Prediction of Bitcoin Prices

Aashray Yadav
Nicolas Sarquis
Bhavya Vashisht
Sai Kannan Sampath
Mubarak Abdul Kader
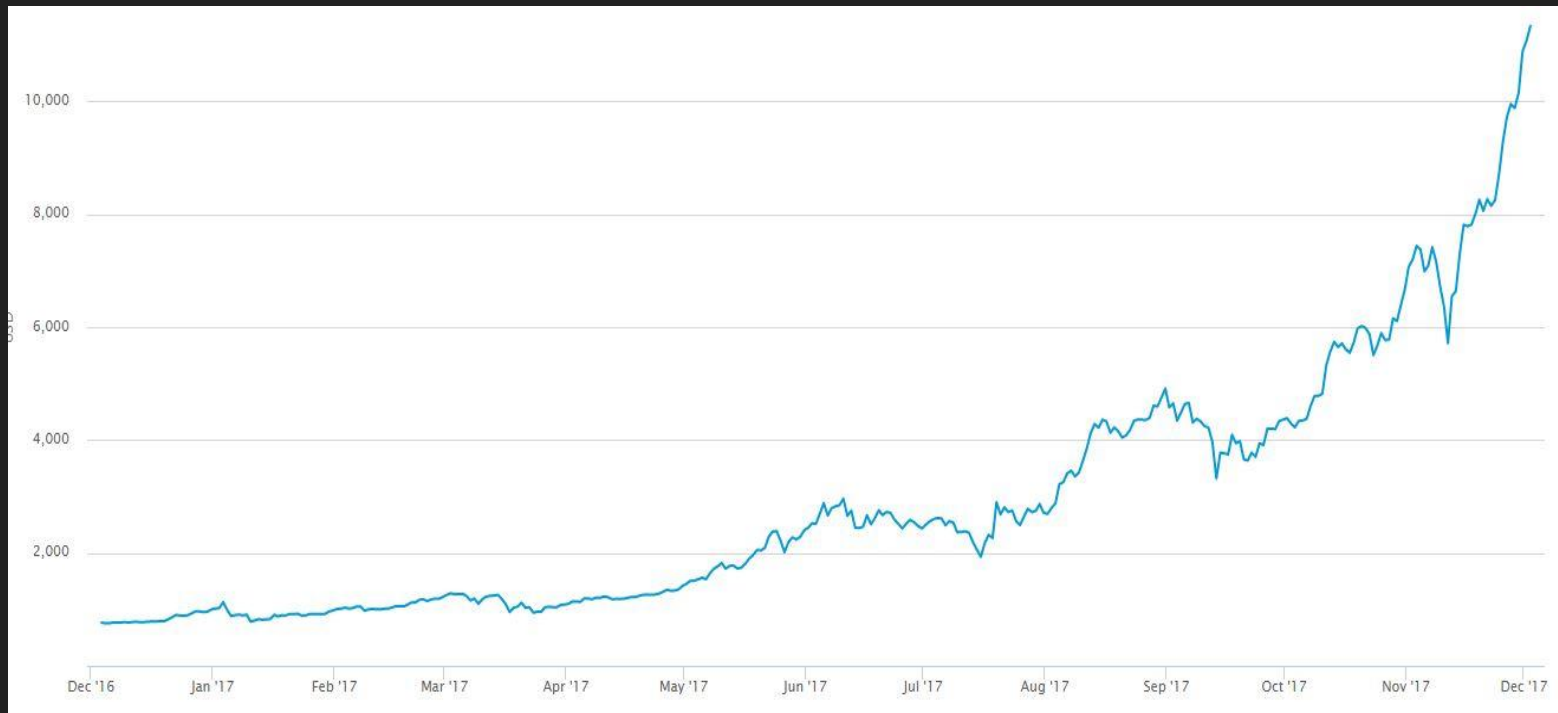
UC Berkeley | Data-X

Bitcoin - The Beast

- ★ Market Cap of $200B
- ★ Currently valued at approx. $11500
- ★ 16.7 M coins in circulation (21 M total)
- ★ Value appreciation of 1000% in 2017 (64% in the last 40 days)
- ★ 400,000 transactions per day
- ★ 500,000 blocks in the bitcoin blockchain
- ★ 8 blocks per hour
- ★ 170GB - Total size of the blockchain
- ★ 300,000 new users on Coinbase over the Thanksgiving Weekend
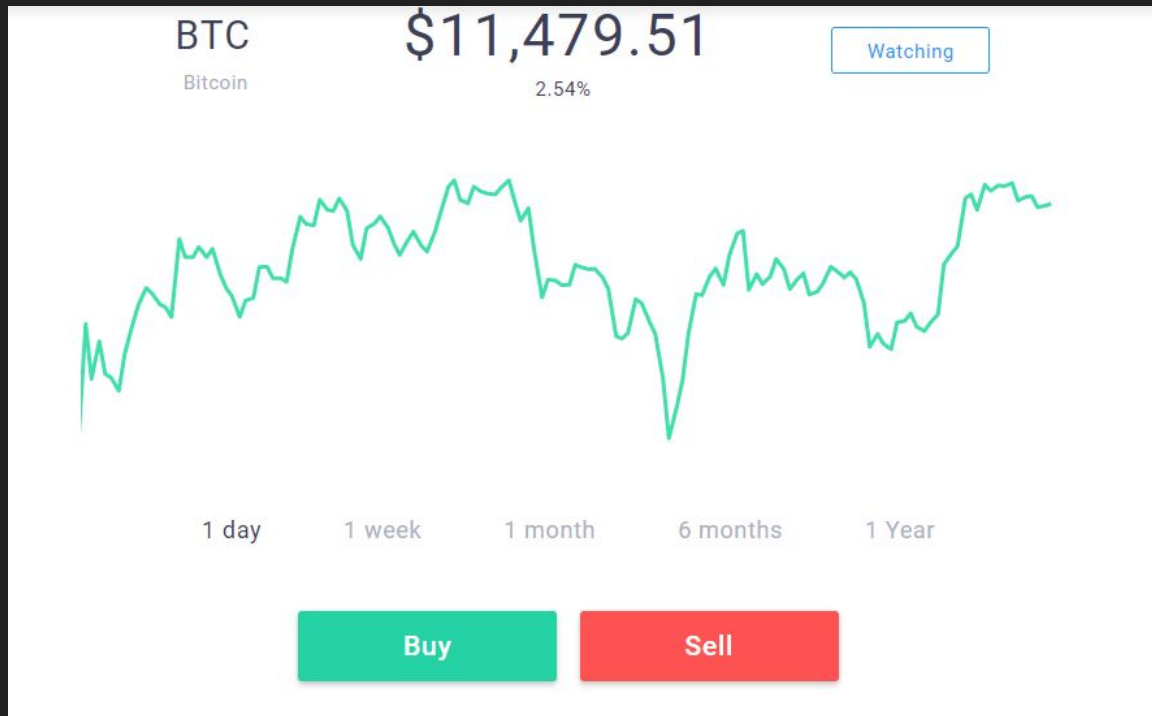
HOW DO WE TAME THIS BEAST?!

UC Berkeley | Data-X

Approach

- ★ Correlation wrt the Stock Market Indices & Other Cryptos
- ★ Basic Linear Prediction Models
- ★ Google Trends data
- ★ Trading Algorithms
- ★ Sentiment Analysis
- ★ Micro-level Analysis - Twitter
- ★ Future steps

Data Analysis

Relationship with Stock Indices

	Bitcoin	Nasdaq	S&P 500	Dow Jones
Bitcoin	1.000000	0.707971	0.688724	0.771410
Nasdaq	0.707971	1.000000	0.989797	0.968166
S&P 500	0.688724	0.989797	1.000000	0.978780
Dow Jones	0.771410	0.968166	0.978780	1.000000

Relationship with Other Cryptos

	bitcoin	ethereum
bitcoin	1.000000	0.946654
ethereum	0.946654	1.000000

	bitcoin	litecoin
bitcoin	1.000000	0.954806
litecoin	0.954806	1.000000

	bitcoin	ripple
bitcoin	1.000000	0.843308
ripple	0.843308	1.000000

	bitcoin	dash
bitcoin	1.00000	0.97258
dash	0.97258	1.00000

Basic Predictive Modelling:
Classifying Movement(Fluctuation)

Challenges Faced:

★ Lack of features
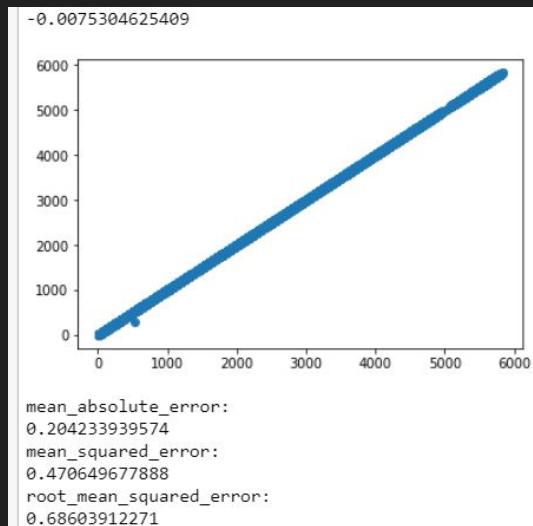★ Sparse dataset

★ Attained 69% accuracy using Logistic Regression

	Date	Price	Sentiment
0	2016-11-30	742.696762	0
1	2016-12-01	752.246125	1
2	2016-12-02	772.437250	1
3	2016-12-03	764.339625	0
4	2016-12-04	764.816250	1

	precision	recall	f1-score	support
0	0.00	0.00	0.00	34
1	0.69	1.00	0.82	76
avg / total	0.48	0.69	0.56	110

0.690909090909

Linear Regression Model

★ Elaborate Data set(~194M)

★ Minimal Error

-0.0075304625409

mean_absolute_error:
0.204233939574
mean_squared_error:
0.470649677888
root_mean_squared_error:
0.68603912271

	Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price
0	1325317920	4.39	4.39	4.39	4.39	0.455581	2.0	4.39
1	1325317980	4.39	4.39	4.39	4.39	0.455581	2.0	4.39
2	1325318040	4.39	4.39	4.39	4.39	0.455581	2.0	4.39
3	1325318100	4.39	4.39	4.39	4.39	0.455581	2.0	4.39
4	1325318160	4.39	4.39	4.39	4.39	0.455581	2.0	4.39

Google Trends(Search Hits)

Challenges Faced:

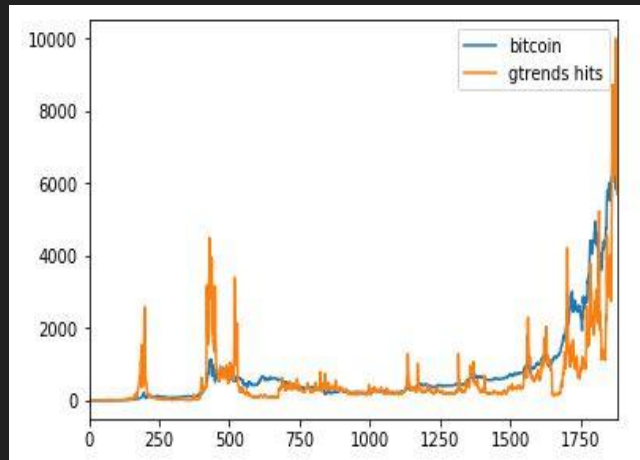Getting acquainted with R and Google Trends package

Learning how to lag data to get desired results

Preprocessing data and getting it in structured format

Results:

Positive Correlation between Google Hits and Bitcoin Price Fluctuation

73% accuracy

UC Berkeley | Data-X

Google Trends(Search Hits) + N Digital Wallets


```
Training Accuracy: 0.913029368048
Accuracy of the model on unseen validation data:  -0.826534132436
```

Sentiment Analysis Using NLTK

Aim

To predict the fluctuation of bitcoin prices using online forum posts

Libraries Used

Beautiful Soup, Scikit-learn

Results

Original Posts | Validation Accuracy - 60.79%
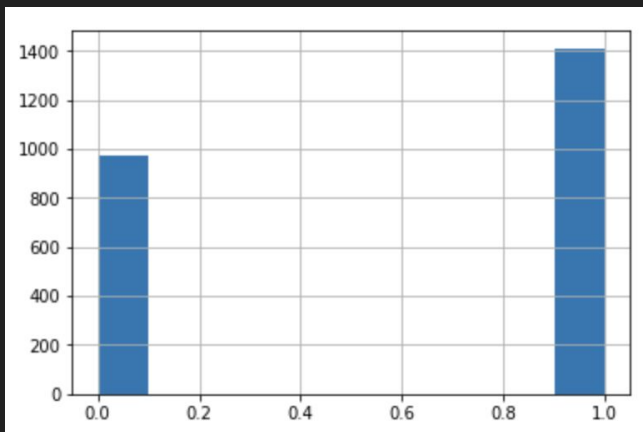Lemmatized Posts | Validation Accuracy - 57.44%

UC Berkeley | Data-X

Sentiment Analysis Using NLTK

Challenges faced

Class Imbalance

- 60:40 Split

Small Dataset

- 2400 posts
- 4 months of data

Sentiment Analysis Using NLTK

Improvements

Larger Dataset

- bitcointalk.org
- 19,000 posts
- 1 year of data

New Results

Original Posts | Validation Accuracy - 73.61%

Lemmatized Posts | Validation Accuracy - 74.23%

Sentential Analysis(Vader)

Computed compound values of Sentences scraped from coindesk to evaluate overall sentiment of a given sentence by making minor additions to Vader's existing Lexicon

```
One of the developers behind the Drivechain project for bringing sidechains to bitcoin is seeking feedback on the project's c
ode. {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Announced on Tuesday, Circle will be launching a digital investment and storage product for a variety of cryptocurrencies in
2018.  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Sorting the signal from the noise may be harder in the cryptocurrency space than almost anywhere else. {'neg': 0.0, 'neu': 1.
0, 'pos': 0.0, 'compound': 0.0}
Researchers are trying to improve prediction markets as potentially valuable tools for measuring community support or opposit
ion to protocol changes. {'neg': 0.0, 'neu': 0.661, 'pos': 0.339, 'compound': 0.8271}
Online banking service Swissquote has launched a bitcoin exchange-traded certificate that it claims will curb the cryptocurre
ncy's volatility. {'neg': 0.0, 'neu': 0.914, 'pos': 0.086, 'compound': 0.128}
Billionaire Mike Novogratz has restated his belief that bitcoin will end the year at $10,000, while ethereum could hit $500.
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Often trumpeted as the future of bitcoin, the success of the Lightning Network may come down to economic forces, say research
ers. {'neg': 0.0, 'neu': 0.844, 'pos': 0.156, 'compound': 0.5719}
The Segwit2x bitcoin fork may have been formally called off, but as many as 150 nodes still running its code have stopped acc
epting transaction blocks {'neg': 0.156, 'neu': 0.731, 'pos': 0.113, 'compound': -0.0772}
Swiss bank Vontobel and Leonteq Securities announced that they will start trading Switzerland's first two mini futures to sho
rt bitcoin on Friday. {'neg': 0.0, 'neu': 0.905, 'pos': 0.095, 'compound': 0.296}
Bitcoin's astonishing price rally this year continues to attract skeptics, including the CIO of Japan Post Bank. {'neg': 0.07
```

Bollinger Bands

Bollinger Bands


```
print(simulation_traiding(nueva,invest=10000))
```

```
('Profit: 3990.678327276828; Number of transactions:8; Returns: 1.3990678327276829',)
```

Tweet Scraping for Sentiment Analysis

- Crowd behaviour and sentiment is the bedrock of Bitcoin price (i.e. there is no inherent "value" to Bitcoin)

- A real-time sentiment analyzer would need the following:
  - Real-time Tweet feed
  - Historical tweet with corresponding Bitcoin prices

- We made a developer account on Twitter

- Tweepy was used to stream real-time Tweets

UC Berkeley | Data-X

Tweet Scraping for Sentiment Analysis

- Historical Tweets are much more difficult to source

- Twitter API does not provide historical tweets beyond the last 7 days

- Twitter website search provides all historic Tweets (all public tweets have been indexed)

- We got all Tweets displayed with hashtag #Bitcoin since 2009

- We attempted scraping Tweets directly from the Twitter search web page

UC Berkeley | Data-X

Tweet Scraping for Sentiment Analysis

- We used Beautiful Soup and urllib to open URL and scrape Tweets

- The AJAX method of Twitter feed supports infinite scrolling - we decided to scroll!

- We used Selenium to simulate page scrolls (up to 2,000 pages) and scraped 3,486 tweets

- The physical memory is limited on local machine - the page loads videos and images too - Datahub doesn't allow Chromedriver in PATH folder

UC Berkeley | Data-X