Big Data Applications Symposium - Fall 2017

Project Name: Bitcoin Price Predictor

Team: Anshu Tomar, Amala Deshpande, Nikita Bhargava

Abstract: *In the recent years, cryptocurrencies especially the Bitcoin have gained a lot of attention from consumers and businesses alike. We aim to provide a prediction model built using the coalesce of tweets centring around bitcoin, "bitcoin" keyword searches over the Google engine and bitcoin pricing in the past year. As per our analysis the trends in Twitter data and Google Search data reflects trends in Bitcoin prices. A regression model is generated using Tweets and Google Searches to predict the future trends in bitcoin pricing. Time lagged data was also used to predict the future prices. The model generated is able to predict the values of bitcoin price with 81 percent accuracy. This prediction model will be beneficial to predict the rise and fall of bitcoin price in the coming days.*

Motivation

Who are the users of this application?
People interested in bitcoin pricing and looking to invest in bitcoins as well as existing bitcoin traders

Who will benefit from this application?
People dealing with buying and selling bitcoins who would like to keep up with bitcoin price trends on a regular basis.

Why is this application important?
Depending on current day's bitcoin related tweets and bitcoin related Google searches, the application will predict whether tomorrow the price of bitcoin will rise or fall and by what value.
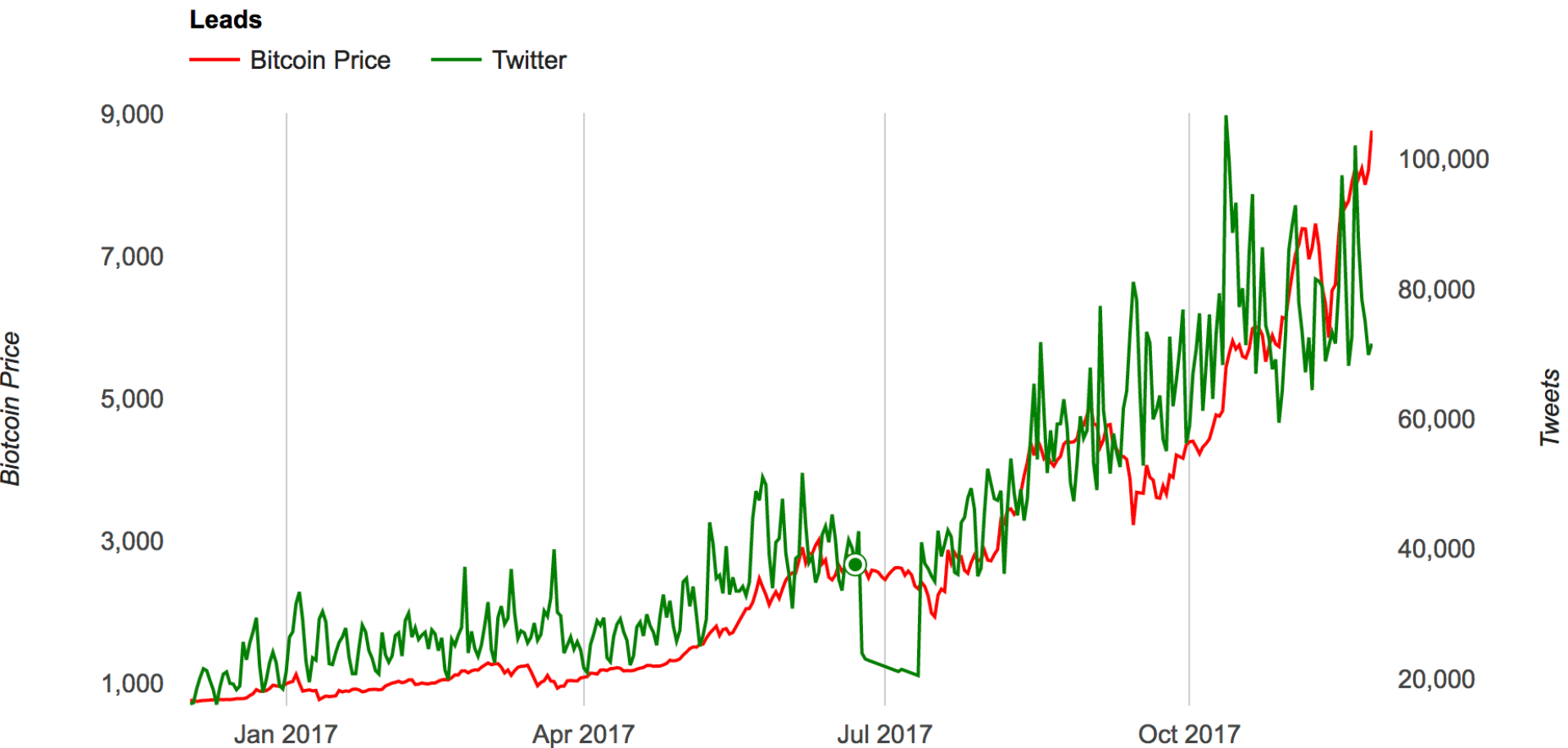
Bitcoin Price Predictor

Remediation

What actuation(s) or remediation actions are performed by the application?

Actionable Insight: Bitcoin price trends mimic the trends in the number of bitcoin related tweets and bitcoin related Google searches per day. Thus, these can be leveraged to predict future bitcoin prices.

Actuation Step: Application will notify the user whether the next day - the bitcoin price is likely to rise or fall and by what amount. This information will be valuable to the user to make bitcoin buying/selling decisions.

Bitcoin Price Predictor
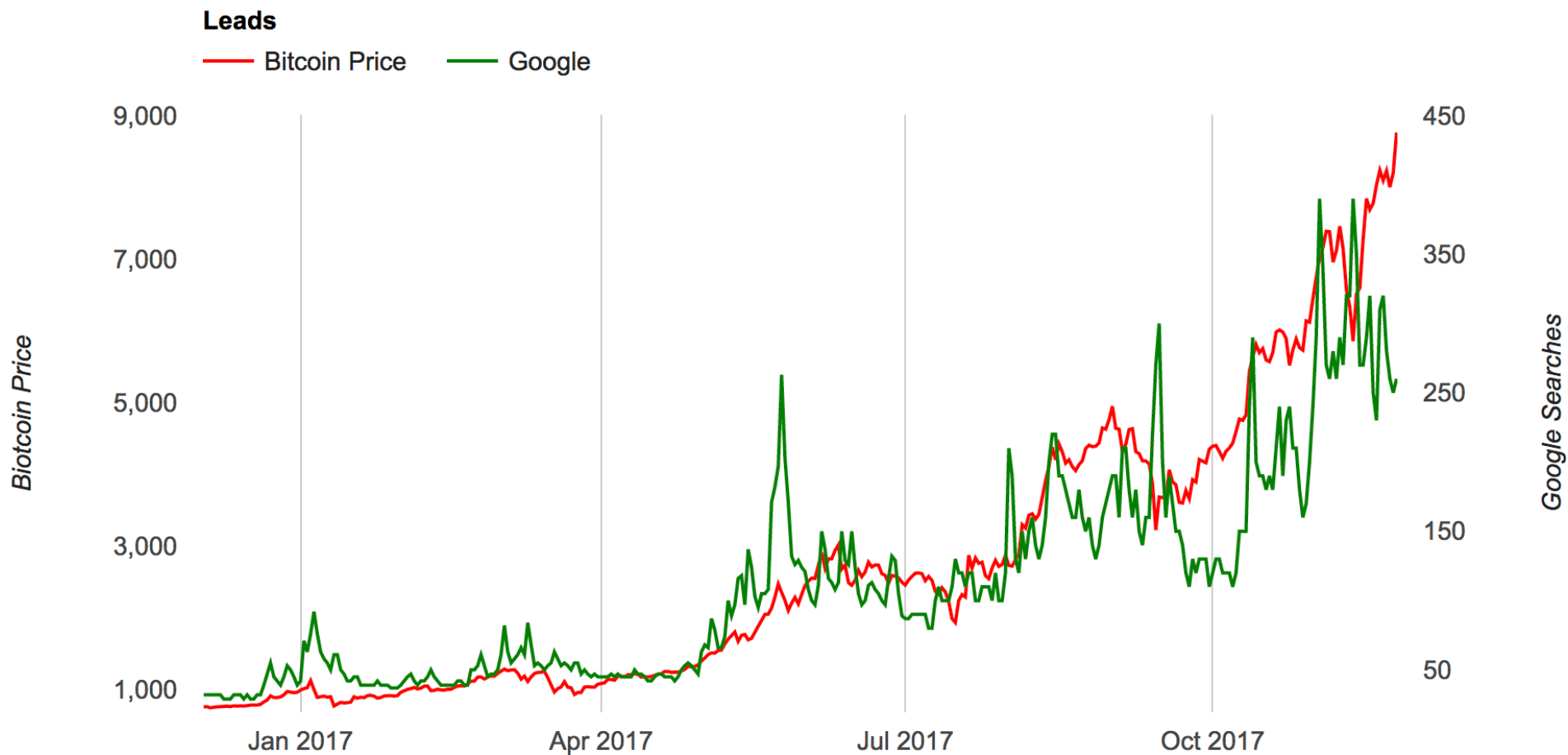
Bitcoin Price Predictor

Data Sources

Name: CoinDesk

Description: This source provides Bitcoin price data for the last 12 months.

Size of data: 365 data points (~ KBs)

Name: Twitter

Description: This source provides number of tweets related to bitcoin for
                    the last 12 months.

Size of data: 365 data points (~ KBs)
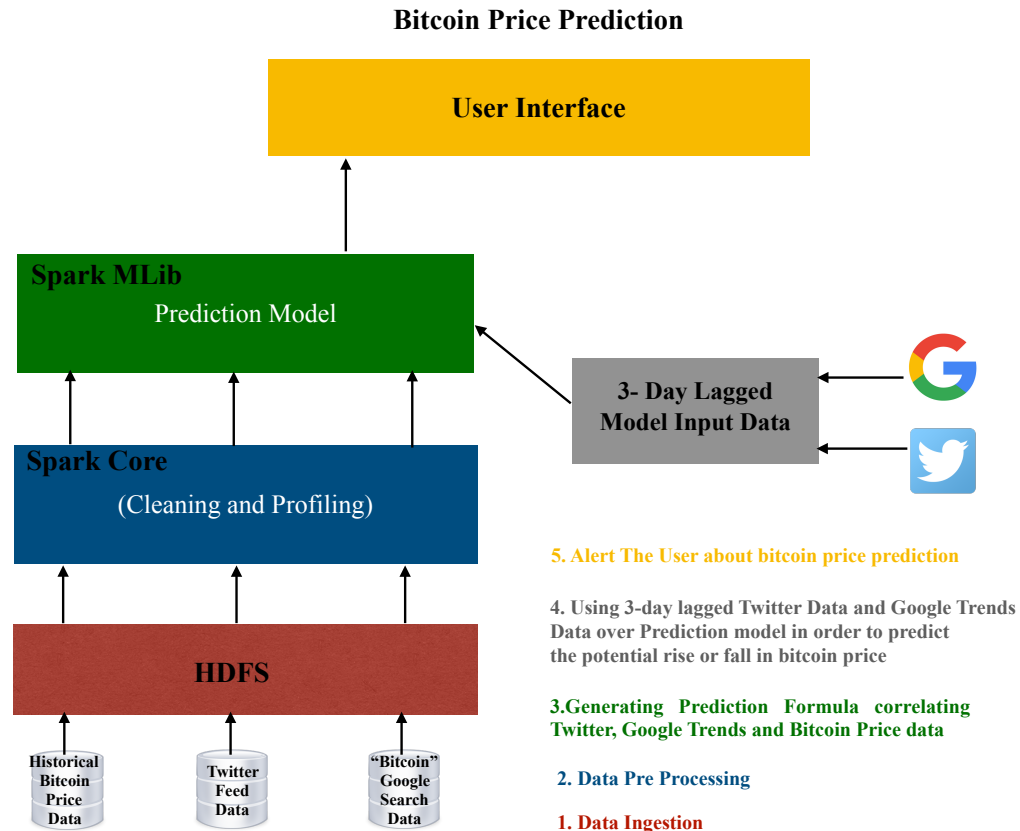
Name: Google Trends

Description: This source provides number of google searches related to
                    bitcoin for the last 12 months.

Size of data: 365 data points (~ KBs)

Bitcoin Price Predictor

Design Diagram

Bitcoin Price Prediction

Platform(s) on which the application runs:

Local Machine (Website + Model), Dumbo (Data Processing)

Bitcoin Price Predictor

Experiments/Results

1. A predictor model was generated using Linear Regression Algorithm by leveraging Apache Spark's MLlib library. For modelling, the Bitcoin Price was taken as dependent variable and Number of Tweets and Number of Google Searches were taken as independent variables.

2. The predicted model can be mathematically represented as follows-

Bitcoin Price = w1 * NumberOfTweets + w2 * NumberOfGoogleSearches + intercept

Where, intercept= 1.076, w1=11.9, w2=20.05

From this formula we can predict the future bitcoin price, given the present Twitter and Google data

3. Bitcoin's actual price was compared with the price predicted by the model. An 81.55% accuracy on an average was obtained.

4. The experimental results showed that variation in number of tweets and google searches per day follow the bitcoin variation trend. Hence, they could be effectively used as features in linear regression modelling.

5. We could also conclude during analysis that Volume of Bitcoin Transactions did not have any direct effect on Bitcoin Price change

Stages for All Jobs
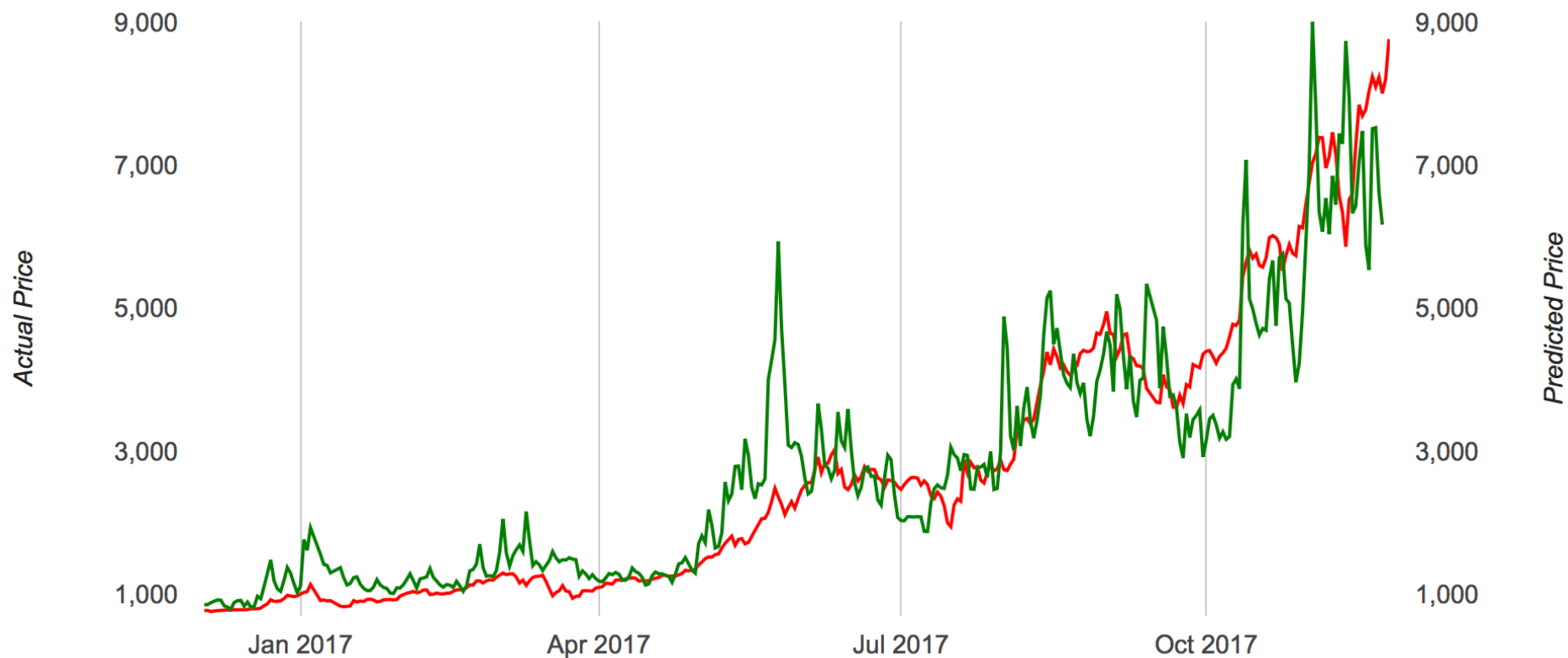
Completed Stages: 107

Completed Stages (107)

Page: 1 2 >

2 Pages. Jump to 1 . Show 100 items in a page. Go

Stage Id ▾	Description		Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
106	saveAsTextFile at GLMRegressionModel.scala:56	+details	2017/12/12 01:29:27	0.1 s	1/1		112.0 B		
105	mean at GenerateModel.scala:99	+details	2017/12/12 01:29:27	31 ms	1/1	6.6 KB			
104	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
103	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	6 ms	1/1	30.9 KB			
102	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	4 ms	1/1	30.9 KB			
101	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
100	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	4 ms	1/1	30.9 KB			
99	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
98	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
97	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	6 ms	1/1	30.9 KB			
96	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
95	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	6 ms	1/1	30.9 KB			
94	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
93	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	4 ms	1/1	30.9 KB			
92	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	4 ms	1/1	30.9 KB			
91	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	3 ms	1/1	30.9 KB			
90	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
89	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
88	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			
87	treeAggregate at GradientDescent.scala:239	+details	2017/12/12 01:29:27	5 ms	1/1	30.9 KB			

Spark 2.2.0

Jobs   Stages   Storage   Environment   Executors   SQL

GenerateModel application UI

Bitcoin Actual Price vs Predicted price from 3 day Lag data

Legend: Bitcoin Price (red), Predicted 3 (green)

10

Obstacles

1. The biggest obstacle was that without paying, we could not get access to the amount of data we wanted. The Twitter API allows access to tweets only for a few days for free.

2. Twitter historical data is not freely available. Initially we tried two methods of fetching historical tweets, but both data sources were found to be inaccurate. We finally settled on a third data source which could fetch the data but only for one year.

3. The google search data obtained from google trends does not give the actual number of searches. It gives a normalized data for a certain time period.

Bitcoin Price Predictor

Summary

In this project, we attempted to predict bitcoin price on a particular day based on the number of tweets regarding bitcoin and number of google searches with keyword "bitcoin" 3 days back. We were able to successfully predict whether there will be a rise or fall in the bitcoin price on a particular day with a 81.55% accuracy rate. The difference in price can be accounted to the fact that we need to segregate the positive and negative tweets and google searches and then use only the positive tweets and/or google searches as features for predicting a more accurate future price trend for bitcoin. We observed that there is a direct correlation between the frequency of tweets and google searched relating to bitcoin and the bitcoin price by running regressions on these aforementioned datasets spanning a year. We exploited this fact to get out result. An important observation was that there is no correlation between bitcoin price and number of bitcoin transactions.

Acknowledgements

Bitcoin Price Predictor

References

- Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, October 2008

- Martina Matta, Ilaria Lunesu, Michele Marchesi, The Predictor Impact of Web Search Media on Bitcoin Trading Volumes, August 2016

- Devavrat Shah, Kang Zhang, Bayesian regression and Bitcoin, October 2014

- Martina Matta, Ilaria Lunesu, Michele Marchesi, Bitcoin Spread Prediction Using Social And Web Search Media , June 2015

- Ahmet Onur Durahim, Mustafa Coşkun, #iamhappybecause: Gross National Happiness through Twitter analysis and big data, October 2015

- Tao-Hung Chang and Davor Svetinovic, Data Analysis of Digital Currency Networks: Namecoin Case Study, 2016

- Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, Learning Spark : Lightening Fast Data Analysis, February 2015

- Tom White, Hadoop: The Definitive Guide, April 2015

Bitcoin Price Predictor

Thank you!