Data-X @ UC Berkeley

Predicting Influence of News Articles on Crypto Prices using Sentiment Analysis

Sudarshan Gopalakrishnan (EECS), Parth Sanghvi (M.Eng), Johannes Kukula (MSc.),
Kenneth Choong (IEOR), Kshiti Bachlaus (M.Eng), Srujay Korlakunta (EECS)

Background

Our Contributions

Analyzing Sentiment of Articles

Understand the relation of new articles to cryptocurrency prices using Natural Language Processing. Using Recurrent Neural Networks, we will classify articles as having a positive, negative or no effect.

Predict Effect of Articles on Crypto Prices

Using time series analysis and RNN-powered sentiment assignment to predict effect of cryptocurrency prices based on types of articles.

Potential Use Cases

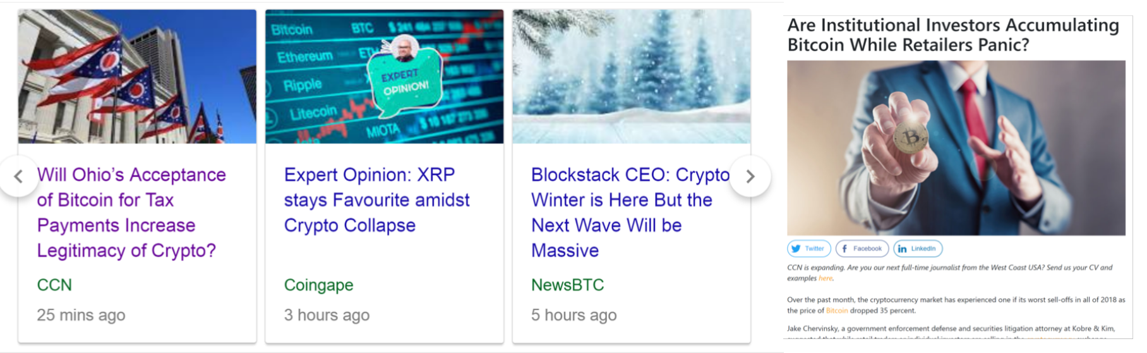
Algorithmic trading

Algorithmic trading (automated trading, black-box trading or simply algo-trading) is the process of using computers programmed to follow a defined algorithm for placing a trade in order to generate profits at a speed and frequency that is impossible for a human trader.

Ranking significance of Articles

Sort and rank articles to allow human readers to sieve through unimportant articles to allow faster information gathering and decision making

Data



Will Ohio's Acceptance of Bitcoin for Tax Payments Increase Legitimacy of Crypto?
CCN
25 mins ago

Expert Opinion: XRP stays Favourite amidst Crypto Collapse
Coingape
3 hours ago

Blockstack CEO: Crypto Winter is Here But the Next Wave Will be Massive
NewsBTC
5 hours ago

Are Institutional Investors Accumulating Bitcoin While Retailers Panic?

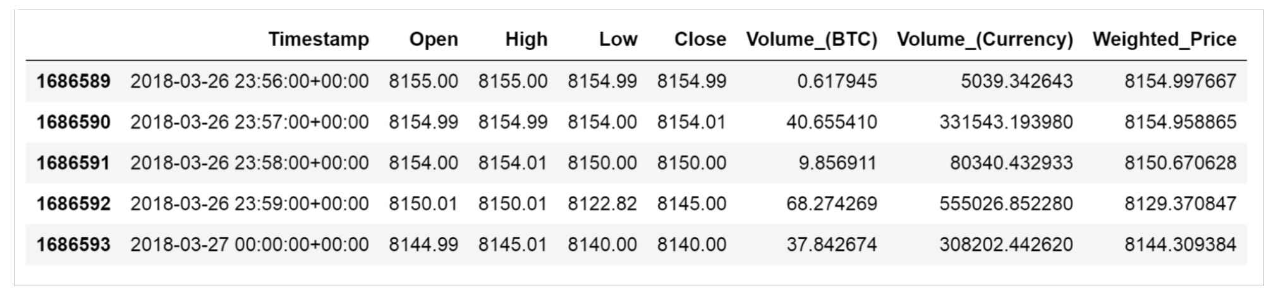
CCN is expanding. Are you our next full-time journalist from the West Coast USA? Send us your CV and examples here.

Over the past month, the cryptocurrency market has experienced one if its worst sell-offs in all of 2018 as the price of Bitcoin dropped 35 percent.

Jake Chervinsky, a government enforcement defense and securities litigation attorney at Kobre & Kim,

Article Data: Full news articles related to bitcoin were scraped from the web

- **Sources**: News APIs (CNBC, Bitcoin Magazine etc.), Kaggle
- **Information**: Author, source, timestamp, title, content
- **Timeframe**: 2018-01-06 to 2018-04-22



	Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price
1686589	2018-03-26 23:56:00+00:00	8155.00	8155.00	8154.99	8154.99	0.617945	5039.342643	8154.997667
1686590	2018-03-26 23:57:00+00:00	8154.99	8154.99	8154.00	8154.01	40.655410	331543.193980	8154.958865
1686591	2018-03-26 23:58:00+00:00	8154.00	8154.01	8150.00	8150.00	9.856911	80340.432933	8150.670628
1686592	2018-03-26 23:59:00+00:00	8150.01	8150.01	8122.82	8145.00	68.274269	555026.852280	8129.370847
1686593	2018-03-27 00:00:00+00:00	8144.99	8145.01	8140.00	8140.00	37.842674	308202.442620	8144.309384

(Snapshot of Bitcoin data from Coinbase)

Bitcoin Price Data: Minute-by-minute Bitcoin (BTS) price data in USD

- **Information**: Timestamp, open, high, low, close, vol., BTC price
- **Timeframe**: 2014-12-01 to 2018-03-27

Acknowledgements

Mentorship of Paradigm, a platform that automates Over-The-Counter (OTC) crypto trading for institutional crypto traders within a native chat application.

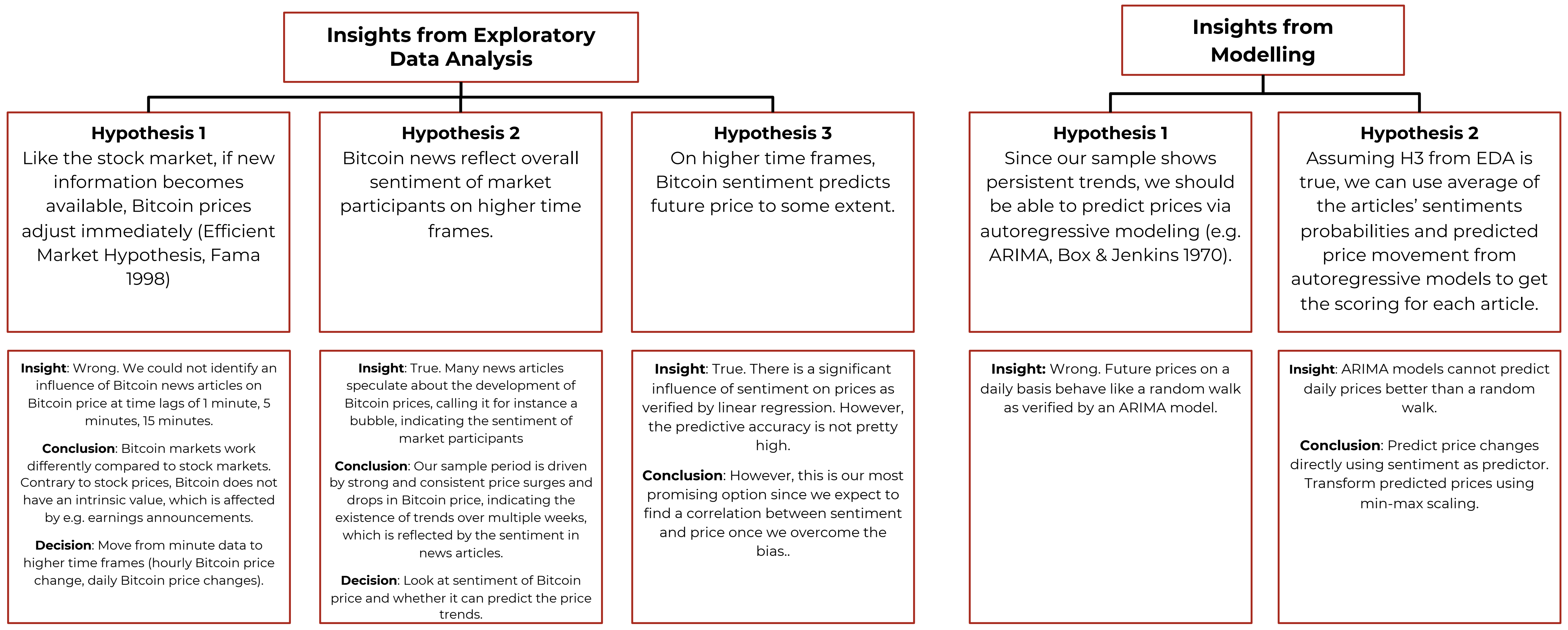
Project built on top of pre-existing work from our mentor, Elias Castro's, Data-X Spring 2018 project.

Box, G. E. P, and GM Jenkins (1970) Time Series Analysis, Forecasting, and Control. *Francisco Holden-Day*.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance1. *Journal of financial economics*, *49*(3), 283-306

Seth, Shobhit. "Basics of Algorithmic Trading: Concepts and Examples." *Investopedia*, Investopedia, 22 Oct. 2018, www.investopedia.com/articles/active-trading/101014/basics-algorithmic-trading-concepts-and-examples.asp.

Hypotheses



Insights from Exploratory Data Analysis

Hypothesis 1
Like the stock market, if new information becomes available, Bitcoin prices adjust immediately (Efficient Market Hypothesis, Fama 1998)

Insight: Wrong. We could not identify an influence of Bitcoin news articles on Bitcoin price at time lags of 1 minute, 5 minutes, 15 minutes.

Conclusion: Bitcoin markets work differently compared to stock markets. Contrary to stock prices, Bitcoin does not have an intrinsic value, which is affected by e.g. earnings announcements.

Decision: Move from minute data to higher time frames (hourly Bitcoin price change, daily Bitcoin price changes).

Hypothesis 2
Bitcoin news reflect overall sentiment of market participants on higher time frames.

Insight: True. Many news articles speculate about the development of Bitcoin prices, calling it for instance a bubble, indicating the sentiment of market participants

Conclusion: Our sample period is driven by strong and consistent price surges and drops in Bitcoin price, indicating the existence of trends over multiple weeks, which is reflected by the sentiment in news articles.

Decision: Look at sentiment of Bitcoin price and whether it can predict the price trends.

Hypothesis 3
On higher time frames, Bitcoin sentiment predicts future price to some extent.

Insight: True. There is a significant influence of sentiment on prices as verified by linear regression. However, the predictive accuracy is not pretty high.

Conclusion: However, this is our most promising option since we expect to find a correlation between sentiment and price once we overcome the bias..

Insights from Modelling

Hypothesis 1
Since our sample shows persistent trends, we should be able to predict prices via autoregressive modeling (e.g. ARIMA, Box & Jenkins 1970).

Insight: Wrong. Future prices on a daily basis behave like a random walk as verified by an ARIMA model.

Hypothesis 2
Assuming H3 from EDA is true, we can use average of the articles' sentiments probabilities and predicted price movement from autoregressive models to get the scoring for each article.

Insight: ARIMA models cannot predict daily prices better than a random walk.

Conclusion: Predict price changes directly using sentiment as predictor. Transform predicted prices using min-max scaling.

Final Conclusion
Move to daily price data. Use sentiment analysis of news articles to identify influence of news article on Bitcoin price.

Methods

Sentiment Analysis Using RNN

Long short-term memory (LSTM) in RNNs provide a more effective sentiment analysis, given that they solve the vanishing gradient problem and hence are able to remember and process long term contexts. Given the subtle tone of financial articles, LSTMs would yield better results on our dataset.

Primary Analysis Method 1 – Training RNN using Human-Assigned Data

A data-set of 300 articles (150 positive and 150 negative articles) will be read and manually (by human) assigned sentiment in order to create a training set for supervised learning. RNN will be trained and the articles in database will be assigned a probability of sentiment between 0 and 1 using softmax regression, which would be used to assign value to assign a qualitative value of assignment.

Primary Analysis Method 2 – Training RNN using Word Dictionaries

The RNN will be trained using word sets for positive and negative sentiments in financial news articles, and the articles dataset would be assigned a probability of sentiment between 0 and 1 using softmax regression, which would be then converted to a qualitative sentiment.

The RNN will be then testing using a data-set of 300 articles (150 positive and 150 negative articles) will be read and manually (by human) assigned sentiment in order to create a training set for supervised learning.

Trends Analysis Methods for Prediction of News Articles' Significance based on Bitcoin Price Trend

Approach 1: Use predicted sentiment per article from the RNN and directly translate it into the significance of influence of a news article's sentiment on price (e.g. significance of influence on price based on predicted sentiment probabilities from the RNN).

Approach 2: Combine predicted sentiment and other features from news articles (e.g. author/source) to predict price trends. Use autoregressive time series modeling to predict price movements. Use price predictions as a score of the significance of article on price (e.g. predicted price change of greater than 5% → very strong positive influence of article on price).

Takeaways in the Current Approach

Neither does a relationship between price and news on short time lags (1, 5 and 15 minutes) exist (cf. 3.1.1) nor can prices be predicted as verified by autoregressive time series modeling.

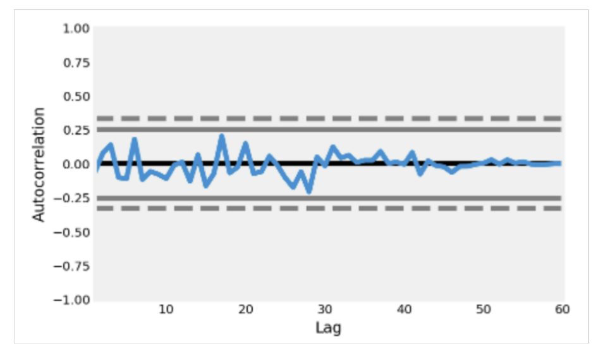
We moved to higher time lags (daily data) to overcome this; however we still couldn't identify an influence of news sentiment on price trends, which is most likely due to biased sentiment labeling of our articles (cf. 4.1.1)

Given that we can identify an influence of sentiment on price, we will go with the primary approach. Otherwise, we will go with the secondary approach, making use of autoregressive time series modeling to come up with our price predictions.

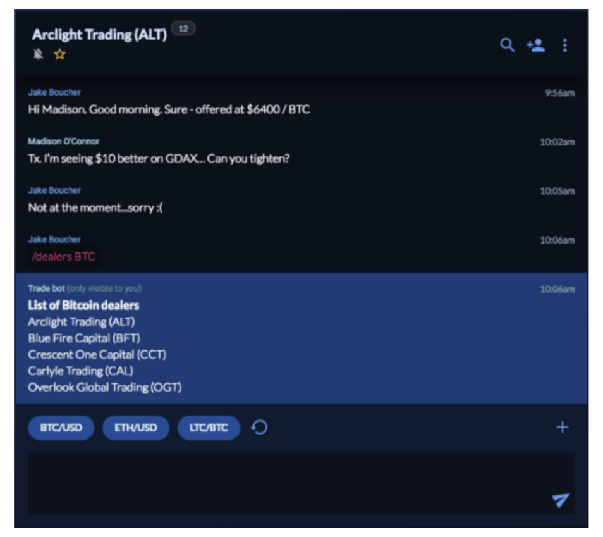
Results

The two results that we measured with our approaches were 1) the accuracy of the RNN model in scoring articles in the test set, and 2) using an ARIMA model to forecast the price of Bitcoin in a given future period. .

For our RNN, we achieved an accuracy of ~71% on the test data, which we deemed accurate enough to proceed with an ARIMA model. We can visualize our ARIMA model through an autocorrelation plot; if it shows a correlation, we can see that the past prices have predictive power. In our plot, because the autocorrelation on the y-axis never passes the horizontal significance boundaries on the x-axis, past prices don't have predictive power.



Weighing the outcomes of the RNN and ARIMA models each at 50%, we give each article a final score, that captures both the sentiment and the accuracy of the article. We then rank the articles by their effectiveness, and display the articles in a chat interface for traders to make better decisions when buying/selling crypto-assets.



Arclight Trading (ALT)

Jake Boucher 9:56am
Hi Madison. Good morning. Sure - offered at $6400 / BTC

Madison O'Connor 10:02am
Tx. I'm seeing $10 better on GDAX... Can you tighten?

Jake Boucher 10:05am
Not at the moment...sorry :(

Jake Boucher 10:06am
/dealers BTC

Trade bot (only visible to you) 10:06am
List of Bitcoin dealers
Arclight Trading (ALT)
Blue Fire Capital (BFT)
Crescent One Capital (CCT)
Carlyle Trading (CAL)
Overlook Global Trading (OGT)

Conclusion

The final model we envisioned gives a rating on an arbitrary scale to each article that is processed through it. The rating given to each article helps analysts and financial traders to sort through the plethora of articles posted every second in order to identify the articles that according to our model could significantly affect the price of cryptocurrency assets.

In addition to the significance of each article, the model also assigns the rating while accounting for the direction of price change. That is, the rating provided by the articles conveys whether the price of the Bitcoin cryptocurrency will change in the positive direction, or the negative.

Limitations

The fact that out RNN was trained using data from a previous team, also means that there are certain biases carried forward in our model. This was necessary because we required labeled datasets in order to have a semi-supervised approach.

If however, we had access to a labelled dataset which was labelled by professional analysts or some sort of a credible entity, there would be a greater reliance achieved by the model.

Next Steps

The final step in this project would be to create a model with a significantly high accuracy and reliance that will enable financial entities to eliminate the need for analysts to intervene and act upon the articles that are rated significant by the model.

A completely reliant model such as this would enable non-professionals to understand the impact of news and sentiment simply based on the user interface of the product.