wołk1

1Polish-Japanese Academy of Information Technology, Warsaw, MZ 02-008 Poland

Corresponding author: Krzysztof Wołk (e-mail: kwolk@pja.edu.pl).

ABSTRACT Over the last few years, bitcoin has become a fundamental aspect of financial systems. Bitcoin is one the largest cryptocurrency in terms of capital share markets but not the only one. Therefore, using sentiment analysis as a computational opinion can be used to predict bitcoin and other crypto currency prices at different intervals of time. One the key characteristics of crypto market is that the fluctuation of its prices does not depend on institutional money regulation but relies on people’s perception and opinions. Therefore, analysing the relationship between social media and web search is crucial for a cryptocurrency price prediction. In this research Twitter and Google Trends are used to forecast the short-term price of main crypto currencies as they are used to influence purchasing decision. This research article uses multi model approach, that is interpolated, to analyse the impact social media on cryptocurrency prices. Summing up we prove that psychological and behavioural attitude of population has great impact on cryptocurrency prices that are very speculative.

INDEX TERMS Sentiment analysis, Machine learning, cryptocurrencies, Social Media, Speculative Models

I. INTRODUCTION

Bitcoin, Ethereum, Electroneum, Ripple, ZEC Cash and Monero (crypto in short) are cryptocurrencies that are an electronic form of currency transaction. Crypto is a decentralized form of currency transaction which takes place without an intermediary. It was introduced in the market in 2008 by Satoshi Nakamoto (as Bitcoin project) and it can be circulated in the market on peer-to-peer networking transaction. Crypto is different from traditional form of currency transaction such as banking system as it allows its users to enjoy transaction without operation fee and following any rules and authorities of financial institutional which are full of fraud and intense corruption.

Bitcoin and other crypto, is one of the most growing form of digital transaction in the world. At the start of year 2017 the price of bitcoin was $863 but it rose to around $17,000 which is about 2000% increase at the end of the same year. This massive unprecedented rise has captured worldwide attention in digital currency transaction. Basing on the previous research studies, it is evident that bitcoin possesses unique characteristics as compared to traditional mode of transaction such as banking. This is because its price fluctuation depends on people’s perception and opinions instead of following institutional regulations.

However, the value of crypto is volatile and its price keep on fluctuating with time and it’s uncertain for investors and people who wish to use them as a currency. Twitter is one of the most widely form of social media platform which collects multidimensional views and perspectives of different people in the whole world. Therefore, Twitter is used as marketing tool for crypto transactions and hence it can be used to predict their prices. Also, web search tool such as Google Trends is also one of the most widely used research platform that provides wide range of information and therefore it is used as a marketing tool for price for crypto and is used to predict the future prices. This research study analyses correlation between Twitter as a social media platform as the number of Tweets and the prices of the crypto. In addition, the effect of web search data like Google Trends on price of crypto will be examined as well.

Therefore, using sentiment analysis we can predict the price change of crypto on different time interval using different computational and statistical models such as linear regression technique, boosting methods and neutral networks and determine the significance of coefficient of determination using Twitter data and Google Trends. By using linear modelling which takes the number of Tweets and Google Trends, we will be able to accurately make a prediction towards the direction of the changes of prices of crypto. According to the past research studies, sentiment analysis technique can be a good modelling to the capital market and cryptocurrencies.

In order to establish the usefulness of the data, only data that contains a certain set of keywords (cryptocurrency full name and abbreviation) is analysed. The underlying assumption is that the sentiment correlates with the movement of the financial instrument, such as Bitcoin. There is solid research to suggest this correlation exists. Google Trends data consists of relative search volume scores for a given search term, during a given time interval. Several researchers have focused on using Google Trends data to predict the stock market. Many searches for bitcoin or some other keyword could indicate a reaction to current events or predict a future event.

When carrying out analysis of sentiment about opinions and perceptions of Twitter users and google researchers regarding the price of bitcoin, the problem statement that emerges and need to be solved is to determine whether there is correlation between Twitter data and the price fluctuation of crypto. Also, can a prediction of naivety model regarding sentiment changes yields better output as compared to random accuracy.

II. PREVIOUS RELATED WORK

This research paper has been built on a wide range of related research ideas and topics. Some economists which are behavioural in nature articulated that decisions regarding financial systems are influenced by emotional ethics and not by value of the capital alone. This idea of behaviour and emotions was also supported by Dollan [1] who argued that decision making is influenced by emotions. Basing on these researches there is an open possibility to find beneficial tools such as sentiment analysis which shows that the price of a commodity may be impacted by other values such as emotions other than economic fundamentals.

Recent research has pointed out clearly that decisions for purchase made by people are being influenced by the information found in the website and social media. A research study conducted by Gallen Thomas showed that Twitter data analysis correlated with views of the people towards the price of cryptocurrency. In addition to that he also pointed that social media such Twitter sentiments has a great impact to the final users of cryptocurrency as compared to the emotional state of the users. Bollen et al [2] conducted a research on Twitter sentiment against stock market. In his research study, he used neural networks and casualty analysis to determine where the price of the cryptocurrency was heading to. From his results there was ability to predict change of capital market for some days for instance almost one week.

Another research was carried out by Prosky et al [3] who used tensor networks in order to formulate a model for learning. He concentrated on Twitter data in order to carry out sentiment analysis on it and see if they could formulate the results and see the relationships with other different stochastic events. Also, Rather et al [4] developed a recurrent neutral networks and a multiple linear regression to formulate a hybrid model, which tried to provide a solution to each model used and solved the associated limitations. According to researchers these three methods could be useful for the prediction of the price of the cryptocurrency and especially bitcoin.

A research done by Nie et al [5] showed that comments done in social media such as Twitter comments and web search such as Google Trends were very important information that could be used to predict the price of bitcoins and other types of cryptocurrencies. Nie found that these factors produced a very high significance with a very low p-value for search terms related to mining and block chains which are important aspects of cryptocurrencies.

Karalevičius [6] used sentiment analysis to predict intraday Bitcoin movements in social media forums. His conclusion was that short-term price fluctuations could be predicted with some degree of accuracy which diminished as time was increased. The significance to this research project is that our time frame is short, using mostly ten or sixty minute time frames. Garcia et al. [7] showed that it was possible to use a combined strategy to predict Bitcoin price using standard financial modelling techniques and social media signals. These signals included target words, sentiment, and other features that describe the changing environment of social media such as post frequency and comments. These researchers implemented a strategy that yielded 32.29 percent daily gain. Valence measures alone yielded a 0.1183 daily gain. With enough capital, at these rates, trading Bitcoin could be profitable. The researchers performed back testing on their results which add some confidence to their prediction model.

Also, a model formulated by a scholar called Kristoufek et al [8] showed that Google Trends as one of the factors which affect the price of Bitcoin had a strong positive correlation with the price of Bitcoin which achieved a very low p-value with a high significance during the study testing. He also used vector auto regression technique which showed that Wikipedia information was also a good predictor to produce a considerate model for the prediction of the price of the Bitcoin. Stenquist Evita and Lonno Jacob wrote a paper titled “Predicting the price of Bitcoin fluctuation using Twitter sentiment analysis” who collected tweets relating to the price of Bitcoin and formulated a model which was useful to predict the price of Bitcoin [12]. They used Valence Aware Dictionary and Sentiment Reasoner (VADER) to analyse the effect of each tweet and classified them as either positive, negative, or neutral. They only kept those tweets that were negative or positive and thus were used for analysis.

III. METHODS

In this project we have applied different predictive and descriptive models which are important for data analysis. The work was initiated using two predicted models essential for predicting the price of the cryptocurrency with the help of Twitter sentiments and Google Trends. These models are least square linear regression and Bayesian Ridge Regression Model. These models are embedded in Python language library called SKLEARN[[1]](#footnote-1). This model was explained intensively by Kuchibhotla et al [9] who argued that high dimensional data and methods have proliferated throughout the literature for the last two decades.

When data is expressed as a linear combination of a product of independent variables and a coefficient matrix, least squares linear regression seeks to minimize any necessary error that occurs. To determine the coefficients, we use array of independent variables and dependent variables. The relationship between the predicted values Y based on the coefficients and the inputs of the array X is expressed as:

To calculate the better coefficient matrix, we use the following formula:

Another important technique used in analysis is Bayesian Ridge Regression modelled by Nie and Ji (2014) who claimed that future learning refers to learning the transformation of the raw data into useful and analytical data and other purposes. Feature learning techniques can be either supervised or unsupervised, which commonly include auto-encoders, dictionary learning, restricted Boltzmann machine, k-means clustering and many other approaches. During the past few years, restricted Boltzmann machine draws more and more attention from researchers due to its capability of handling different kinds of data and its efficient learning method.

Bayesian Ridge Regression is similar to LSLR, but it adds a lambda parameter to the input values that penalizes the beta coefficients and shifts them towards zero. Bayesian ridge regression returns a probabilistic model with a Gaussian parameter. MacKay [10] describes the Bayesian model with a Gaussian probability parameter in the following equation:

Using the effects of precision of Gaussian, we choose alpha and lambda which are chosen to be gamma distribution. To examine the default parameter in the model for alpha and lambda we use 10-6. These can be adjusted to the data for modeling using the SKLEARN package. Bayesian Ridge Regression assigns coefficient values using the equation:

In the equation above “I” resembles the identity matrix, and the lambda term is applied across only the diagonal elements of the input array.

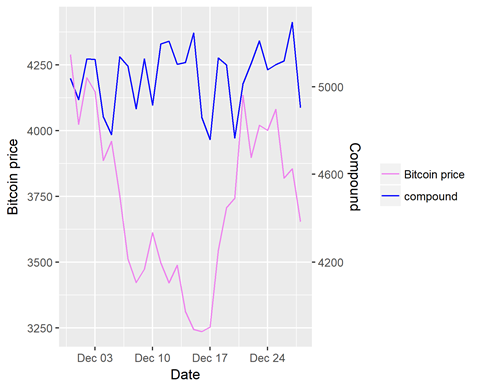
Boostings algorithms were also employed, specifically AdaBoost and Gradient boosting. In general, these boosting algorithms work by minimizing the error. Equation below illustrates how this procedure is done:

where E is the error during each iteration, and alpha\*h(x) is the weak learner for the classifier function. Each result is also weighted. When implementing gradient boosting, the model applies steepest descent (or gradient descent), updating the model by computing the derivative of the residuals (loss) and a multiplier,

The information regarding crypto was retrieved from a web-based platform known as Crypto Compare[[2]](#footnote-2), which provides historical prices for various cryptocurrencies. During data processing, data was time indexed, concatenated, and averaged. The many data models were employed to make predictions. The complete list of data models and their results is given in our results section.

As it was discussed earlier, we used the VADER[[3]](#footnote-3) sentiment analysis tool which is impressive in terms of its ability to be sensitive to nuances in the text, such as punctuation, capitalization, negation, and amplification of lexicon values. Data processing was the most time-consuming aspect of the research, in addition to variable transformation. All variables were included in the final model, given that they all had at least moderate correlation coefficients, and there was no logical reason to exclude them given their potential predictive ability.

IV. MEASURES OF FIT AND RESULTS

A bagging method of many different models was used to generate the final prediction. In this bagging method the result from different categories are collected and are either summed of averaged or the probability of their occurrence was identified. To reduce errors that can be occurred in one particular model, we found that having an ensemble method of learning was beneficial. Comparing linear regression and ensemble method we found that the latter performance was better compared to linear regression model. To test measures of goodness of fit we found that mean error and correlation coefficients were our measures of fit and the other measures of fit will indicate potential profit.

The correlation coefficient R2 was calculated from the set of testing data. In practical application, the full set of data minus the final target value should be trained. Only the final point, or the last unknown price value should be predicted. We are only interested in knowing how the final predicted value differs from the actual value. Therefore, another measure of fit is introduced, and shall be called +-T or dT, the error from our target value in dollars. This is the most useful measure of fit and establishes the potential to be profitable when trading crypto.

V. RESULTS

The hybrid of models that were used are shown below, including their measures of fit. Each model was run on testing data to gather the RSI and ME values, where ME is the mean error, R2 is the correlation coefficient, and +-T is the actual error when predicting the price on a brand-new data point (the final interval). Sampling at first was done in a 10 and 60-minute shifts. Overall, in this empirical experiment the 10-minute shifts results in less error in the hybrid model and it was chosen to be used within the experiments. Table 1-6 show the results provided by different methods we have used for Bitcoin, Ethereum, Electroneum, Monero, ZEC and Ripple.

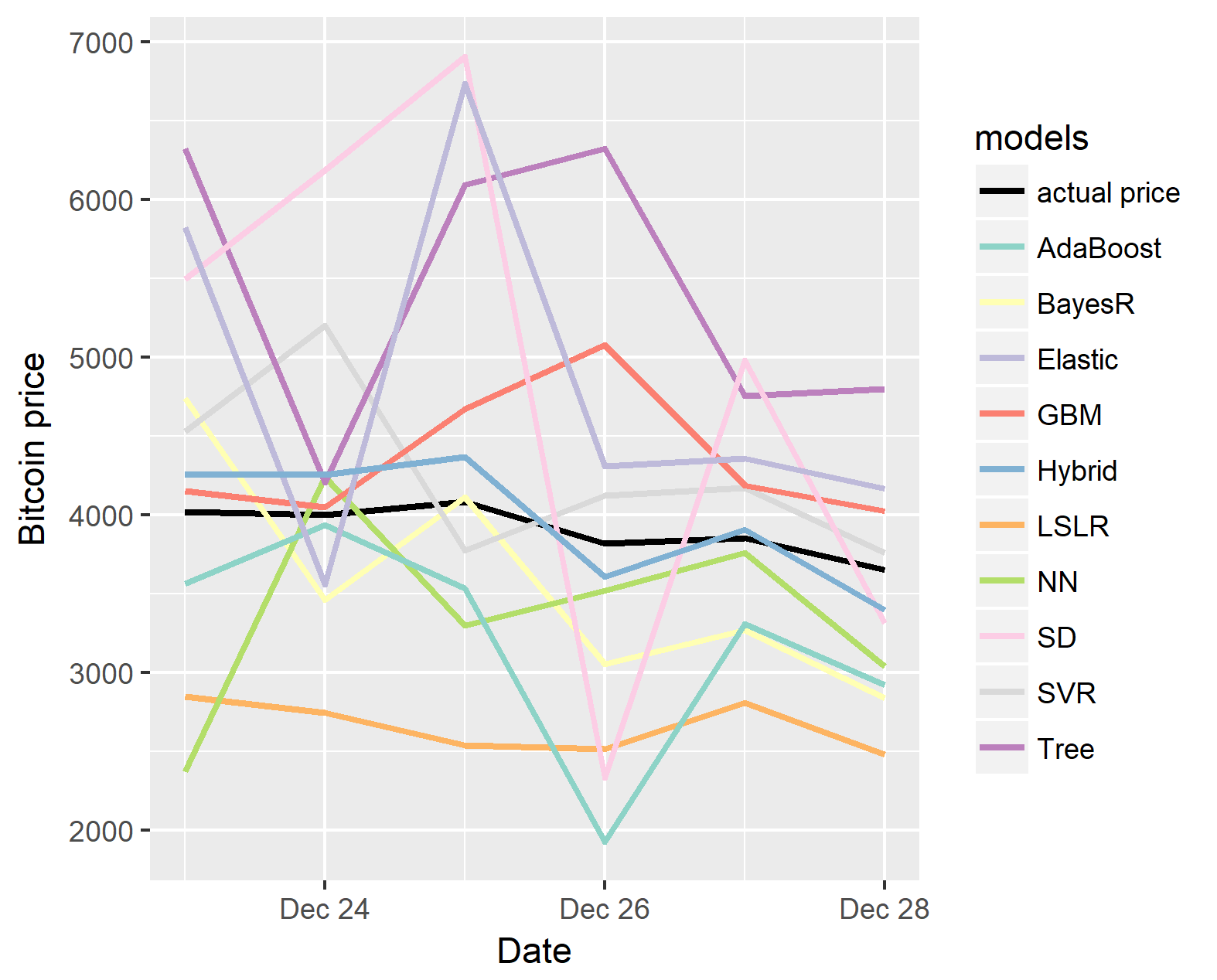
To be more precise the models we used were Support Vector Regression [13], Stochastic Gradient Descent [14], Gradient Boosting Model [15], MLP Neural Network [11], Least Squares Linear Regression [16], AdaBoost [17], Bayesian Ridge Regression [18], Decision Tree [19], ElasticNet [20] and Hybrid is mean of all of them. This mean is actually what was used for prediction. The tweet frequency was added as a transformed variable and graphed against the crypto prices. We found that tweets had a high inverse correlation with the price, as if bad news caused an increase in post frequency. This is illustrated in following Figures (1-25). On the figures we also compared the Google Trends data with Crypto Data as well as Tweet Frequency Data with Crypto Data. These results imply that there are meaningful relationships between those entities.

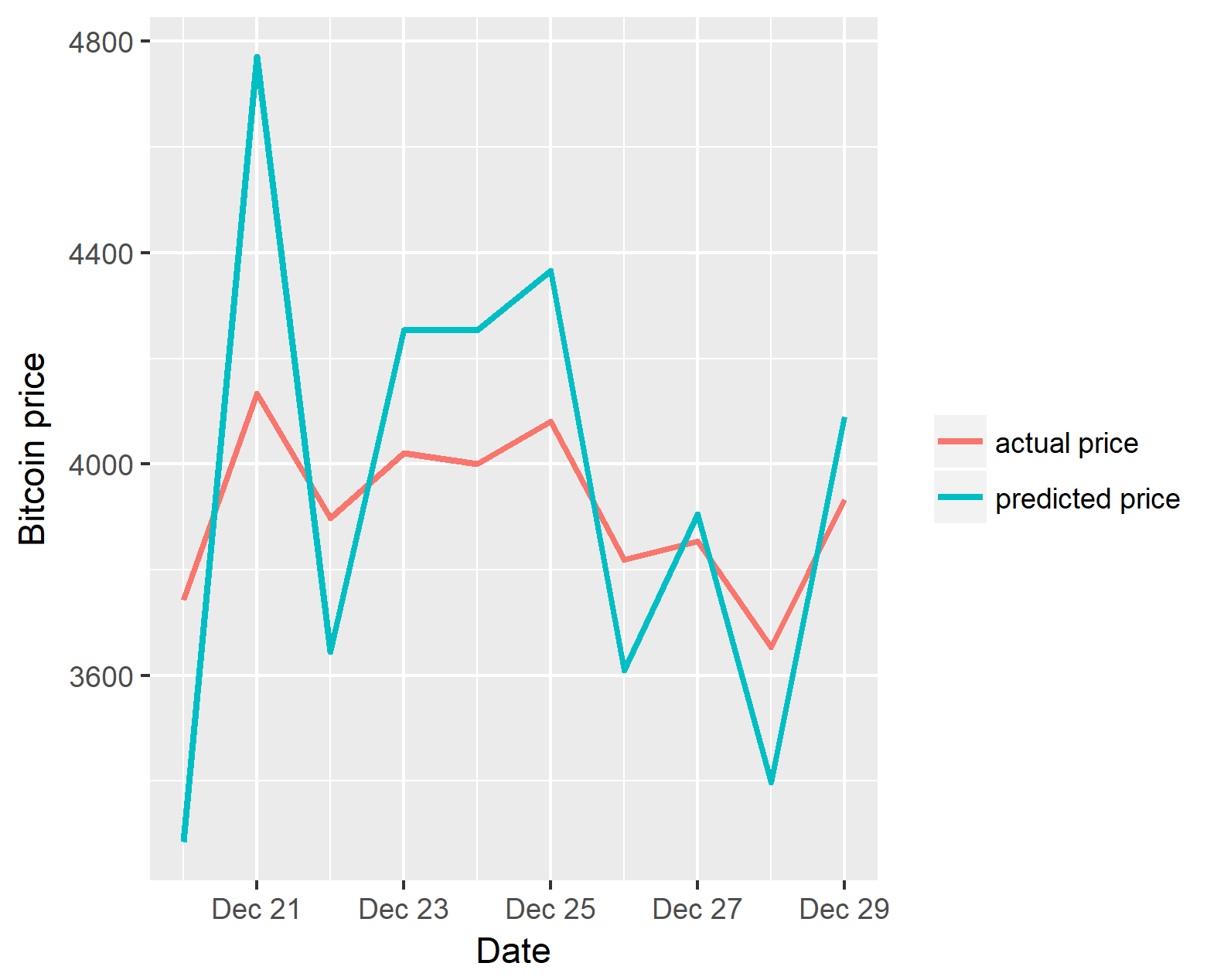
TABLE I

Model results for Bitcoin

|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 1357.482 | 0.706722 | -384.664 |
| Stochastic Gradient Descent | 8706.509 | 0.684718 | -254.258 |
| Gradient Boosting Model | 1370.471 | 0.704768 | -209.353 |
| MLP Neural Network | 1382.774 | 0.703728 | -117.398 |
| Least Squares Linear Regression | 2000.951 | 0.685587 | 395.1489 |
| AdaBoost | 1986.594 | 0.676574 | -5.40525 |
| Bayesian Ridge Regression | 1234.013 | 0.720673 | 48.95157 |
| Decision Tree | 8791.874 | 0.678843 | 359.0313 |
| ElasticNet | 2280.217 | 0.769856 | 313.6858 |
| Hybrid (Mean) | 498.6117 | 0.94169 | 151.6282 |

**FIGURE** **1**. **Bitcoin price vs number of tweets**



**FIGURE** **2**. **All models vs Bitcoin price**

**FIGURE** **3**. **Bitcoin price vs predicted price**

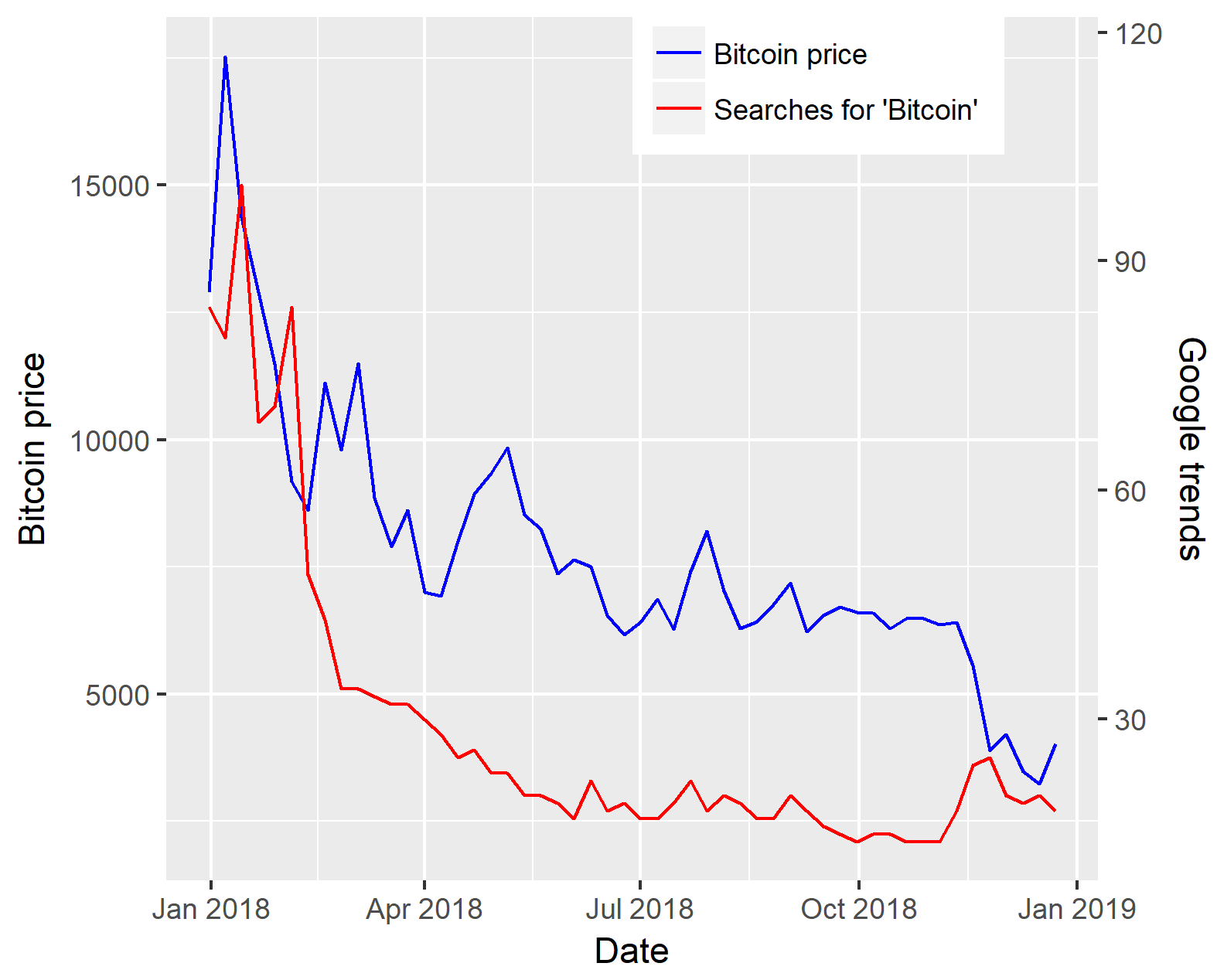
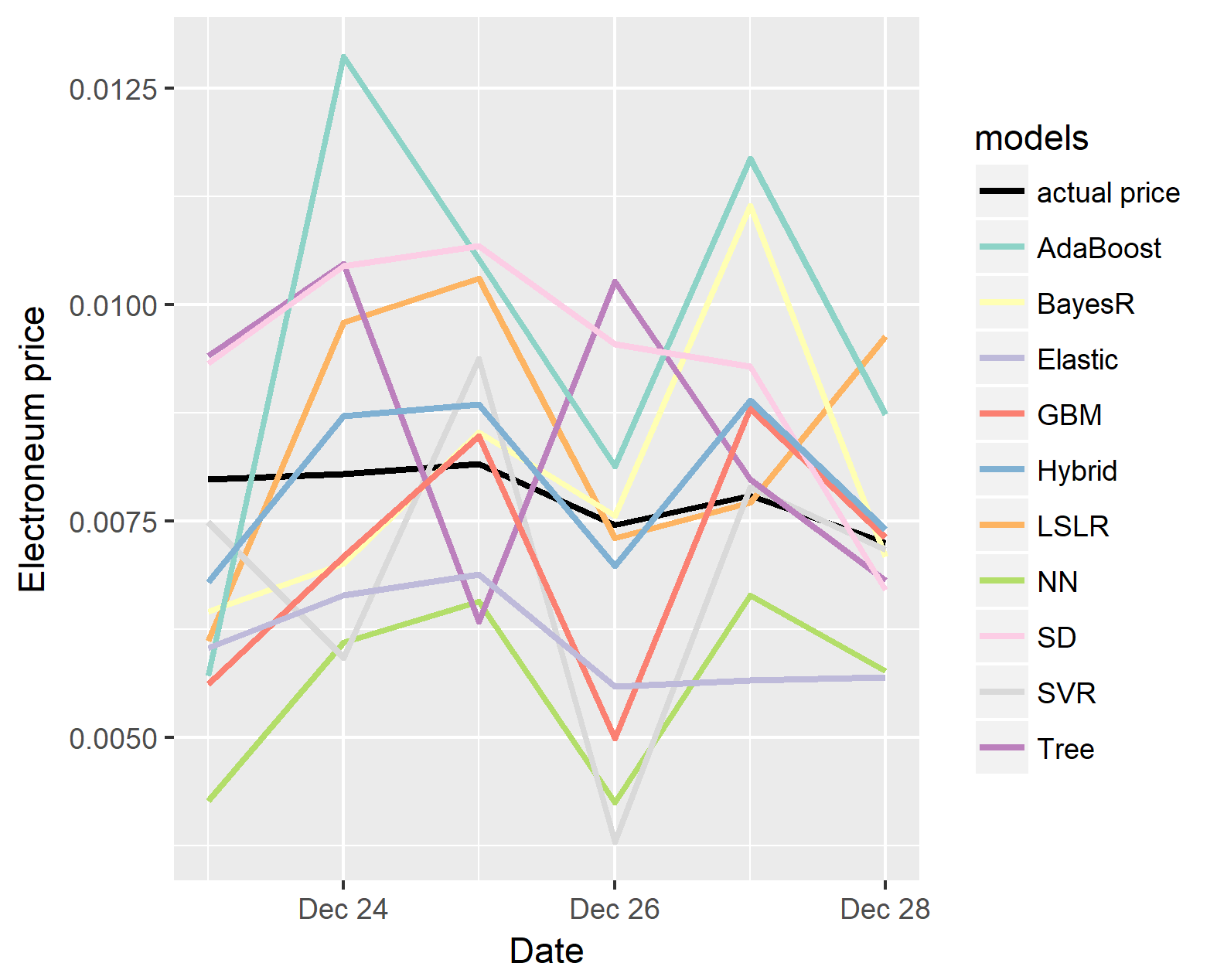
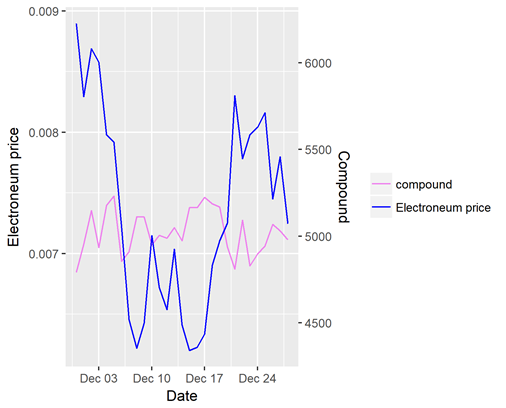
**FIGURE** **4**. **Bitcoin price vs Google trends**

TABLE II

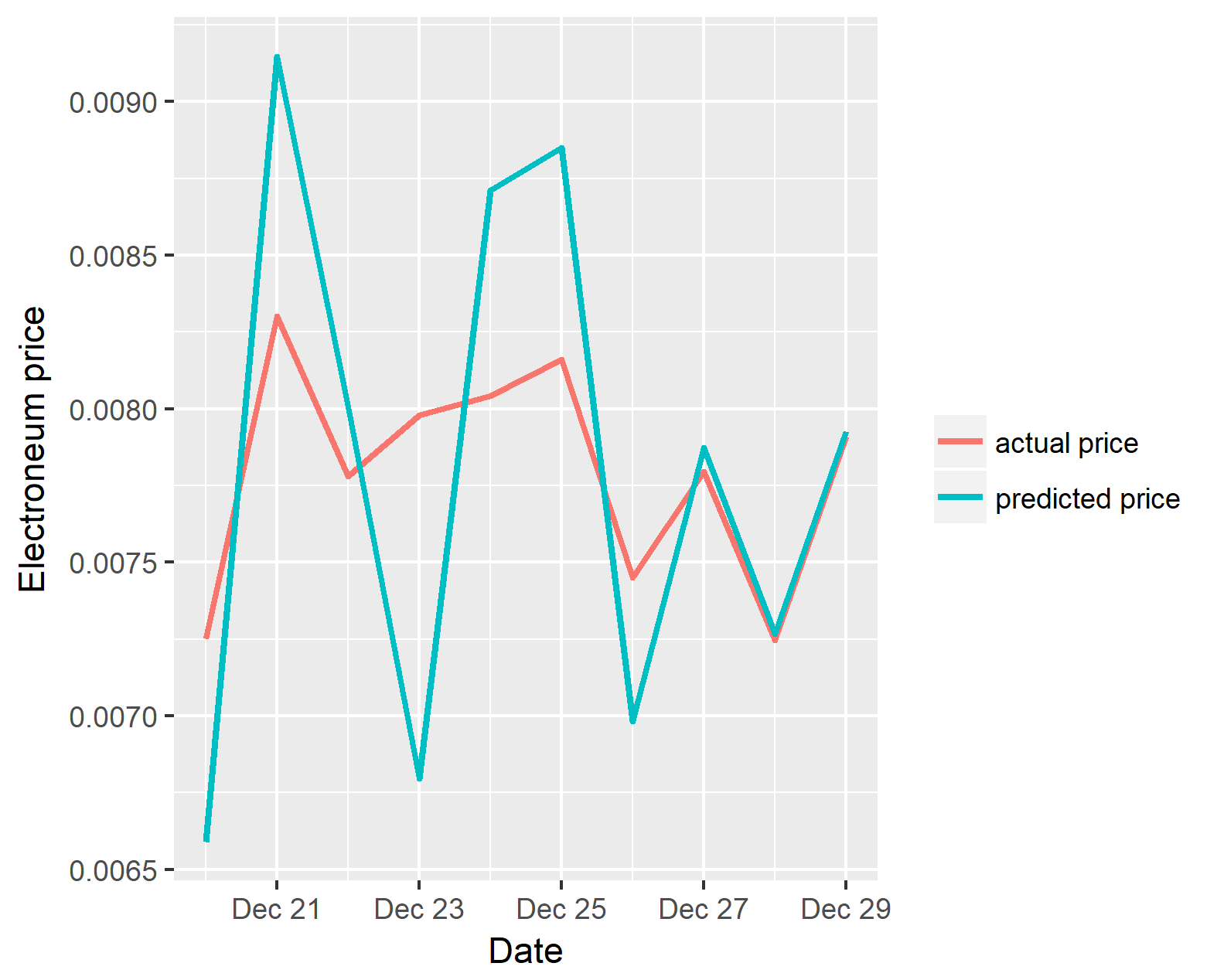
Model results for Electroneum

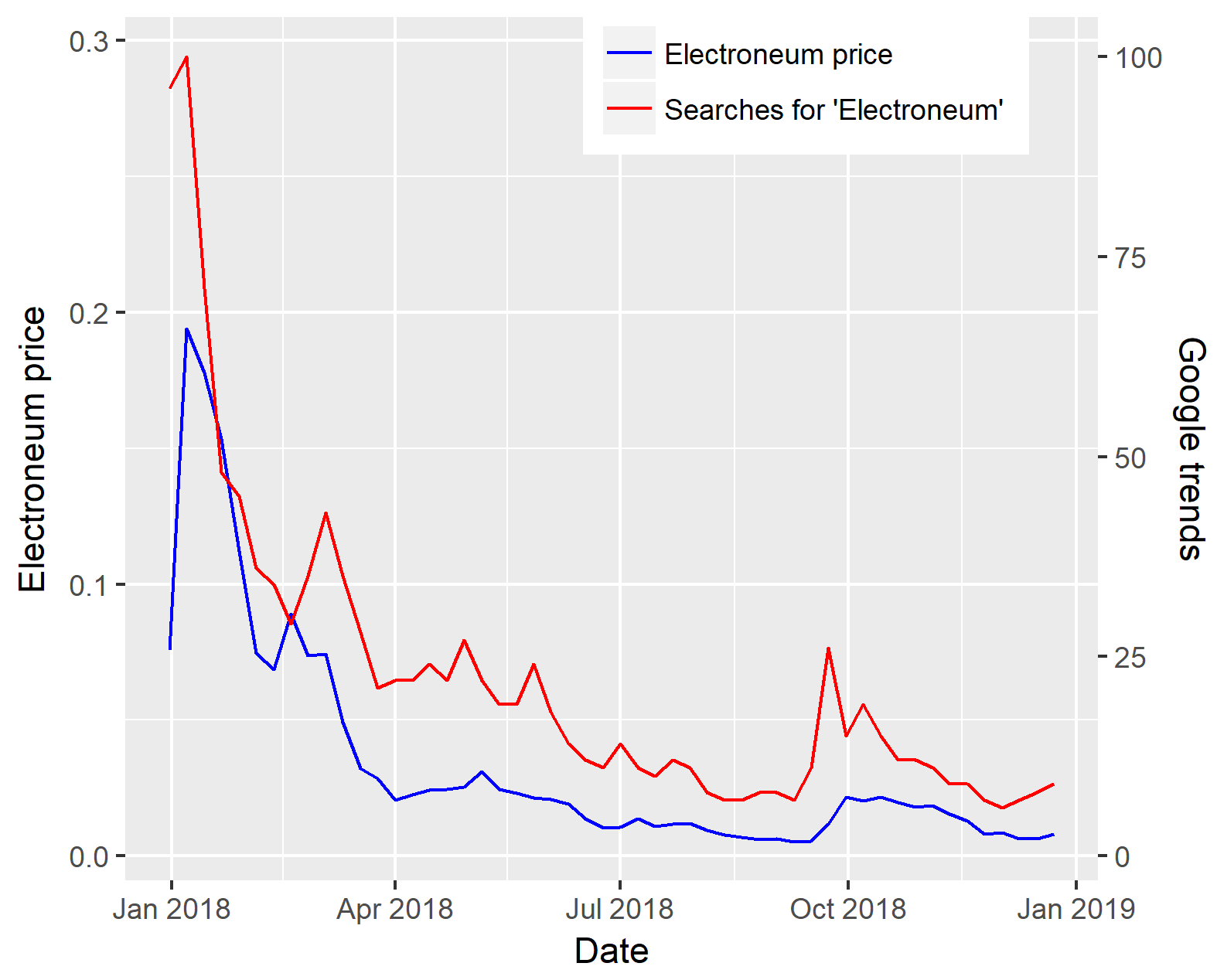
|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 0.004487 | 0.946137 | 0.000601 |
| Stochastic Gradient Descent | 0.036371 | 0.922049 | -0.00082 |
| Gradient Boosting Model | 0.005131 | 0.932071 | 0.000303 |
| MLP Neural Network | 0.008413 | 0.935374 | 0.000488 |
| Least Squares Linear Regression | 0.004657 | 0.953642 | 0.001144 |
| AdaBoost | 0.009122 | 0.927278 | -0.00137 |
| Bayesian Ridge Regression | 0.006355 | 0.898442 | 0.001081 |
| Decision Tree | 0.033998 | 0.952623 | 0.00044 |
| ElasticNet | 0.007629 | 0.9364 | -0.0008 |
| Hybrid (Mean) | 0.001842 | 0.99163 | 0.001 |



**FIGURE** **5**. **Electroneum price vs number of tweets**

**FIGURE** **6**. **All models vs Electroneum price**



**FIGURE** **7**. **Electroneum price vs predicted price**

**FIGURE** **8**. **Electroneum price vs Google trends**

TABLE III

Model results for Ethereum

|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 75.8912 | 0.964188 | -17.1537 |
| Stochastic Gradient Descent | 364.9388 | 0.966607 | -12.644 |
| Gradient Boosting Model | 126.1168 | 0.968592 | -3.84195 |
| MLP Neural Network | 45.18472 | 0.964776 | -12.367 |
| Least Squares Linear Regression | 126.2375 | 0.963377 | -3.38566 |
| AdaBoost | 73.22109 | 0.969224 | 16.75794 |
| Bayesian Ridge Regression | 73.57855 | 0.964501 | -2.00044 |
| Decision Tree | 615.5554 | 0.961733 | 17.85768 |
| ElasticNet | 124.8832 | 0.968191 | -6.17257 |
| Hybrid (Mean) | 16.02903 | 0.994549 | 7.823218 |

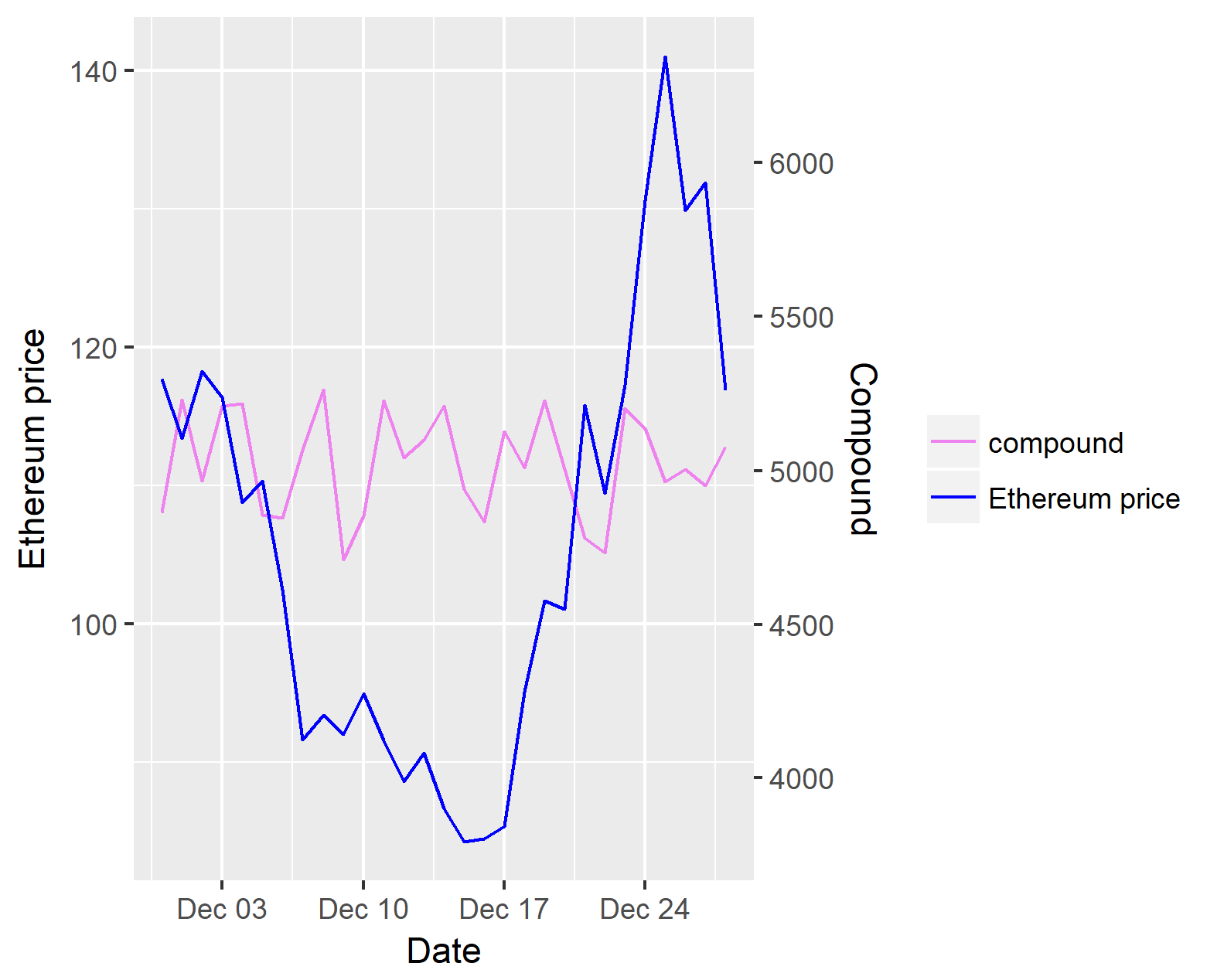


FIGURE 9.  Ethereum price vs number of tweets.

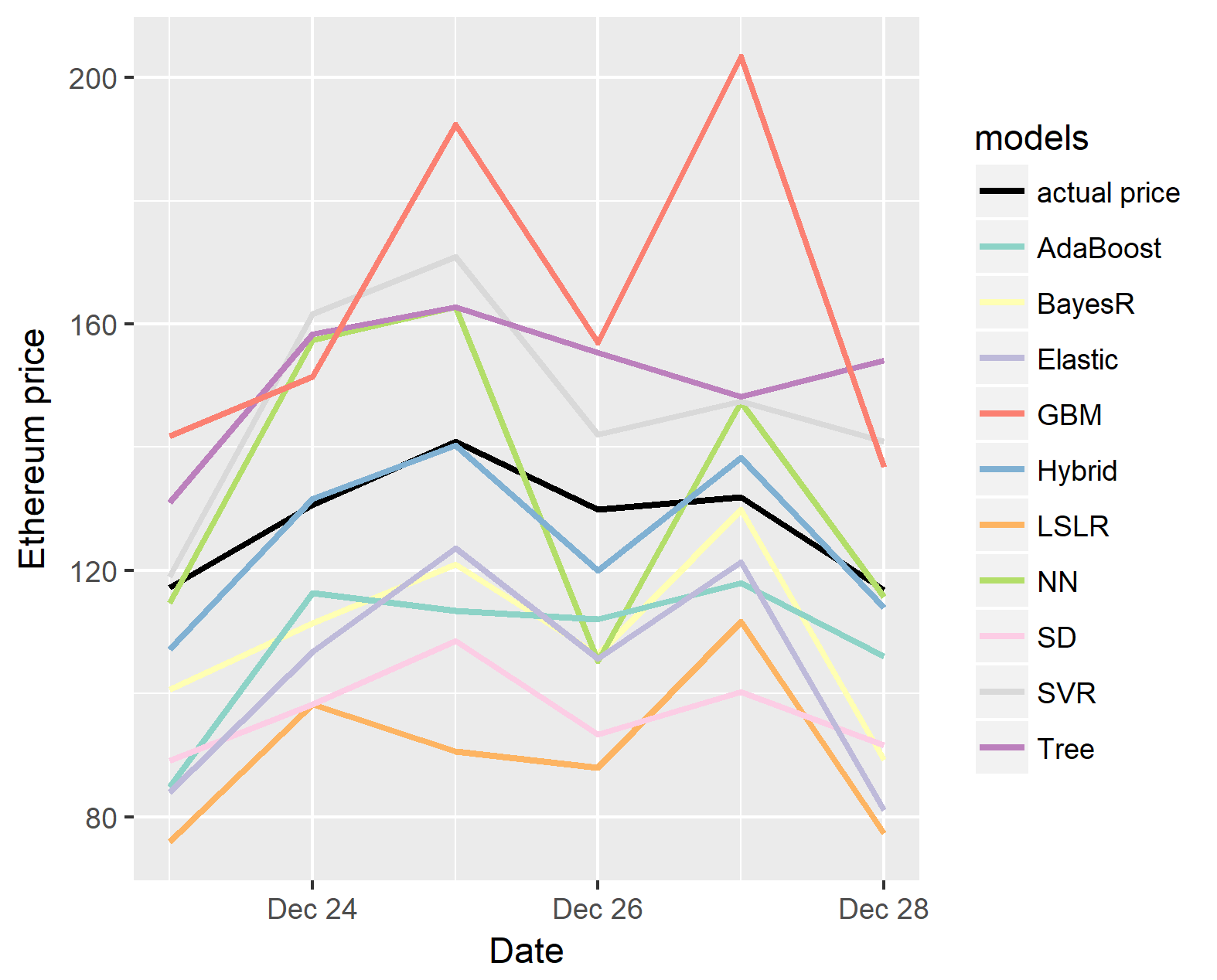


FIGURE 10.  All models vs Ethereum price.

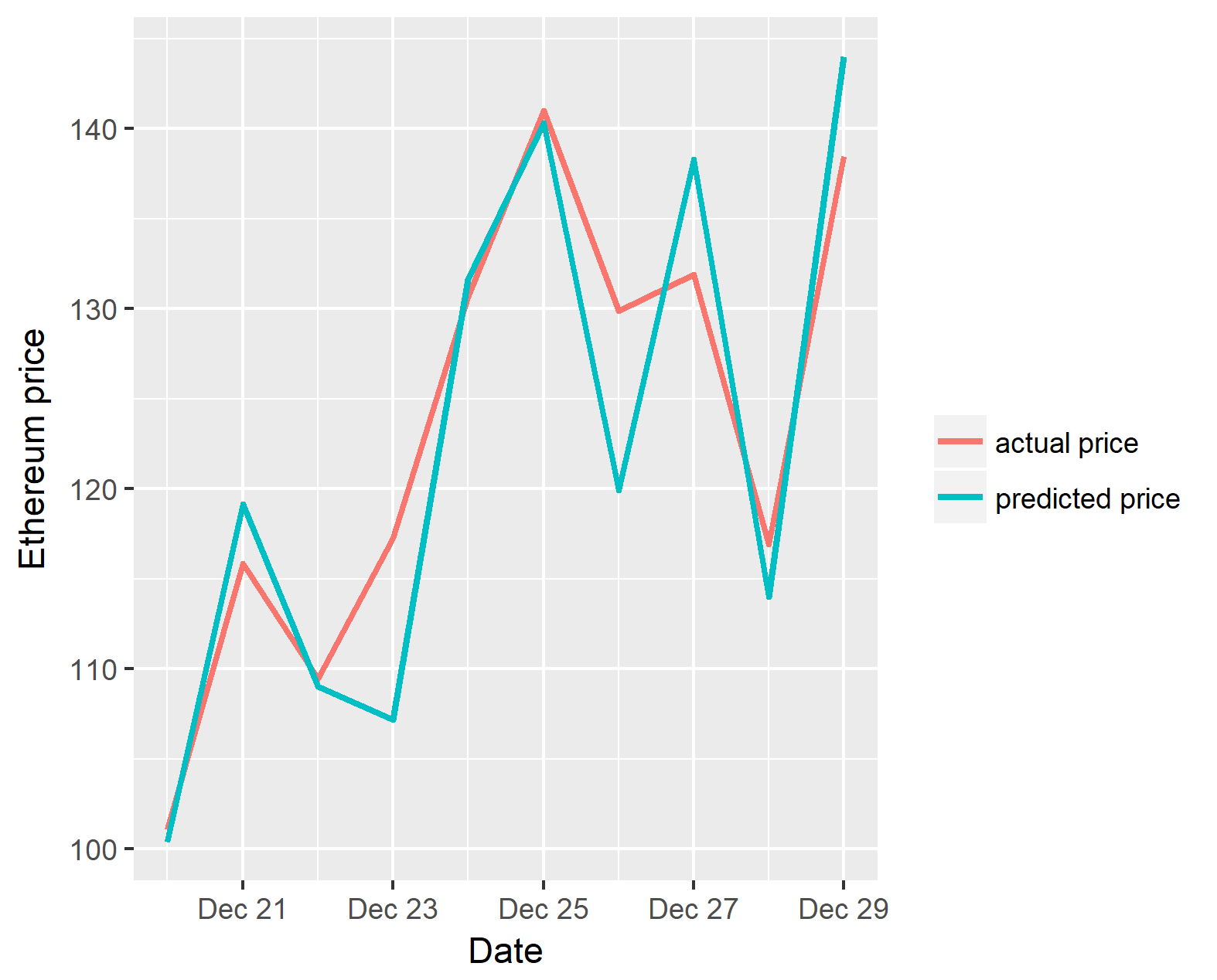


FIGURE 11.  Ethereum price vs predicted price.

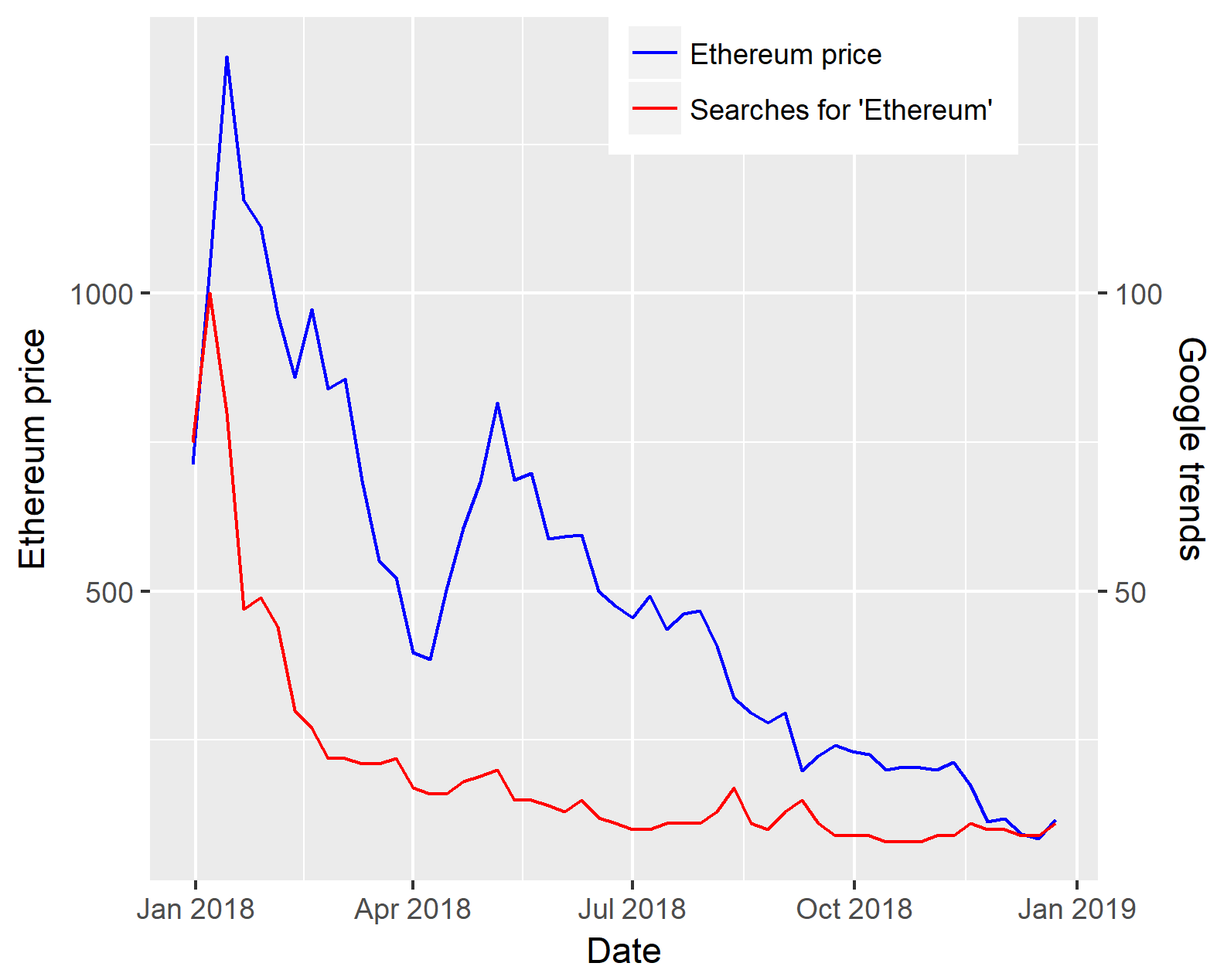


FIGURE 12.  Ethereum price vs Google Trends.

TABLE IV

Model results for Monero

|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 32.21549 | 0.839515 | -2.96132 |
| Stochastic Gradient Descent | 194.4944 | 0.815937 | -2.35604 |
| Gradient Boosting Model | 45.53463 | 0.843566 | -4.13285 |
| MLP Neural Network | 32.52362 | 0.845643 | 8.301083 |
| Least Squares Linear Regression | 26.08474 | 0.862746 | 4.746065 |
| AdaBoost | 30.88287 | 0.859943 | 2.887843 |
| Bayesian Ridge Regression | 50.84047 | 0.84862 | 3.112188 |
| Decision Tree | 196.3985 | 0.859415 | 3.559582 |
| ElasticNet | 44.01616 | 0.856563 | -3.38396 |
| Hybrid (Mean) | 13.79168 | 0.978258 | -8.01742 |

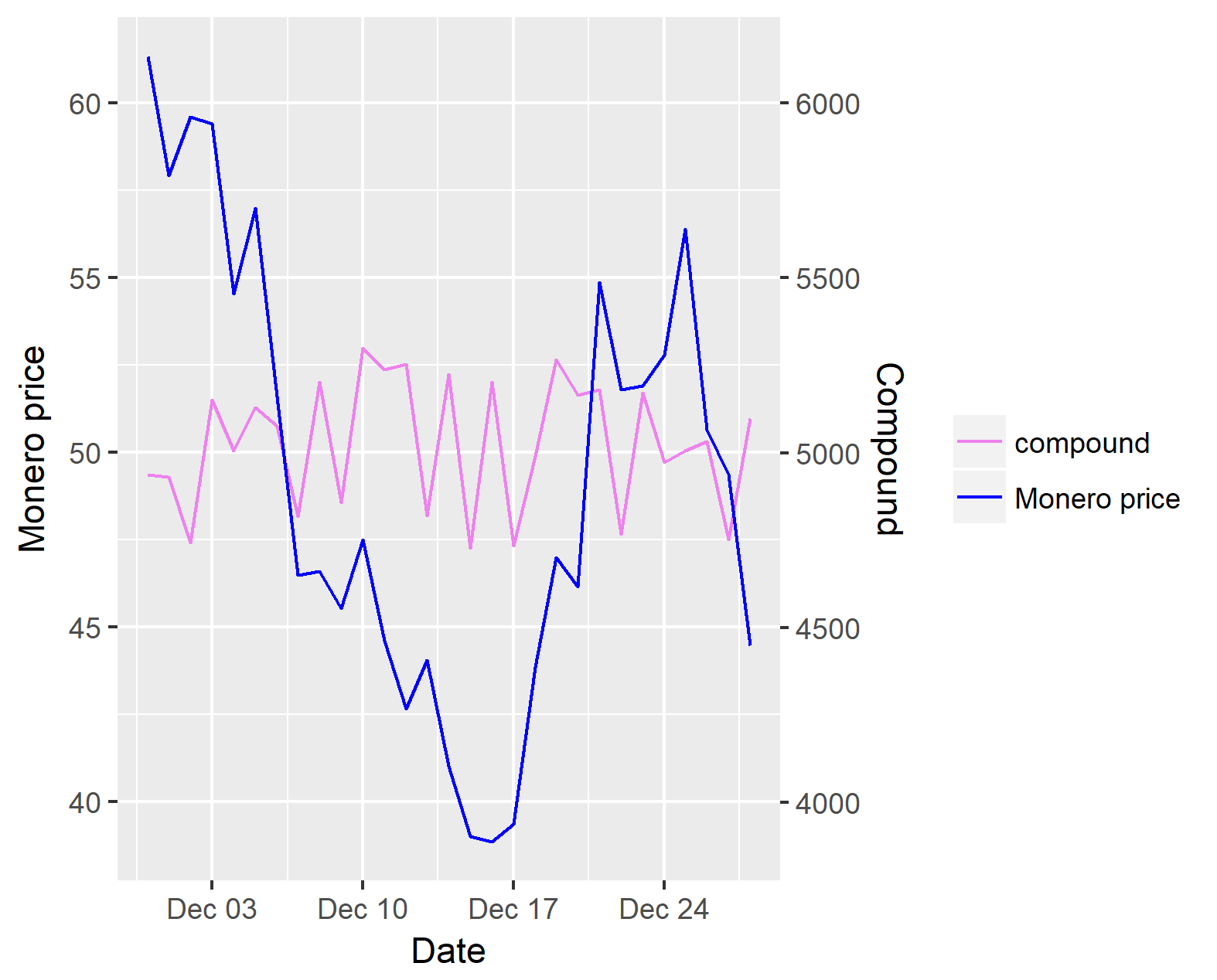


FIGURE 13.  Monero price vs number of tweets.

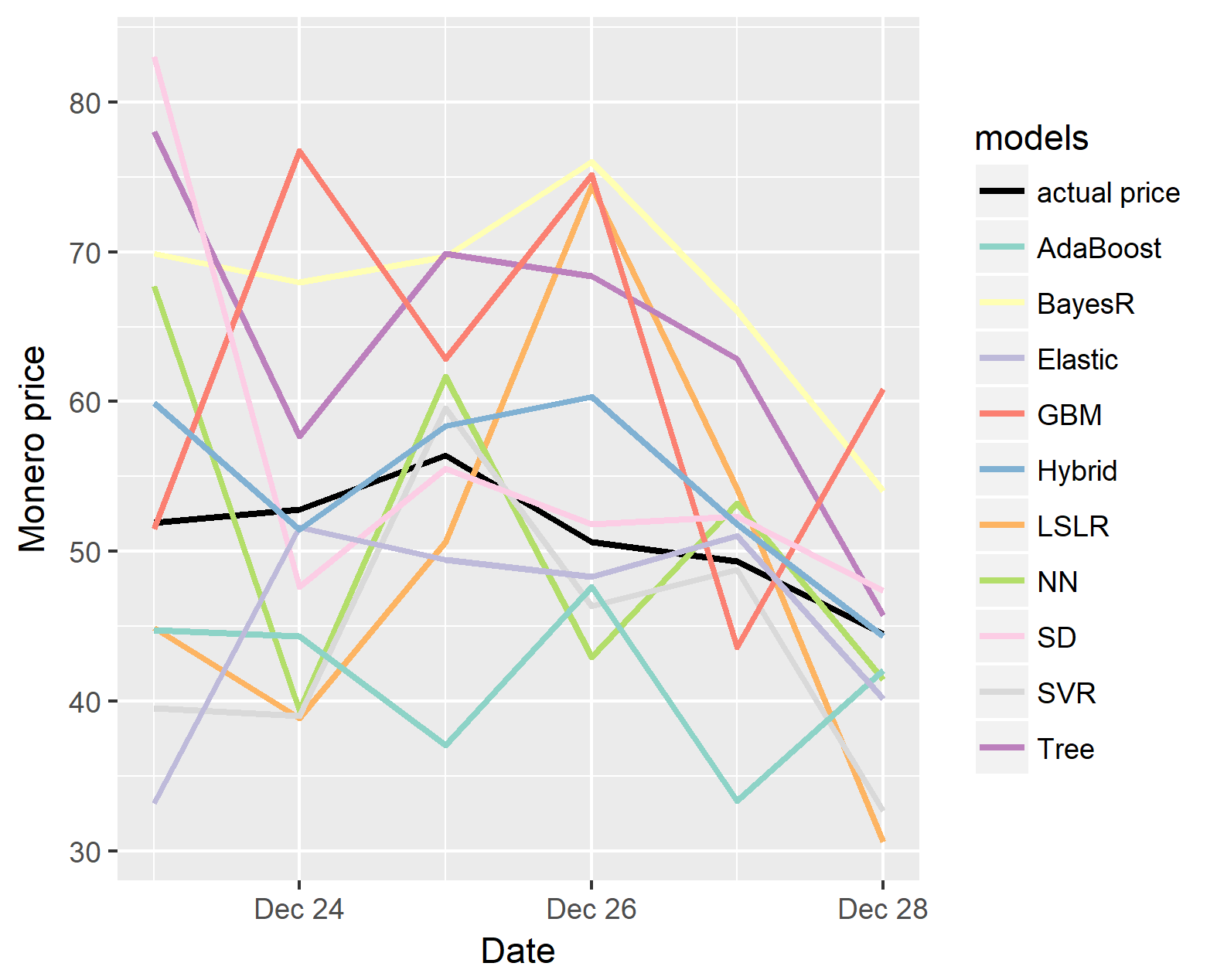


FIGURE 14.  All models vs Monero price.

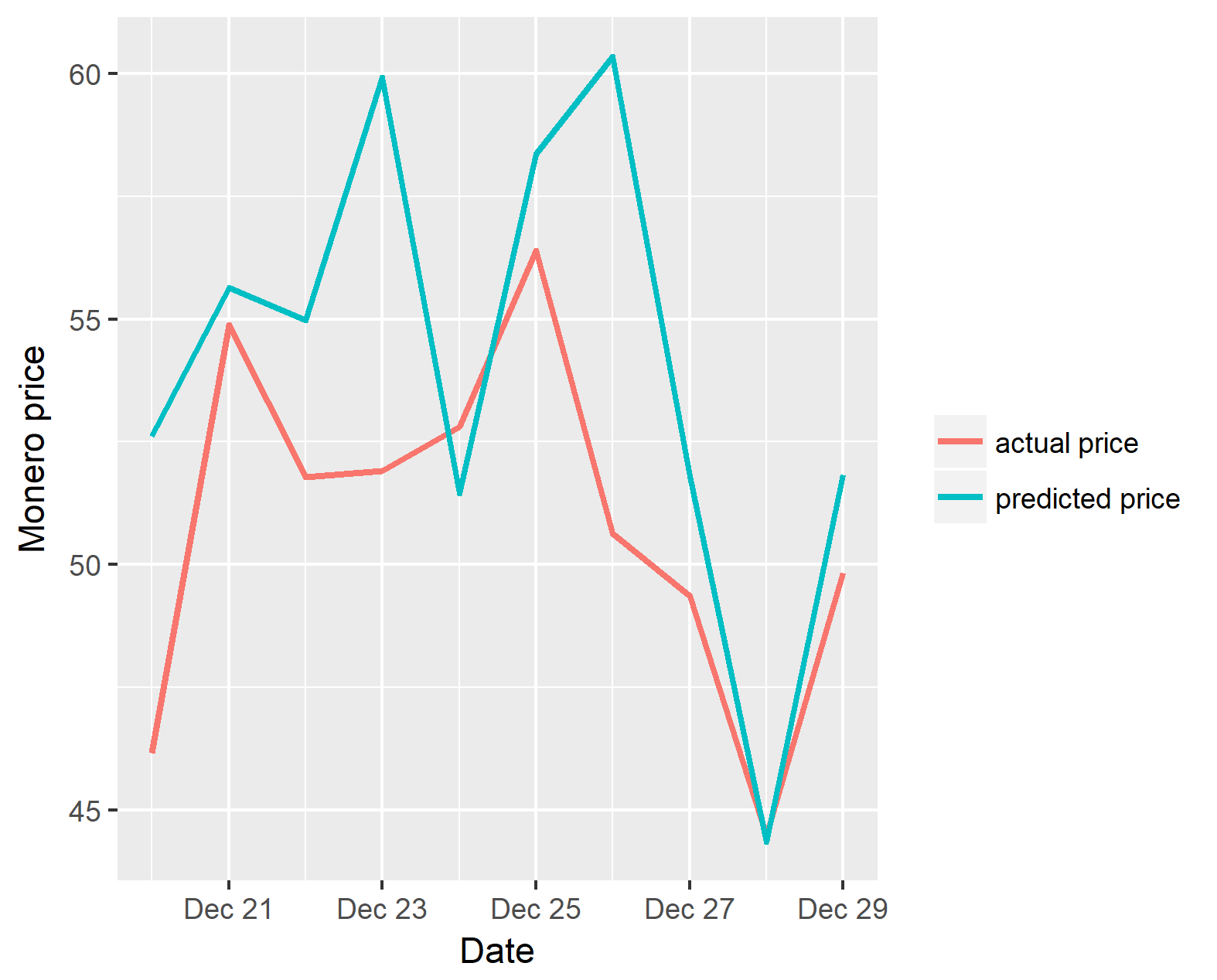


FIGURE 15.  Monero price vs predicted price.

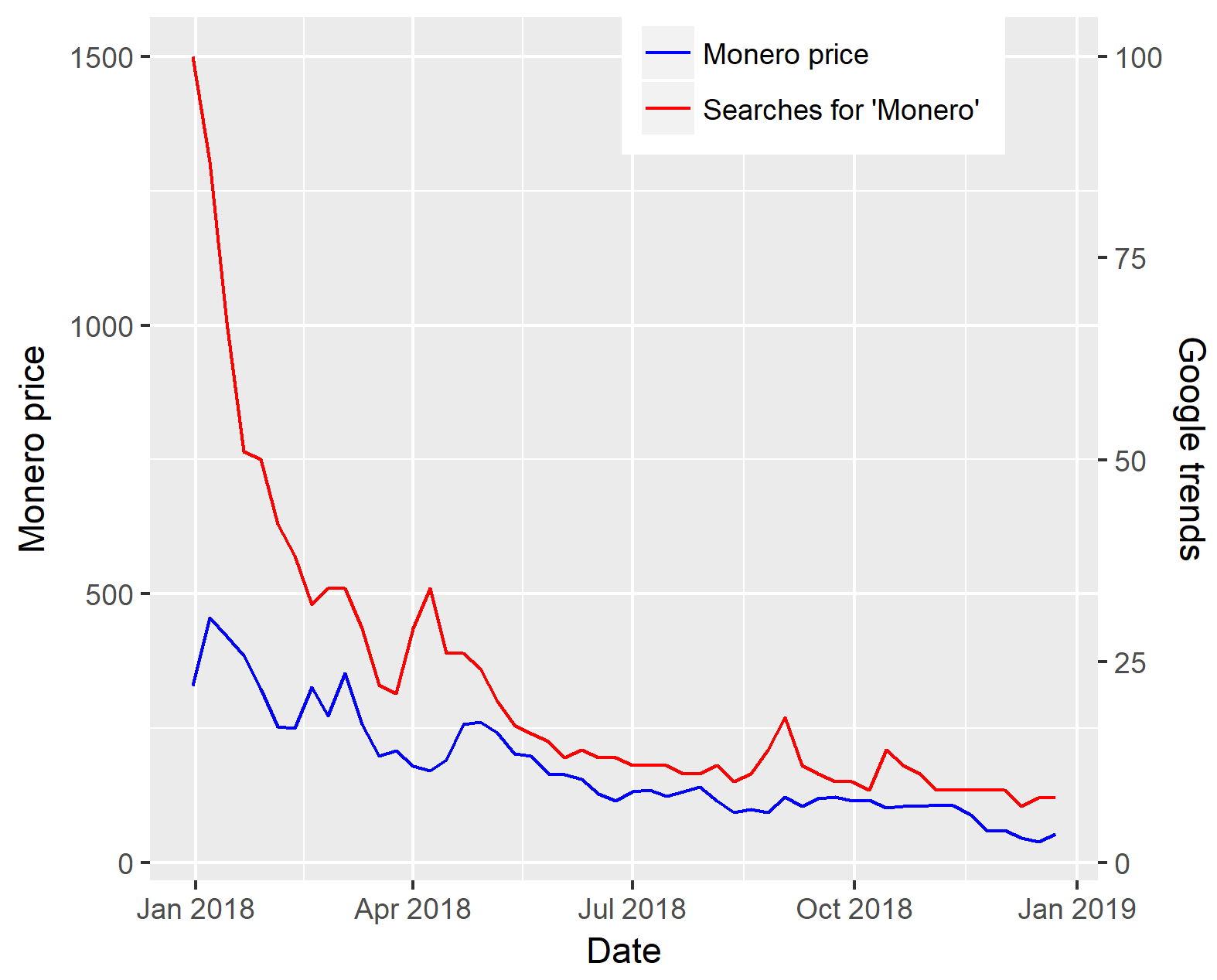


FIGURE 16.  Ethereum price vs Google Trends.

TABLE V

Model results for Ripple

|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 32.21549 | 0.839515 | -2.96132 |
| Stochastic Gradient Descent | 194.4944 | 0.815937 | -2.35604 |
| Gradient Boosting Model | 45.53463 | 0.843566 | -4.13285 |
| MLP Neural Network | 32.52362 | 0.845643 | 8.301083 |
| Least Squares Linear Regression | 26.08474 | 0.862746 | 4.746065 |
| AdaBoost | 30.88287 | 0.859943 | 2.887843 |
| Bayesian Ridge Regression | 50.84047 | 0.84862 | 3.112188 |
| Decision Tree | 196.3985 | 0.859415 | 3.559582 |
| ElasticNet | 44.01616 | 0.856563 | -3.38396 |
| Hybrid (Mean) | 13.79168 | 0.978258 | -8.01742 |

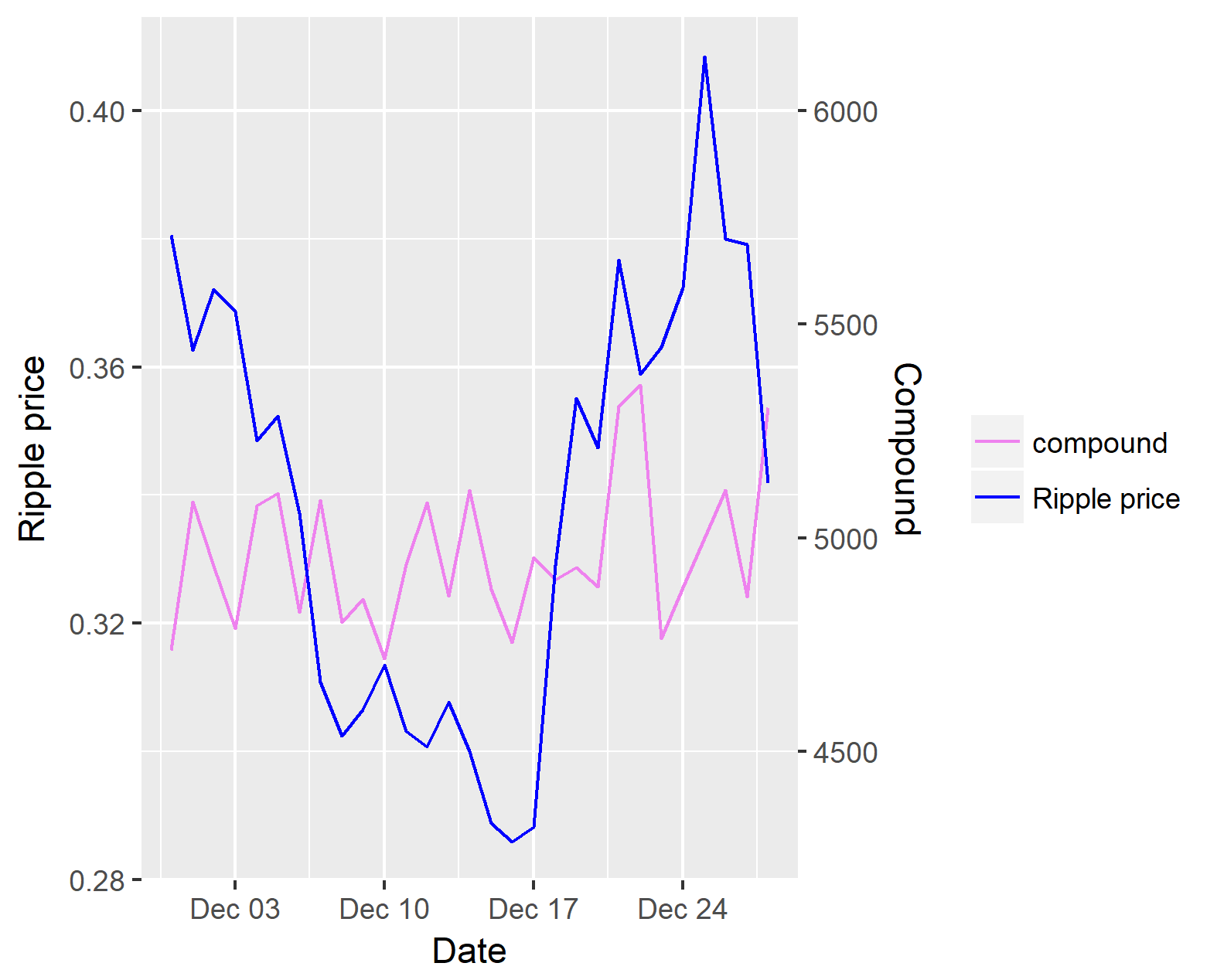


FIGURE 17.  Ripple price vs number of tweets.

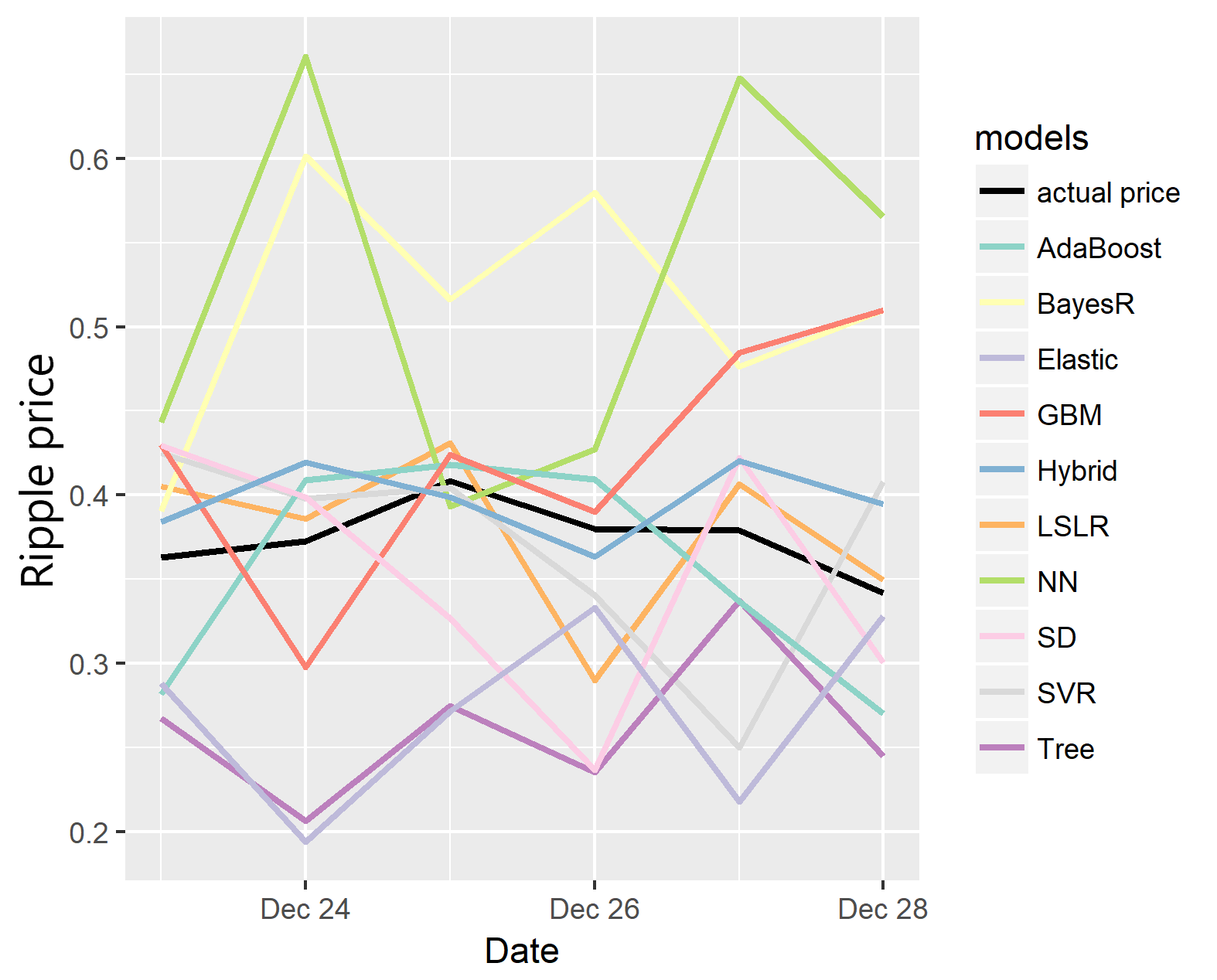


FIGURE 18.  All models vs Ripple price.

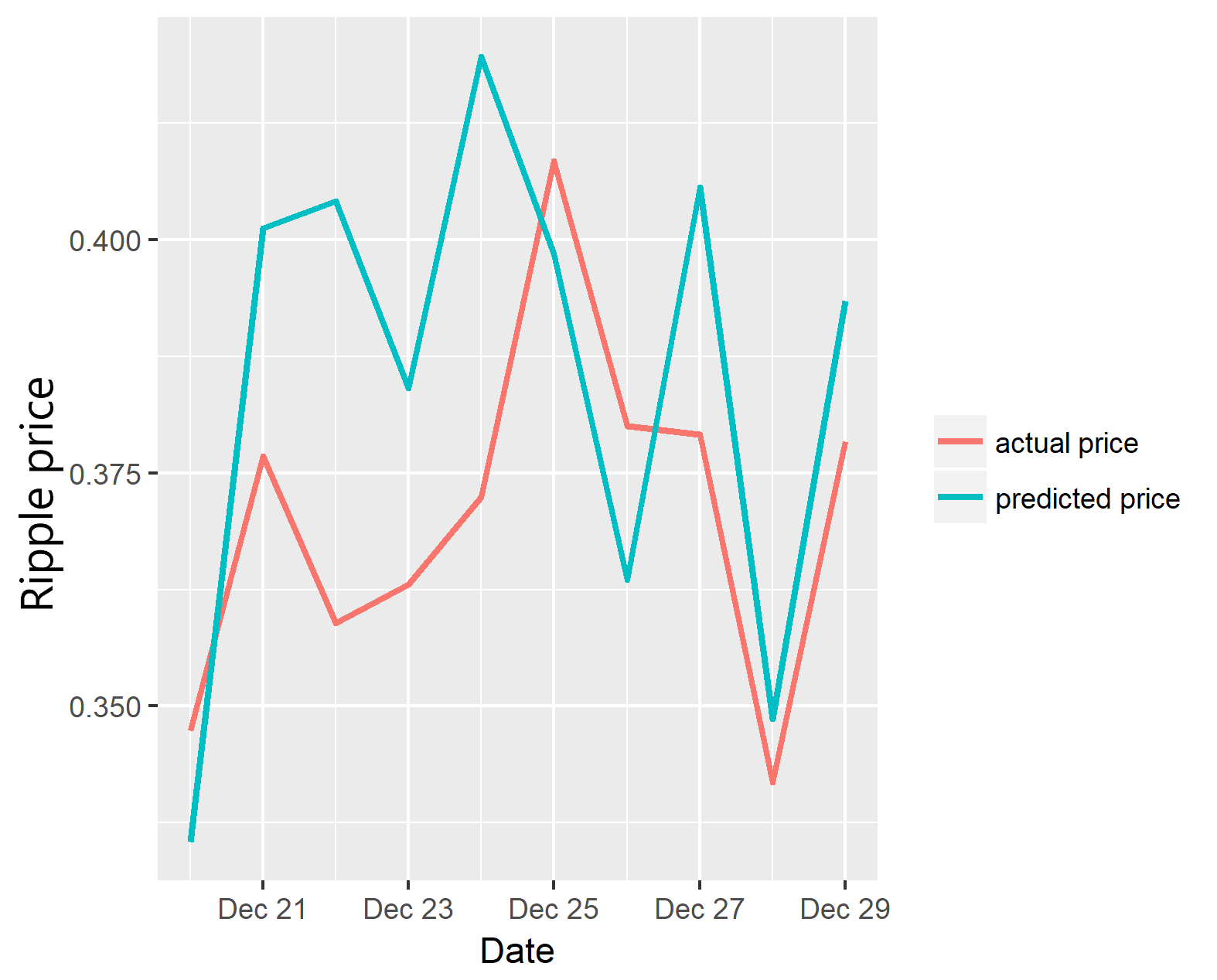


FIGURE 19.  Ripple price vs predicted price.

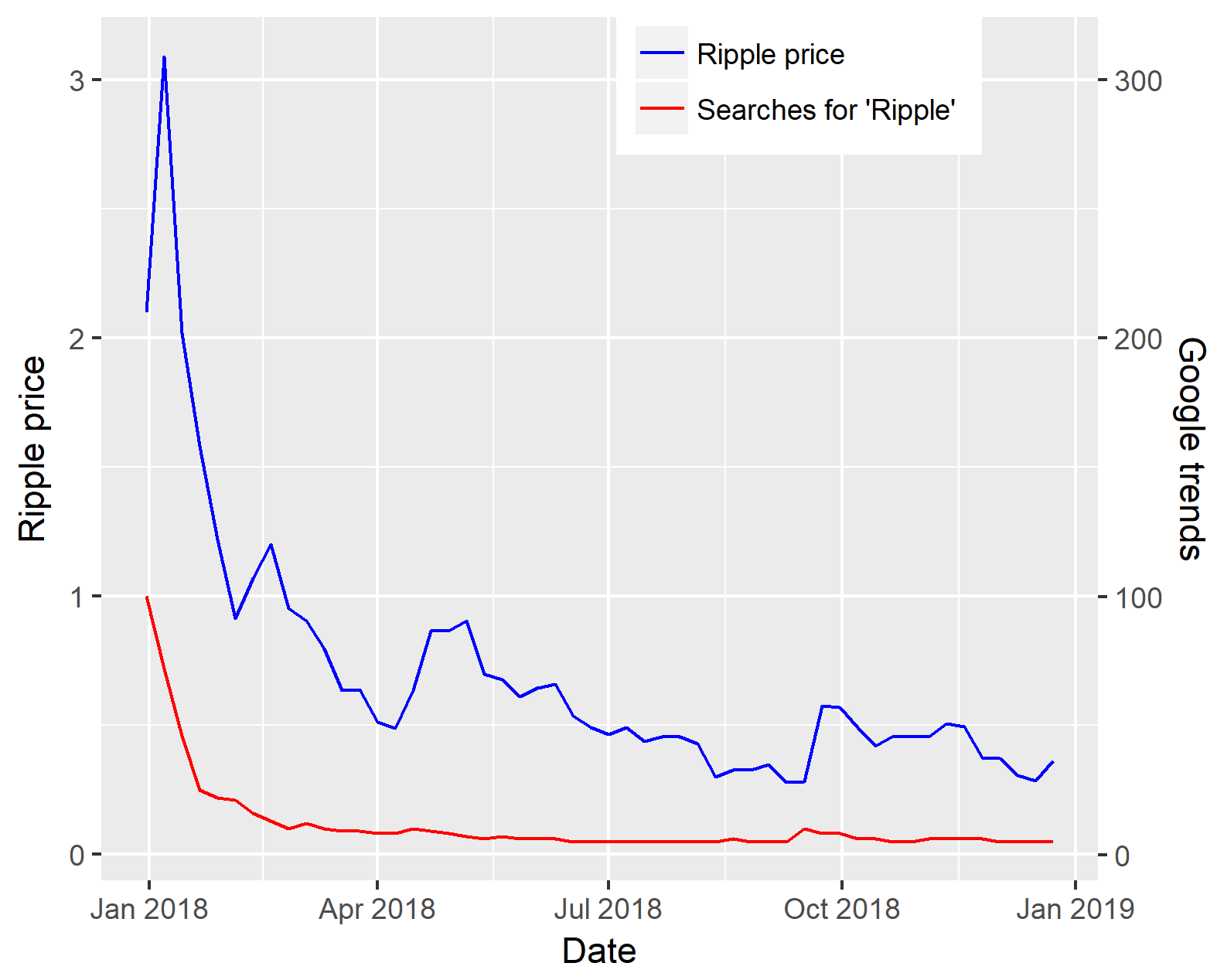


FIGURE 20.  Ripple price vs Google Trends.

TABLE VI

Model results for Zcash

|  |  |  |  |
| --- | --- | --- | --- |
| Models | ME | R² | T.s. |
| Support Vector Regression | 32.21549 | 0.839515 | -2.96132 |
| Stochastic Gradient Descent | 194.4944 | 0.815937 | -2.35604 |
| Gradient Boosting Model | 45.53463 | 0.843566 | -4.13285 |
| MLP Neural Network | 32.52362 | 0.845643 | 8.301083 |
| Least Squares Linear Regression | 26.08474 | 0.862746 | 4.746065 |
| AdaBoost | 30.88287 | 0.859943 | 2.887843 |
| Bayesian Ridge Regression | 50.84047 | 0.84862 | 3.112188 |
| Decision Tree | 196.3985 | 0.859415 | 3.559582 |
| ElasticNet | 44.01616 | 0.856563 | -3.38396 |
| Hybrid (Mean) | 13.79168 | 0.978258 | -8.01742 |

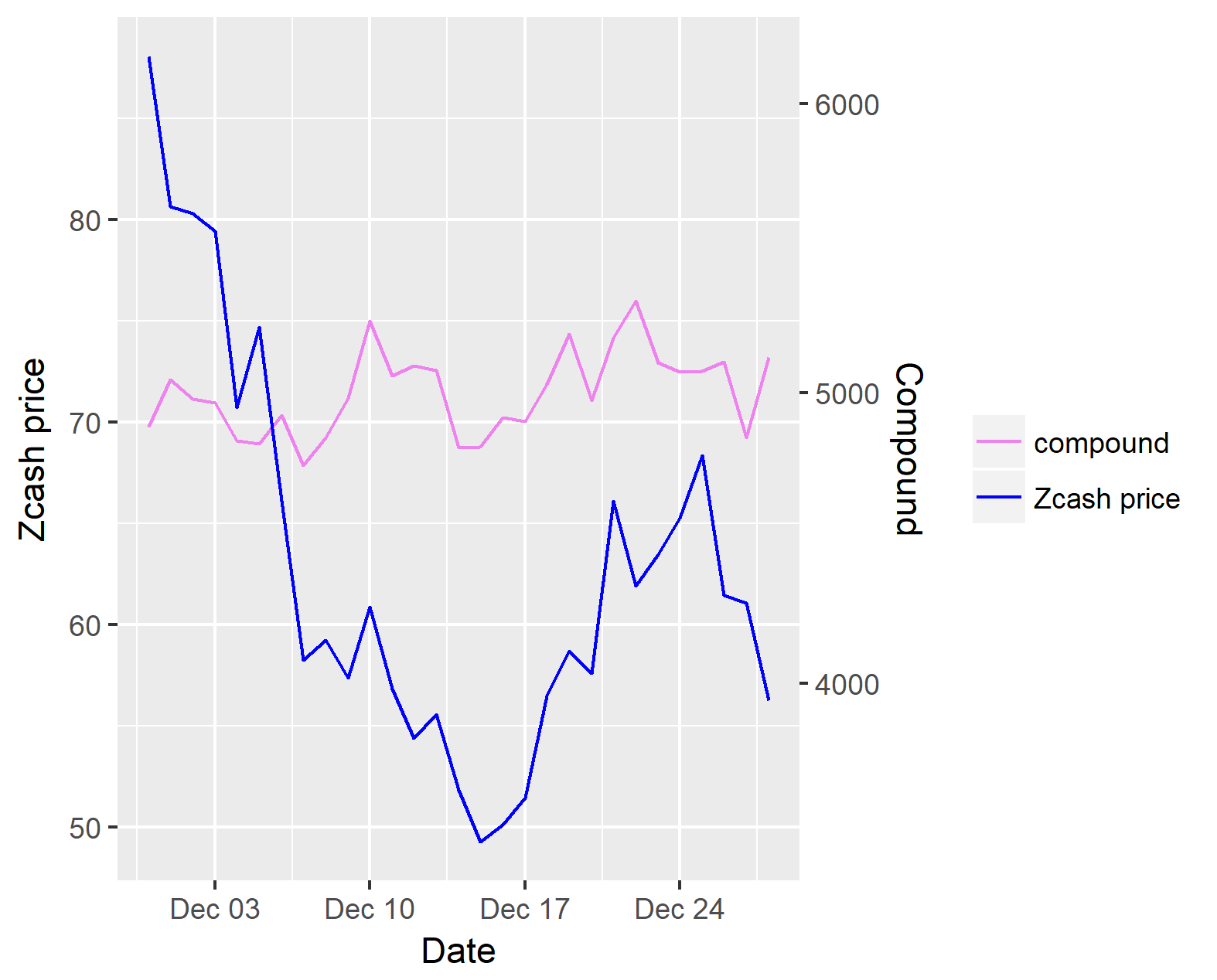


FIGURE 21.  Zcash price vs number of tweets.

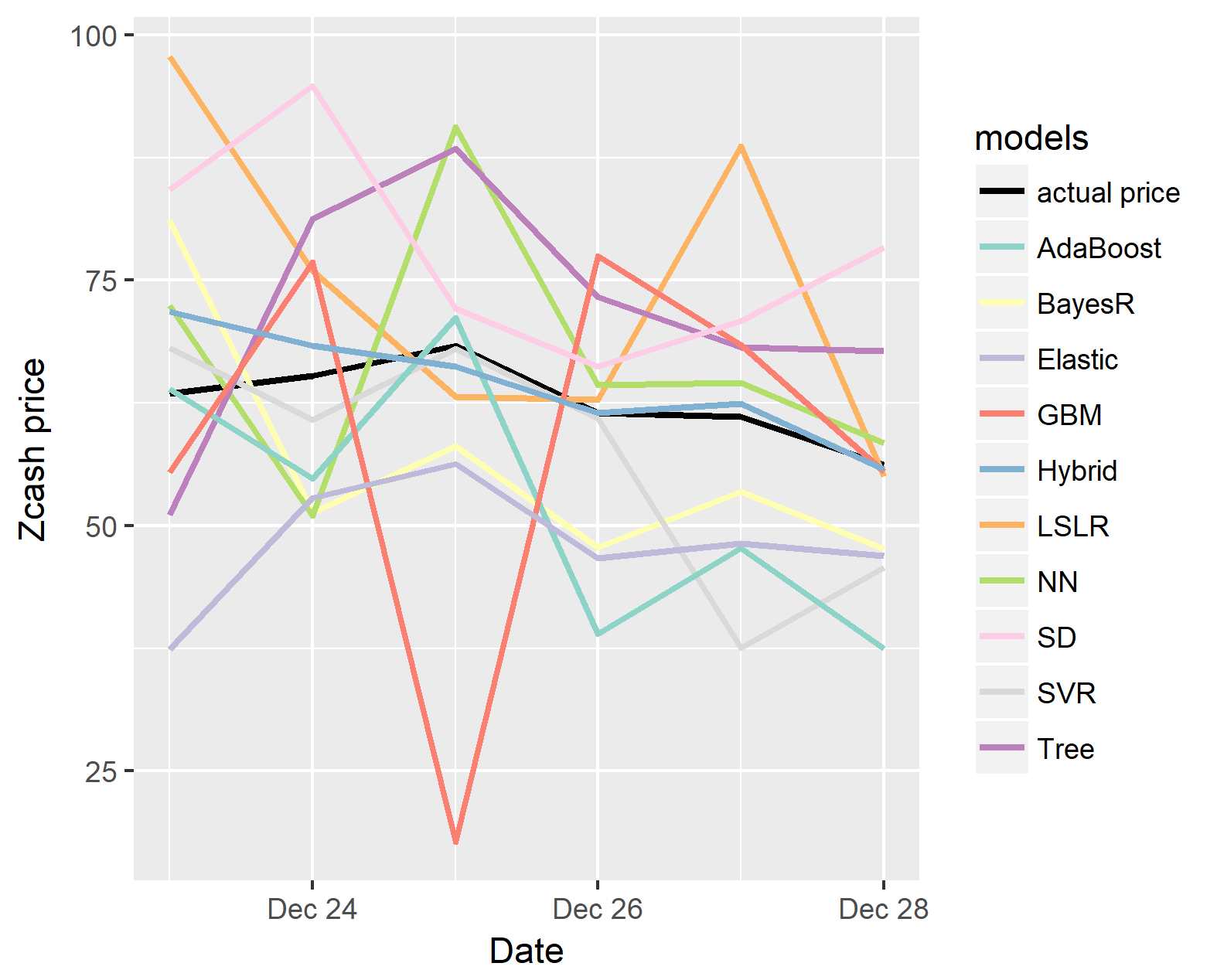


FIGURE 22.  All models vs Zcash price.

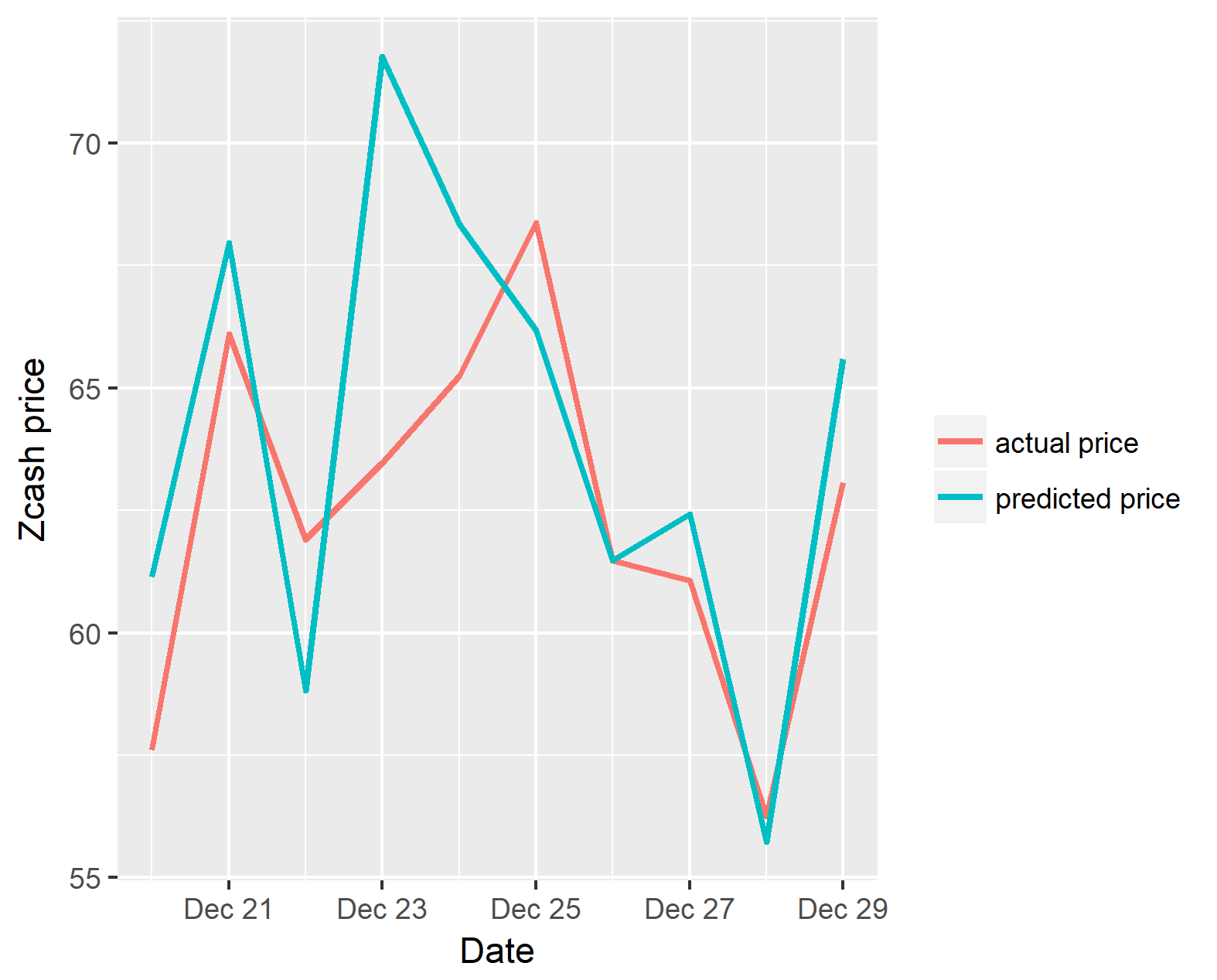


FIGURE 23.  Zcash price vs predicted price.

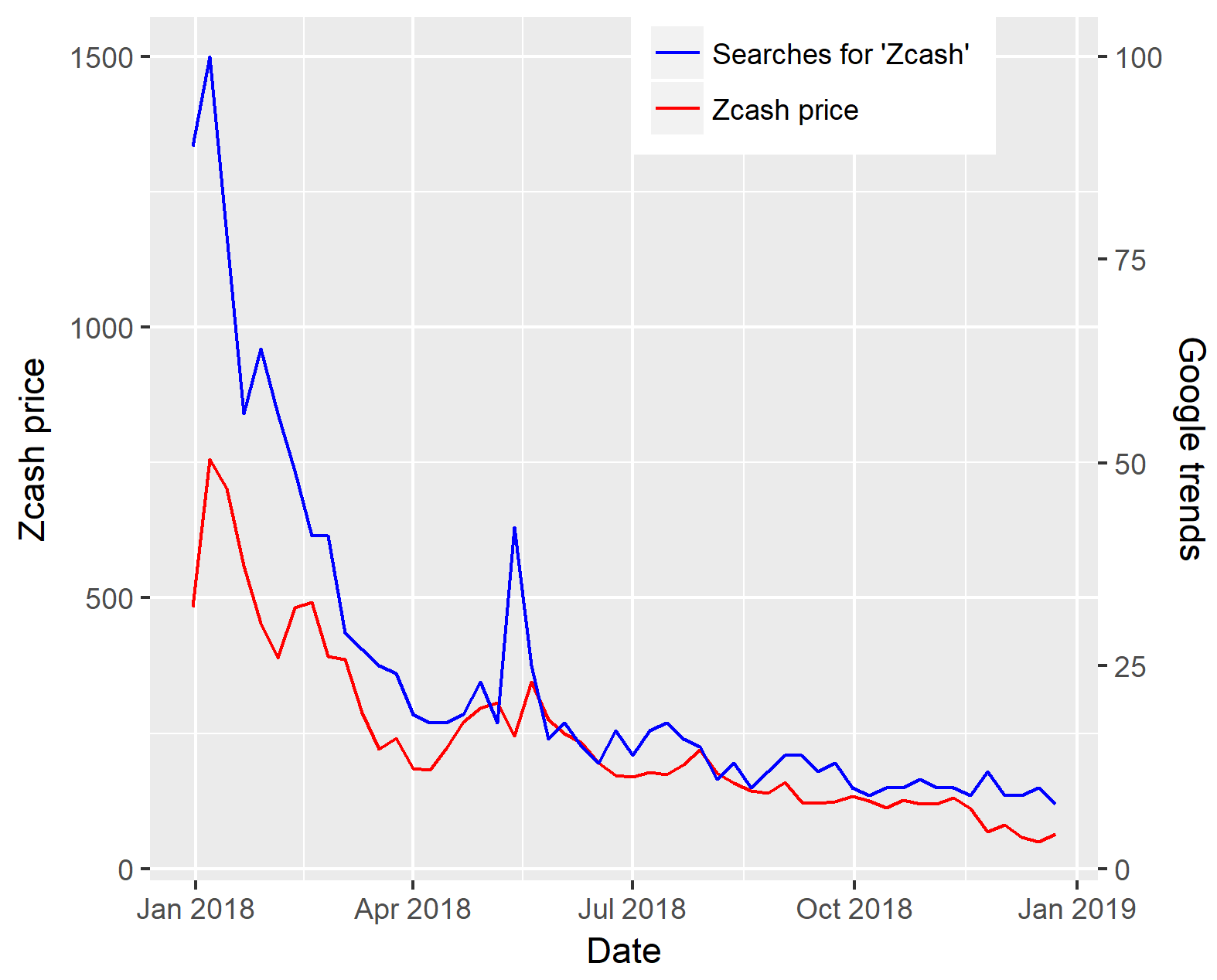


FIGURE 24.  Zcash price vs Google Trends.

From the above analysis which was performed by comparing sentiment analysis on twitter data against crypto prices for a certain time frame. We showed on the figures for each crypto in analysis the comparison and experiments, the performance of the models against testing data, the results of these models for a interval of 10 minutes and the implementation of our model against the full data set. The last value in our predicted values represents the true performance of the model when making a final prediction. We also showed separately the results of the hybrid model. The best result gives an error of less than $6, when predicting the final price point. The volatility of crypto has resulted in daily price swings that are much greater than our total error. This suggest the model could be profitable.

Finally, we did an empirical experiment on investing 100$ and trading it using BitBay Cryptocurrency exchange for a period of one month. For this we implemented the Python script that automatically gathered predictions every 10 minutes, and if it found it profitable to buy, sell or exchange crypto (taking into account BitBay fees) the recommended action was taken. After a month the account balance showed 114.82$, what conforms method is profitable, especially crypto market is on its down currently. Our bot did about 1-3 transaction per day. In contrast we also used well recognized KryptoBot that managed to convert 100$ into 102.45$ within the same period of time.

V. CONCLUSIONS

From the data analysis conducted, we can conclude that cryptocurrency fluctuations depend heavily on social media sentiments and web data bases such as Google Trends. In regard on the future price of the cryptocurrency we can conclude that the Twitter sentiments with respect to crypto price tend to be positive. Many people tweet about crypto even if the price of them goes down giving a positive Twitter sentiment. However, we have identified some problems associated with prediction of crypto. One of the problems is the high level of flexibility of the currency due to volatility nature of cryptocurrency in the current market. We also see that bank regulations, political risk and regulatory agencies caused major fluctuations of the currency during the study of this paper. Our hybrid model, as shown in our results, achieved consistently good results even when shown blind testing data. We found the most powerful predictors to be Google Trends data together with general negative sentiment (including weighted sentiment). Negative news and carries a larger weight, as shown by the correlation values during our data exploration phase. We recommend a hybrid model to help alleviate some of the deficiencies of any one model, and most of the research supports this methodology. Summing up we prove that psychological and behavioural attitude of population has great impact on cryptocurrency prices that are very speculative.

Finally, our solution is shared as Python tool on GitHub repository. The script is capable to customize to any currency type, allows custom windows for grouping tweets and averaging sentiment and Google Trends data, it allows custom number of tweets to extract, connects to Google Trends API to get search trends, connects to CryptoCompare API to get currency prices, connects to Twitter API to get tweets containing specified keywords (currencies), performs sentiment analysis on tweets using all described models, builds models using Google Trends and sentiment analysis, makes predictions for future price, gives recommendations for each model and totals the number of buy / sell / hold recommendations for the group.

REFERENCES

1. P. Dolan, R. Edlin, "Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis?." Journal of Health Economics 21.5 (2002): 827-843.
2. J. Bollen, H. Mao, X. Zeng, "Twitter mood predicts the stock market." Journal of computational science 2.1 (2011): 1-8.
3. J. Prosky, X. Song, A. Tan, M. Zhao, "Sentiment Predictability for Stocks." arXiv preprint arXiv:1712.05785 (2017).
4. A.M. Rather, A. Agarwal, V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns." Expert Systems with Applications 42.6 (2015): 3234-3241.
5. S. Nie, Q. Ji, "Feature Learning Using Bayesian Linear Regression Model." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.
6. V. Karalevicius, N. Degrande, J. De Weerdt, "Using sentiment analysis to predict interday Bitcoin price movements." The Journal of Risk Finance 19.1 (2018): 56-75.
7. D. Garcia, F. Schweitzer, "Social signals and algorithmic trading of Bitcoin." Royal Society open science 2.9 (2015): 150288.
8. L. Kristoufek, "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era." Scientific reports 3 (2013): 3415.
9. A.K. Kuchibhotla, L.D. Brown, A. Buja, E.I. George, L. Zhao, "A Model Free Perspective for Linear Regression: Uniform-in-model Bounds for Post Selection Inference." arXiv preprint arXiv:1802.05801 (2018).
10. D.J.C. MacKay, "Bayesian interpolation." Neural computation 4.3 (1992): 415-447.
11. A. Salinca, "Convolutional Neural Networks for Sentiment Classification on Business Reviews." arXiv preprint arXiv:1710.05978 (2017).
12. Y.B. Kim, J. Lee, N. Park, J. Choo, J.H. Kim, C.H. Kim, "When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation." PloS one 12.5 (2017): e0177630.
13. A.J. Smola, B. Schölkopf, "A tutorial on support vector regression." Statistics and computing 14.3 (2004): 199-222.
14. L. Bottou, "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT'2010. Physica-Verlag HD, 2010. 177-186.
15. J.H. Friedman, "Stochastic gradient boosting." Computational Statistics & Data Analysis 38.4 (2002): 367-378.
16. S. Wold, A. Ruhe, H. Wold, W.J. Dunn, "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses." SIAM Journal on Scientific and Statistical Computing 5.3 (1984): 735-743.
17. M. Collins, R.E. Schapire, Y. Singer, "Logistic regression, AdaBoost and Bregman distances." Machine Learning 48.1-3 (2002): 253-285.
18. A.E. Hoerl, R.W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems." Technometrics 42.1 (2000): 80-86.
19. R. Kohavi, "Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid." KDD. Vol. 96. 1996.
20. H. Zou, T. Hastie, "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67.2 (2005): 301-320.

1. SKLEARN: http://scikit-learn.org/ [↑](#footnote-ref-1)
2. https://www.cryptocompare.com/ [↑](#footnote-ref-2)
3. https://github.com/cjhutto/vaderSentiment [↑](#footnote-ref-3)