**Deep Learning to Predict Cryptocurrency Prices**

**IS 698 Capstone Project Spring 2018**

*Edwige Talla Badjio, Shyam Balagurumurthy Viswanathan,*

*Thomas Detzel, Joseph Elikishvili*

**Abstract**

Deep learning algorithms excel at pattern recognition and increasingly drive prediction science in speech and visual recognition, recommender systems, and AI applications such as driverless cars. Finance adopted deep learning to predict stock and other asset prices. This paper examines the use of Long-Short Term Memory recurrent neural networks, a form of deep learning, to predict cryptocurrency prices based on market indices, expert sentiment, and economic uncertainty. Results show that LSTM models can achieve accuracy on par with traditional models and deserve more study to refine methods and broaden data inputs.

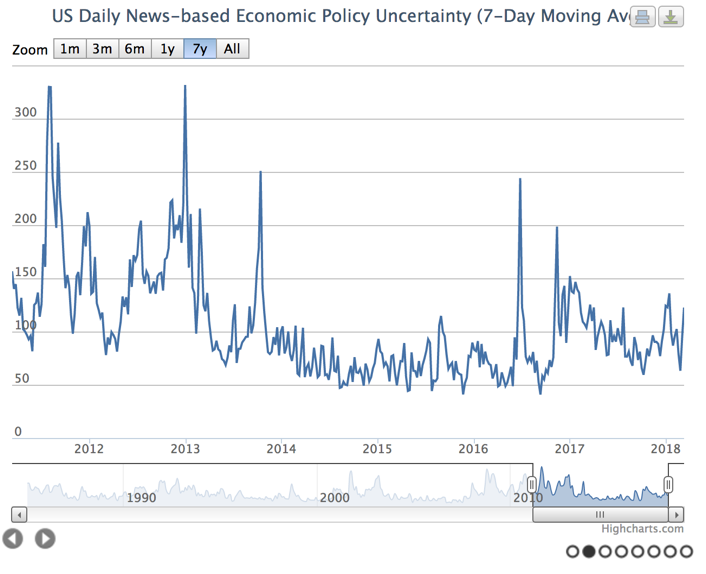
**1. Introduction**

A plethora of research demonstrates that neural networks can be used to significantly bolster predictive accuracy in financial markets, particularly for equity prices, equity indices movements and futures/options. By comparison, relatively little research has investigated the application of neural networks to cryptocurrency prices. Although the predictive challenges appear similar, cryptocurrency markets are less transparent and more volatile than equities. These obstacles hamper the development of reliable algorithmic trading systems for cryptocurrencies. This project explores the use of neural networks combined with semantic analysis and market data to predict cryptocurrency prices. Specifically, we use Long-Short Term Memory (LSTM) recurrent neural networks to predict next-day Bitcoin prices with reasonable accuracy on unseen test data.

**2. Background**

In recent years, deep learning using neural network algorithms for prediction won increasing acceptance in a wide variety of applications that require pattern recognition. Neural networks are used in voice, music and image recognition, to advise consumer choices in marketing, to predict medical outcomes and analyze the human genome, and to power internet search engines and guide driverless cars. More recently, the financial world has adopted deep learning for prediction. Researchers report that neural networks can improve prediction accuracy over traditional methods when targeting stock prices, stock market index movement or the direction of stock options. Neural networks have three key advantages over traditional time series models and other machine learning approaches: 1) They excel at ferreting out hidden patterns; 2) They can provide a precise probability for each prediction in classification problems, and 3) They respond to nonlinear relationships among predictors. These characteristics mean that neural networks can better synthesize a wide variety of predictive features, including sentiment analysis from various sources, a potentially rich vein of information in the cryptocurrency domain.

Machine learning algorithms, including “deep learning” neural networks, already are in use by hedge funds [1] and other traders to predict price change of traditional equity markets. Significant research backs their effectiveness.

At the University of Manchester [2], Nikova M. attained 77 percent accuracy using a neural network and 11 financial indicators to predict long-term stock prices. Deepak, et al. [3], applied neural networks to Bombay Stock Exchange indices and found neural networks to be the most customizable and practical predictive method. Ding, et al. [4], extracted events from news text and two neural networks, one of them convolutional, to model short- and long-term stock prices. The authors achieved a 6 percent increase in predictive accuracy over traditional forecasting methods and used market simulation to show how their methods could power a profitable trading system. Dai and Zhang [5], used machine learning methods to reach 79 percent long-term predictive accuracy. Our project enters an increasingly active research space in which the opportunity for discovery and application of deep learning methods to cryptocurrencies is ripening.

**Figure 1.** One of our predictors is the [Economic Policy Uncertainty Index](https://fred.stlouisfed.org/series/USEPUINDXD/), which broadly captures concern about economic risk based on sentiment from leading newspapers, the dispersion of Federal Reserve forecasts and congressional analysis of expiring U.S. tax code provisions .

**3. Previous studies**

As cryptocurrencies rise in popularity, researchers are quickly following. Our literature review found that multiple machine learning approaches have been studied [12, 13, 14, 15, 16] using a wide array of predictors, from blockchain network data to broad economic indicators and social network metrics. Many of these projects have produced attractive results and, like our study, combined semantic signals with other market indicators and cryptocurrency measures.

McNally [11] used a Bayesian Recurrent Neural Network and LSTM neural net to achieve 52 percent classification accuracy and predict Bitcoin prices at reasonable computation speed. Madan, Saluja, and Zhao [12] used support vector machine (SVM) and Random Forest models to predict the direction of daily Bitcoin price changes with 99 percent accuracy and 10-minute price changes at 55 percent accuracy, suggesting a possible benchmark for neural network methods. Amjad and Shah [13] used neural networks to achieve 70 percent accuracy predicting Bitcoin prices.

Laskowski and Kim [14] demonstrated that the volume of messages on [Twitter](https://twitter.com/) and [IRC chat](https://en.wikipedia.org/wiki/Internet_Relay_Chat) rooms correlates with Bitcoin price movements. Garcia and Schweitzer [15] found that Bitcoin trading volume, price, transaction volume and social signals from Twitter could be used to formulate a profitable algorithmic trading strategy. Their study found that opinion volume and exchange volume both preceded Bitcoin price rises. Kim, et al., [16] used [Google Trends](https://trends.google.com/trends/) data and analyzed Bitcoin-related social media forums to extract keywords and volume data. They combined the social signals with transaction and price fluctuation metrics, then fed a deep learning model to predict price and transaction count. They achieved an ~80 percent accuracy rate for predicting price and volume. Jiang and Liang [6] tested a convolutional neural network with price, volume and asset values as inputs. They developed portfolio management strategies involving the 12 highest-volume cryptocurrencies on the [Poloniex exchange](https://www.poloniex.com/), which hosts 200 currencies. Their back-tested trading model produced the highest returns at the lowest risk. They achieved similar results in a subsequent experiment [7].

**Table 1.** Image of sample tweets from the top 25 cryptocurrency influencers on Twitter after being scored using NLTK’s VADER sentiment lexicon. Scores from five years of tweets were grouped and averaged to create a daily time series of expert sentiment. The sentiment score was included with Bitcoin market and economic metrics in our LSTM models.

Kim, et al., [17] crawled online user communities for the Bitcoin, Ethereum, and Ripple cryptocurrencies and used [VADER sentiment analysis](http://www.nltk.org/_modules/nltk/sentiment/vader.html) to create a five-level, positive-negative score for comments. The authors employed Granger causality tests to show that lagged sentiment predicted price, then simulated an investment strategy based on the model that over 12 weeks earned a 35 percent profit, or triple that of a random trading strategy. Li and Wang [18] reviewed existing studies of Bitcoin exchange rates that evaluated numerous factors. They gathered 13 variables, ranging from Google Trend sentiment to currency prices, Bitcoin technical data such as hash rate and mining difficulty, Bitcoin price and volume, and economic indicators such as [U.S. GDP](https://fred.stlouisfed.org/series/GDP/), inflation, and the [Fed Funds rate](https://apps.newyorkfed.org/markets/autorates/fed%20funds). They differentiated between short- and long-term effects on price. For example, they found that both public opinion and mining difficulty had an impact on the short-term exchange rate. Demir et al., [19] used Ordinary Least Squares (OLS) and the Quantile-on-quantile Regression (QQ) to show that logarithmic Bitcoin returns are negatively associated with the [Economic Policy Uncertainty Index](https://fred.stlouisfed.org/series/USEPUINDXD/), suggesting that Bitcoin could be used as a hedging tool. Shah and Zhang [9] applied Bayesian regression to Bitcoin price prediction and reported high hypothetical profits on back-tests.

**4. Methodology**

*4.1 Long -Short Term Memory Recurrent Networks*

We propose that change in Bitcoin price can be reasonably predicted using a Long-Short Term Memory (LSTM) recurrent neural network trained on publicly available measures of expert sentiment, Bitcoin market data and select economic indicators. Recurrent neural networks have shown special value in modeling time series data. In a vanilla neural network, inputs are fed to one or more hidden layers of neurons. Each hidden layer represents a linear combination of randomly weighted inputs that are transformed by an activation function. Traditional neural networks rely on a process called backpropagation. Data flows forward and then backward through the hidden layers; the information held by neurons and their weights are updated to achieve the lowest possible error. The final output is modeled on a linear combination of the hidden layers. For a regression problem such as ours, backpropagation optimizes on the sum of squared residuals. A standard feed-forward neural network of P predictors and H hidden units estimates a total of H (P + 1) + H + 1 total parameters. Because even simple neural networks typically involve multiple hidden layers, they may require computing hundreds or thousands of parameters as neurons are transformed and reweighted by the backpropagation algorithm.

Two computational problems often encountered in neural networks are vanishing and exploding gradients. Gradients are computed during the optimization process to minimize error, but in some applications, they can become so large or so small that they extinguish the network.

|  |  |
| --- | --- |
| Predictor | Description |
| Volume | *Total prior-day trading volume* |
| Bid-Ask split | *The prior-day mean difference, bid and ask price* |
| Trades per minute | *Prior-day trades-per-minute average* |
| Transactions | *Prior-day transactions using Bitcoin* |
| Lagged price | *Prior-day weighted closing price $USD* |
| Uncertainty index | *Prior-day U.S. Economic Policy Uncertainty index* |
| Expert semantic index | *Consolidated prior-day positive/negative sentiment of top 25 cryptocurrency influencers on Twitter.* |

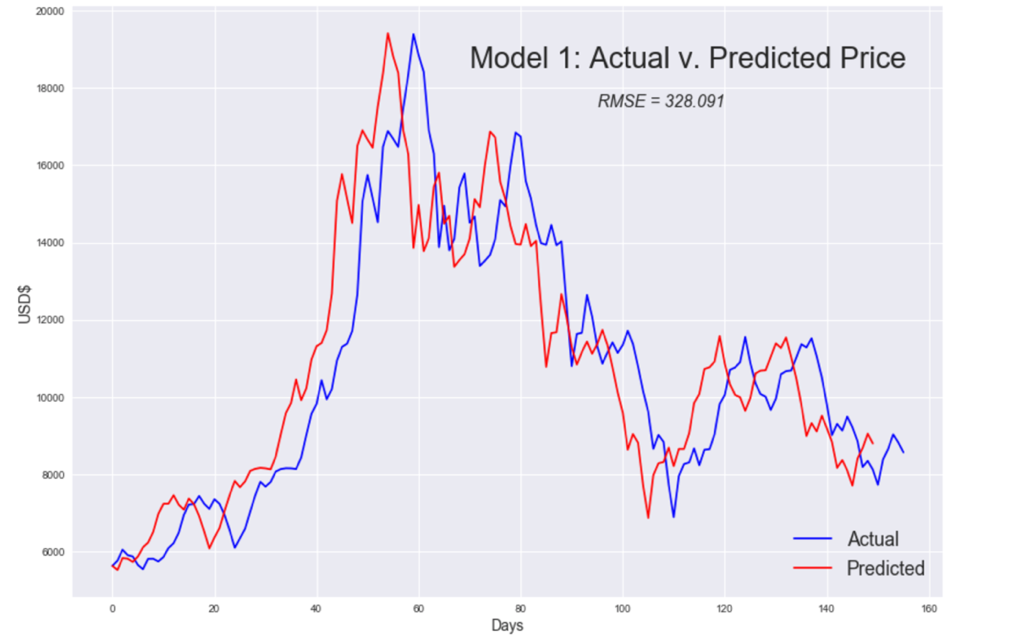
These vanishing and exploding gradients represent a particular obstacle for modeling time series data, which impose a sequencing order. That is, outputs are a function not only of the model predictors but of their time dimensions. LSTM recurrent neural networks are designed to model multiple inputs with awareness and memory of time states. They employ blocks, a special type of neuron that incorporates a memory state governed by a trio of weighted functions called gates. These gates (an input, output and ‘forget’ gate) determine how the block interacts with the larger network. They serve to prevent a block from exploding or vanishing, as with traditional backpropagation algorithms, and can further minimize error by retaining more information. Any neural network is considered to be “deep learning” if it involves more than three hidden layers. Our most successful LSTM models incorporated up to 5 hidden layers, although we also achieved comparably accurate predictions with simpler models.

**Table 2.** Summary table of seven predictors used in our LSTM models. Five are Bitcoin market metrics. The expert semantic index and uncertainty index provide domain-specific and broad market information to the neural network.

*4.2 Sentiment analysis*

We employed natural language programming as implemented in the Scikit-Learn NLTK module for Python 3. From the MarketWatch trade publication, we identified the Top 25 most-influential cryptocurrency experts of 2017 and downloaded all their tweets over a five-year period ending March 24, 2018. We used NLTK’s Valence Aware Dictionary and Sentiment Reasoner (VADER) to compute daily positive or negative sentiment scores from aggregated tweets from each expert. *(See Table 1.)* VADER employs sentiment dictionaries to attach positive or negative values to words. For example, positive words are “rejoiced” (2.0) and “great” (3.1); negative words are “insane” (-1.7) and “tragedy” (-3.4). VADER has the advantage of being able to score many of the slang terms used on Twitter, including emoticons. Scores for each tweet were grouped by day and averaged to compute an expert sentiment time series synchronized with other market inputs.

**Figure 2.** Actual versus predicted prices for Bitcoin during the final 160 days of our study. This LSTM model used five hidden layers of 100, 20, 10, 5 and 1 (output), respectively, over 10 epochs and consistently returned the lowest RMSE.

*4.3 Economic Uncertainty*

Our broad sentiment indicator is the [Economic Policy Uncertainty Index](http://www.policyuncertainty.com/methodology.html) (EPU), which combines Federal Reserve data, newspaper coverage of economic uncertainty and Congressional Budget Office of tax code provisions to estimate levels of concern about the economy. *(See Figure 1.)* The index hit an all-time high of more than 250 during the EU debt crisis of 2011-12; it averaged 109.5 during our study period.

*4.4 Bitcoin technical predictors*

Based on our literature review, we selected a basket of Bitcoin technical and market indicators for inclusion in our predictor set. The data was refreshed on March 24, 2018, from the [Bitcoinity.org](https://data.bitcoinity.org/markets/volume/30d?c=e&t=b), which tracks cryptocurrencies across multiple global exchanges. It should be noted that price, volume and other cryptocurrency attributes may vary depending on the exchange and data collector. Our price data is a weighted average kept by [Bitcoinity.org](https://data.bitcoinity.org/markets/volume/30d?c=e&t=b) across a group of the highest-volume exchanges. Similarly, trade volume, volatility, bid-ask spread, and transactions per minute are computed daily averages.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | RMSE | Layers | | Neurons | Epochs | Compute mins. |
| LSTM 1 | 328.09 | 5 | | 100, 20, 10, 5, 1 | 10 | 2.8 |
| LSTM 2 | 326.01 | 5 | | 50, 10, 5, 3, 1 | 200 | 46.7 |
| LSTM 3 | 362.11 | 3 | | 200, 100, 1 | 100 | 15.8 |
| LSTM 4 | 328.13 | 2 | | 10, 1 | 500 | 116.7 |
| LSTM 5 | 328.26 | 6 | | 100, 50, 30, 20, 10, 5 | 200 | 70.0 |
| Mean | 334.52 |  |  | |  |  |

*4.4 Data exploration and preparation*

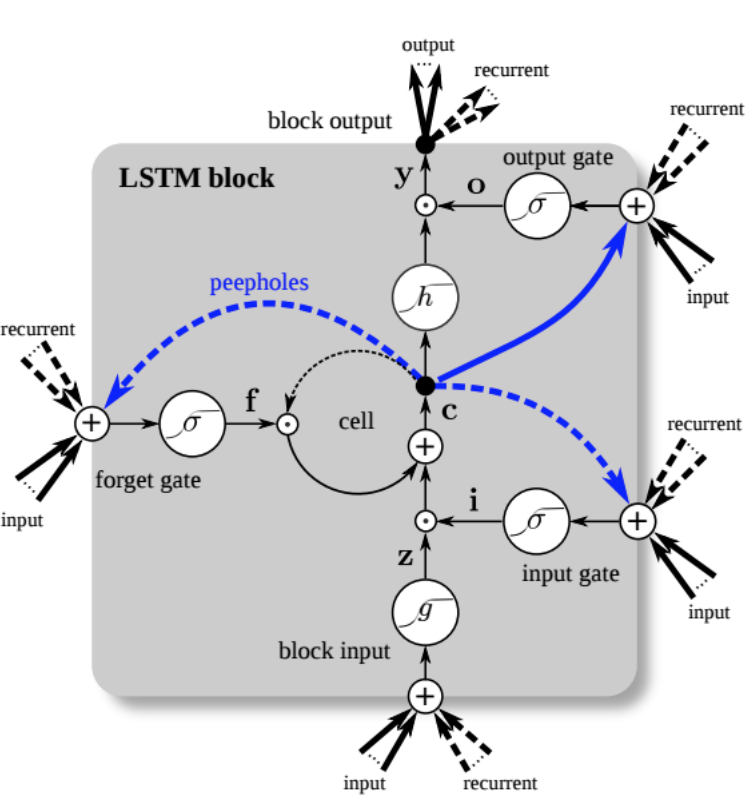
Preparing data for LSTM modeling requires multiple steps. We first examined our predictors for stationarity as recommended in time-series modeling. We plotted each predictor and conducted Augmented Dickey-Fuller tests and differenced predictors that were non-stationary.

**Table 3.** Summary table of accuracy and configuration for five Long-Short Term Memory model used to predict Bitcoin prices. Root mean square error (RMSE) can be read as dollar units. Layers are the number of hidden layers, including the output layer. Neurons is the number of LSTM memory blocks per layer. Each model ran over multiple epochs, or iterations, through the neural network. Approximate compute times are in minutes.

Next, we examined the predictors for collinearity. We eliminated volatility because it was highly correlated (0.70) with volume. We also normalized all our predictors before modeling to accommodate varying scales. Our final models employ seven predictors *(See Table 2.)*: prior day’s price, some transactions, trades per minute, bid-ask spread, volume, economic uncertainty index, and expert semantic score. As a baseline, we also ran an LSTM model predicting Bitcoin price based on prior-day price as a sole predictor. It should be noted that our test data encompassed a period of extraordinary Bitcoin volatility, including a streak of record prices in late 2017 and resulting drops; as such, results may not be reproducible in less volatile periods. Moreover, data compiled from other exchanges using methods that vary from Bitcoinity.org may produce differing outcomes.

**5. Results**

Our array of five LSTM models *(See Appendix A.)* produced an average RMSE on the test data of 334.52, or less than 2 percent of Bitcoin's high price during the study period and less than 4 percent of its final closing price. *(See Table 3 for a summary model results and attributes.)* We found that simpler models run over fewer epochs (each epoch puts all the data through the network) required less compute time and performed better or on par with more complex models. A five-layer LSTM with 100 neurons in the first layer, run over only 10 epochs, repeatedly produced the lowest RMSE. Using fewer layers or adding neurons generally did not have a big impact on RMSE. The range of RMSE for all our models was compact at between 326.01 and 362.11. This suggests that simpler, more computationally efficient LSTMs can produce acceptable results. It also suggests that our LSTM models may simply be seizing on prior day price as the most important predictor and dispensing with complexity.

How good were our predictions? We compared our LSTM results to forecasts made with the popular Facebook Prophet algorithm for analyzing and forecasting business time series. Prophet uses additive regression, and [Facebook says](https://research.fb.com/prophet-forecasting-at-scale/) its default settings produce results as good as those from skilled experts applying refined methods. Using Prophet, we built two models to forecast next-day price based on prior day price and one using a 10-prior-day price window for predictions. RMSE averaged 633.22 for the three models, although the best model reported RMSE of 233.41, or approximately one-third lower than our best LSTM result. That model used logged prior-day price as the predictor *(See Appendix B.)*.

We hypothesized that because LSTM models can retain memory states and excel at pattern recognition, they have distinct advantages when predicting over a large dataset with numerous predictors. This is not clear from our results. Our LSTM models all performed similarly, perhaps because our dataset was relatively small. Prophet, by comparison, performed better when fed less data. Our impression after many model runs and configurations is that there is far more headroom for LSTM models to achieve higher accuracy with more data. On the other hand, LSTM models are computationally demanding. Two of our models took more than 40 minutes to complete on a well-equipped laptop. High-frequency predictions or predictions relying on very large data sets will require scaling to a distributed computing system.

**Figure 3.** Inside an LSTM block. Neurons differ from standard feed-forward networks or recurrent networks. Three gates – input, output and ‘forget’ – use sigmodial or tanh functions and weights to decide whether block state is updated and output to the network, stored or erased. A block’s past state can be fed back to itself or on to the network.[20]

As a further measure, we backtested our best LSTM model by trading a hypothetical $1 million portfolio over the five-year data range using long-only and long-short strategies. *(See Appendix A.)* Both approaches produced gross profits approximately 40 percent greater than a buy-and-hold approach for the period. Again, our study period included an unprecedented run-up in Bitcoin prices in the late 2017 cryptocurrency craze. Though prices fell substantially from the peak, they plateaued at a high level.

**6. Conclusions**

Our small-scale test of deep learning to predict Bitcoin prices shows that LSTM models can perform on a par with other contemporary methods. Further study is needed to determine if LSTM configurations can do so consistently, but we believe that LSTM has the potential to outperform other techniques. Scores of tuning parameters are available to LSTM modelers, including different optimization, activation and loss functions and regularization (dropout, learning rate) switches. Models can be constructed to predict on varying time windows, rather than prior-day information, as in our study. Additional semantic variables could be imagined and collected to enhance accuracy. Our choice to use cryptocurrency expert sentiment, including some naysayers, attempts to cut through advertising and promotional bias in much of the public cryptocurrency commentary. Likewise, our choice to use the Economic Policy Uncertainty Index as a predictor attempts to capture broad sentiment about the direction of the economy.

Our study suffers from some limitations. Bitcoin is the dominant cryptocurrency, and LSTM methods may not generalize to other blockchain coins. Our data is limited to five years of price and trading information. As noted earlier, this period included a fevered period of speculation and historic prices, a large correction and heavy volatility. Our study predicts daily prices at close and may not be applicable or practical for high-speed trading at the minute or millisecond scale. Still, we conclude that LSTMs hold potential for prediction of cryptocurrency prices and deserve more testing with a wider array of predictors and more voluminous data.

*Code and resources on Github at* [*https://github.com/bvshyam/crypto\_trading*](https://github.com/bvshyam/crypto_trading)*.*

**References**

[1] B. Hope, J. Chung, (2017) The Future Is Bumpy: High-Tech Hedge Fund Hits Limits of Robot Stock Picking, The Wall Street Journal, [Online]. Available: https://www.wsj.com/articles/the-future-is-bumpy-high-tech-hedge-fund-hits-limits-of-robot-stock-picking-1513007557

[2] N. Milosevic, “Equity Forecast: Predicting Long-Term Stock Price Movement using Machine Learning”, University of Manchester, 2016. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1603/1603.00751.pdf

[3] R. Deepak et al., “Machine Learning Approach in Stock Market Prediction”, International Journal of Pure and Applied Mathematics, 2017, Vol. 115, No. 8, pp. 71-77. (SNIP: 0.388)

[4] X. Ding et al.,” Deep Learning for Event-Driven Stock Prediction”, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), pp: 2327-2333.

[5] Y. Dai and Y. Zhang, Machine Learning in Stock Price Trend Forecasting, Stanford University, 2013, [Online]. Available:

[6] Z. Jiang and J. Liang, (2016). "[Cryptocurrency Portfolio Management with Deep Reinforcement Learning.](https://arxiv.org/abs/1612.01277)" Xi’an Jiaotong-Liverpool University, Department of Mathematical Sciences. Unpublished student paper.

[7] Z. Jiang et al., (2017). "[A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem.](https://arxiv.org/pdf/1706.10059.pdf)" Department of Mathematical Sciences. Unpublished student paper.

[8] A. Greaves and B. Au, “Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin,”stanford.edu, 08-Dec-2015. [Online]. Available: http://snap.stanford.edu/class/cs224w-2015/projects\_2015/Using\_the\_Bitcoin\_Transaction\_Graph\_to\_Predict\_the\_Price\_of\_Bitcoin.pdf. [Accessed: 03/03/2018].

[9] D. Shah and K. Zhang, “Bayesian regression and bitcoin,” in Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 409–414.

[10] X. Ding et al.,"Deep Learning for Event-Driven Stock Prediction,” Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), pp: 2327-2333.

[11] S. McNally “Predicting the Price of Bitcoin using Machine Learning,” Research Project in Data Analytics, 9th September 2016.

[12] I. Madan et al., “Automated Bitcoin Trading via Machine Learning Algorithms,” 2015. [Online]. Available: https://pdfs.semanticscholar.org/e065/3631b4a476abf5276a264f6bbff40b132061.pdf

[13] M. Amjad and D. Shah. 2017. Trading Bitcoin and Online Time Series Prediction. In NIPS 2016 Time Series Workshop. 1–15.

[14] M. Laskowski and H. Kim (2016) "[Rapid Prototyping of a Text Mining Application for Cryptocurrency Market Intelligence](https://arxiv.org/abs/1611.00315) ." Cornell University, submitted for publication. Available: <http://cs229.stanford.edu/proj2015/029_report.pdf> [Accessed: 03/03/2018].

[15] D. Garcia and F. Schweitzer (2015). "[Social signals and algorithmic trading of Bitcoin."](http://rsos.royalsocietypublishing.org/content/2/9/150288) Royal Society Open Science, 2(9), Royal Society Open Science, 2015, Vol.2(9).

[16] Y. B. Kim et al., "[When Bitcoin encounters information in an online forum: Using text mining to analyze user opinions and predict value fluctuation."](ttp://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177630) PLoS ONE. 5/12/2017, Vol. 12 Issue 5, p1-14.

[17] Y. B. Kim et al., (2016). "[Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies.](http://dx.doi.org.central.ezproxy.cuny.edu/10.1371/journal.pone.0161197)" PLoS ONE, 11(8), E0161197.

[18] X. Li and C.A. Wang, (2017). “[The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin](https://www-sciencedirect-com.central.ezproxy.cuny.edu/science/article/pii/S0167923616302111).” Decision Support Systems, March 2017, Vol.95, pp.49-60. <https://doi.org/10.1016/j.dss.2016.12.001>

[19] E. Demir et al., (2018). "[Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation.](http://eprints.hud.ac.uk/id/eprint/34389/)" Finance Research Letters, 31.01.2018.

[20] Illustration courtesy of Eclipse Deeplearning4j, “A Beginner’s Guide to Recurrent Network and LSTMs.” [Online]. Available: <https://deeplearning4j.org/lstm.html> [Accessed: 3/05/2018].