Cryptocurrency Price Prediction Using News and Social Media Sentiment

Connor Lamon, Eric Nielsen, Eric Redondo
{conlamon, nielsene, eredondo}@stanford.edu

Background

As the economic and social impacts of cryptocurrencies continue to grow rapidly, so does the prevalence of related news articles and social media posts. While there are many causes of cryptocurrency price fluctuation, as with traditional financial markets, there appears to be a relationship between media sentiment and coin prices.

Our goal was to explore whether sentiment analysis on available online media can inform predictions on whether a coin's price (i.e., perceived value) will go up or down.

Method

Data

General

- Time period: January 1, 2017 to October 31, 2017
- Split: 60% train, 20% dev, 20% test

Media Data *(acquired / cleansed via web scraping)*

- 3200+ news article headlines from cryptocoinsnews.com
- 20 Bitcoin, 20 Ethereum, and 20 Litecoin tweets per day
- Labeled based on actual changes in coin prices

Coin Data *(acquired via Kaggle)*

- Daily market data for Bitcoin, Ethereum, and Litecoin (sample below)

Date	Open	High	Low	Close	Volume
31-Oct-17	6132.02	6470.43	6103.33	6468.4	2,311,380,000

Model

The model is comprised of 2 steps: in the first step, a **classifier** assigns labels to each news headline and tweet; in the second step, a **predictor** uses the labels to generate coin price predictions.

Classifier

- **Features are text words**
  - Implemented via spacy library functions
  - News headline and tweet text is cleansed to remove new line characters and stop characters (e.g., 'a', 'the', 'and')
  - Text is then tokenized by single words
- **Logistic regression is used to learn feature weights**
  - Implemented via scikit-learn library functions
  - Linear support vector classification, multinomial Naive Bayes, and Bernoulli Naive Bayes were also tried, but logistic regression produced the best results
- **Feature weights are used to assign 2 labels to each input**
  - Binary labels describe anticipated Bitcoin, Ethereum, and Litecoin price changes 1 day and 2 days into the future

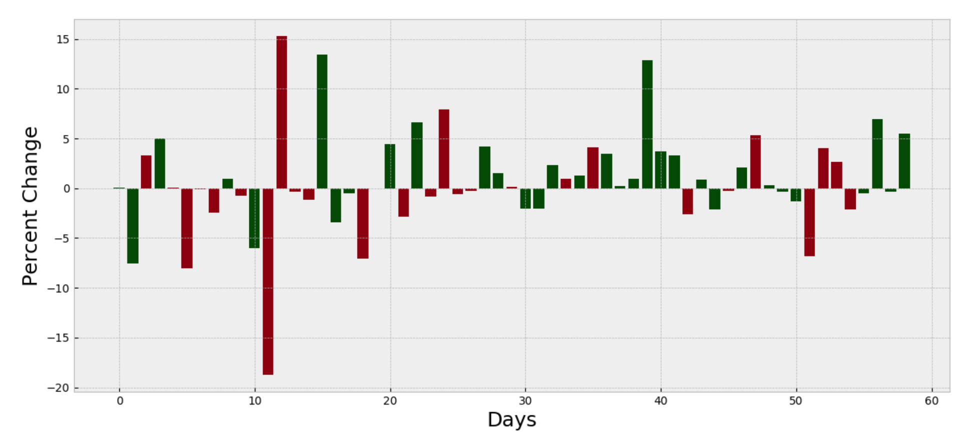
Predictor

- Classifier output is aggregated by day (e.g., all news headlines and tweets from January 1, 2017 are considered together)
- Coin price predictions are generated based on the majority of positive versus negative labels per day

Results

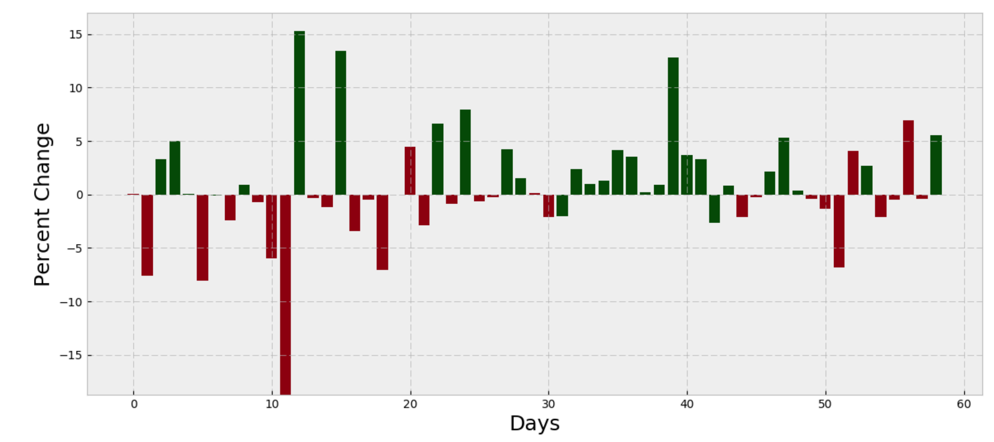
Bitcoin

News Headlines



	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	32 (54%)		27 (46%)	
Price ↑	21 (68%)	+3.8%	10 (32%)	+4.4%
Price ↓	11 (39%)	-2.4%	17 (61%)	-3.2%

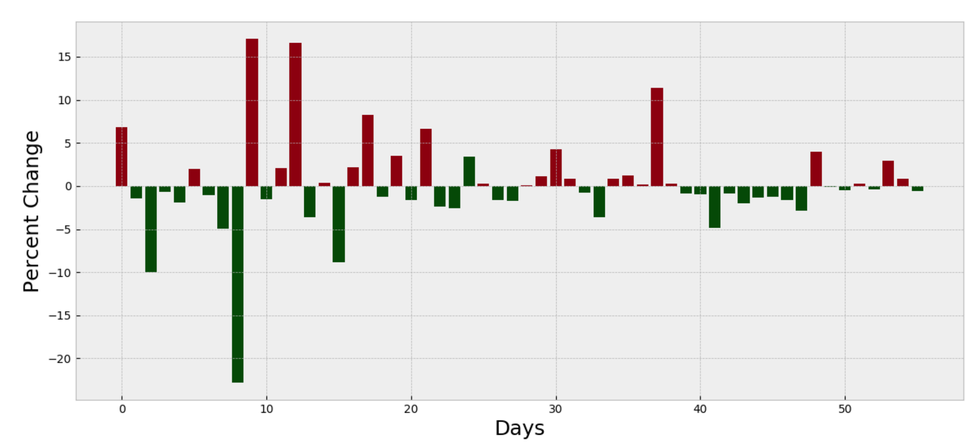
Tweets



	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	30 (51%)		29 (49%)	
Price ↑	26 (84%)	+4.2%	5 (16%)	+3.1%
Price ↓	4 (14%)	-1.2%	24 (86%)	-3.2%

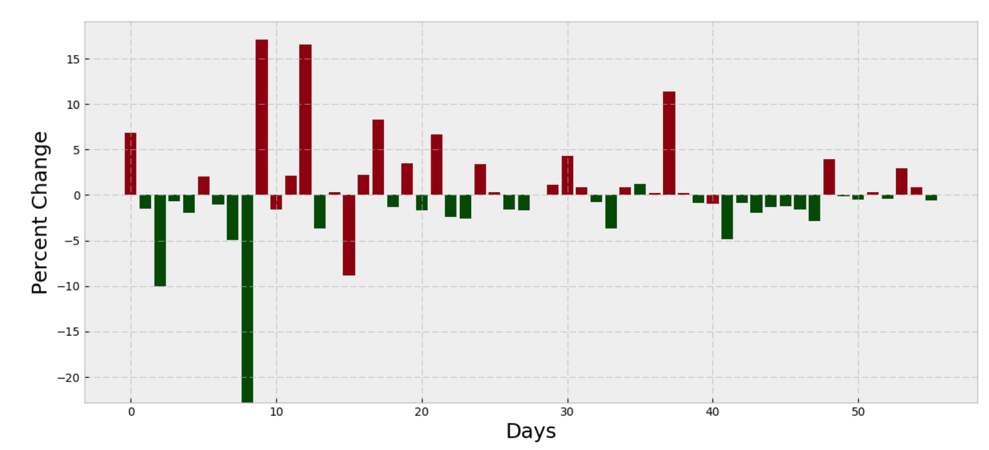
Ethereum

News Headlines



	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	32 (57%)		24 (43%)	
Price ↑	1 (4%)	+3.4%	24 (96%)	+4.0%
Price ↓	31 (100%)	-3.2%	0 (0%)	

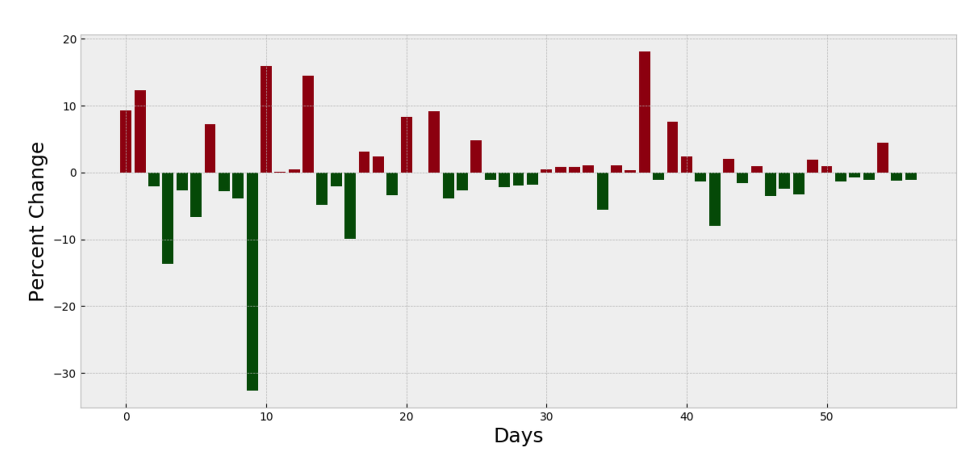
Tweets



	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	29 (52%)		27 (48%)	
Price ↑	1 (4%)	+1.2%	24 (96%)	+4.0%
Price ↓	28 (90%)	-2.8%	3 (10%)	-3.8%

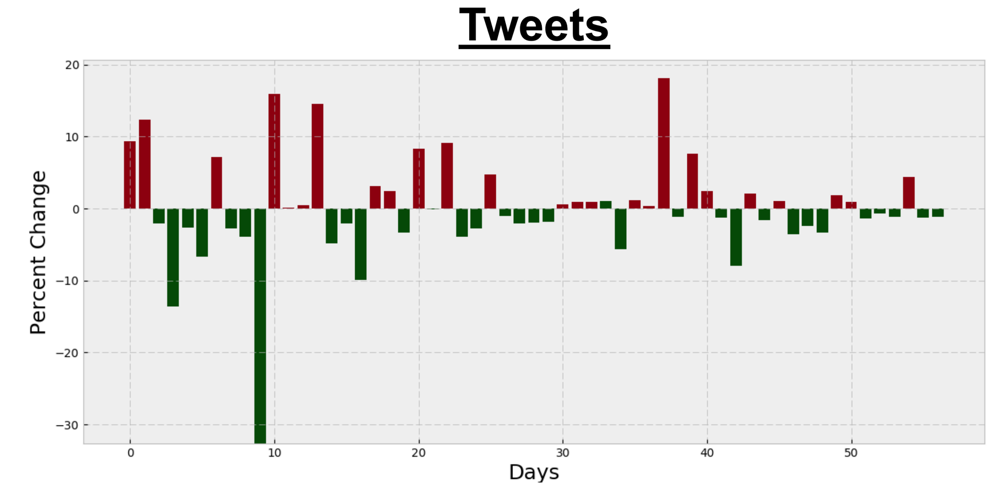
Litecoin

News Headlines



	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	31 (54%)		26 (46%)	
Price ↑	0 (70%)		26 (100%)	+5.0%
Price ↓	31 (100%)	-4.2%	0 (0%)	

Tweets

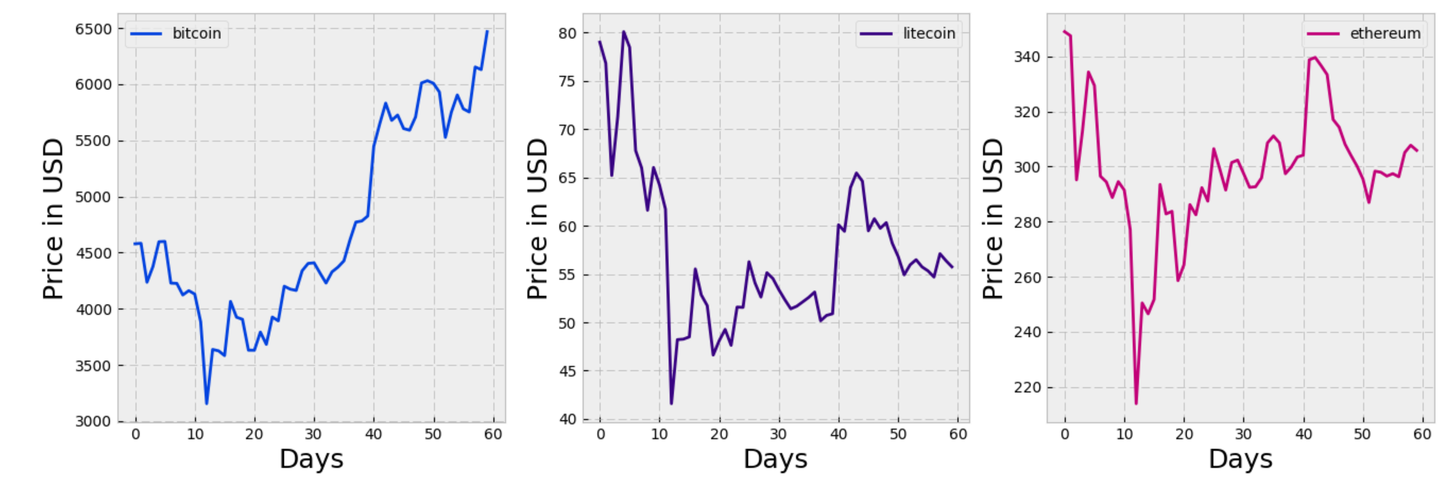


	Correct		Incorrect	
	Count	Avg. price △	Count	Avg. price △
All predictions	32 (56%)		25 (44%)	
Price ↑	1 (4%)	+1.1%	25 (96%)	+5.2%
Price ↓	31 (100%)	-4.2%	0 (0%)	

Note: All shown results for predictions +1 day

Discussion

Coin Prices during Test Set Time Period



Observations

The model appears to identify general trends in coin prices:

- **Bitcoin**: There is a general increase in Bitcoin price during the test set period; the model correctly picks up on this via the text input and most often predicts additional increases
- **Ethereum**: There is a general decrease in Ethereum price during the test set period; the model correctly picks up on this via the text input and most often predicts additional decreases
- **Litecoin**: Similar results to Ethereum

Conclusion

Though the model is still being improved, our two main project objectives were completed.

✓ Develop a model that makes cryptocurrency price predictions using non-technical data
✓ Consistently achieve greater than 50% prediction accuracy

Future Work

We plan to conduct additional experiments and make updates to the model in the future. Current priorities include:

- Training model with a combination of news and twitter data so that classification and prediction will be more robust to different trends in both data sets
- Integrating additional types of media (e.g., from other news sources, Slack channels, subreddits) and larger volumes of input
- Investigating different strategies for labeling training data (e.g., using a NN to label based on text sentiment)
- Using generated price predictions as features in automated cryptocurrency portfolio manager built for CS 221 project

References

[1] S.J. Kumar, "Cryptocurrency Historical Prices" [Data files], 2017. Retrieved from https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory

[2] Twitter API, https://developer.twitter.com/en/docs/tweets/search/overview

[3] Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis; Past 229 project; http://cs229.stanford.edu/proj2015/029_poster.pdf