

NLP Meets Crypto Twitter

Joseph Rodrigues

Goal and Motivation

- Predict what the price of Bitcoin will be in 14 days
- Twitter has a lot of data that might provide an edge on the market
- Across Twitter there are different subcultures that develop
- Crypto Twitter (CT) involves accounts focused primarily on cryptocurrency
- CT has "prominent" accounts
- Use tweets from prominent accounts in addition to historical bitcoin data to try and improve on solely using bitcoin data

-10% or more	-5% to -10%	-5% to +5%	-1% to +1%	+5% to +10%	+10% or more
--------------	-------------	------------	------------	-------------	--------------

Challenges

- Manual data collection
- Rate limits
- 6 labels makes being accurate more difficult
- Possibly not enough data (tweets and price history)

Data

- 2021 Bitcoin data gathered from <https://coincodex.com/crypto/bitcoin/historical-data/>
- ~64k tweets gathered using custom api that interfaces with Twitter api <https://developer.twitter.com/en/docs/twitter-api> (academic access needed)

Date	Open	High	Low	Close	Volume	Market Cap
Feb 08, 2022	\$ 43,838	\$ 45,405	\$ 42,712	\$ 44,082	\$ 73.99B	\$ 830.88B
Feb 09, 2022	\$ 44,017	\$ 44,749	\$ 43,160	\$ 44,284	\$ 68.03B	\$ 832.28B
Feb 10, 2022	\$ 44,376	\$ 45,649	\$ 43,255	\$ 43,524	\$ 65.55B	\$ 839.93B
Feb 11, 2022	\$ 43,511	\$ 43,896	\$ 42,067	\$ 42,378	\$ 86.77B	\$ 818.18B
Feb 12, 2022	\$ 42,375	\$ 42,963	\$ 41,792	\$ 42,126	\$ 65.89B	\$ 802.25B
Feb 13, 2022	\$ 42,182	\$ 42,714	\$ 41,892	\$ 42,236	\$ 39.87B	\$ 802.41B
Feb 14, 2022	\$ 42,046	\$ 42,770	\$ 41,661	\$ 42,595	\$ 47.23B	\$ 800.47B
Feb 15, 2022	\$ 42,550	\$ 44,701	\$ 42,480	\$ 44,603	\$ 58.82B	\$ 831.73B
Feb 16, 2022	\$ 44,523	\$ 44,523	\$ 43,368	\$ 43,998	\$ 51.24B	\$ 833.90B

```
{4}
  text : My tweets marking bottoms as uze?
  created_at : 2021-02-16T04:41:26.000Z
  id : 1361536137769021440
  author_id : 509690453

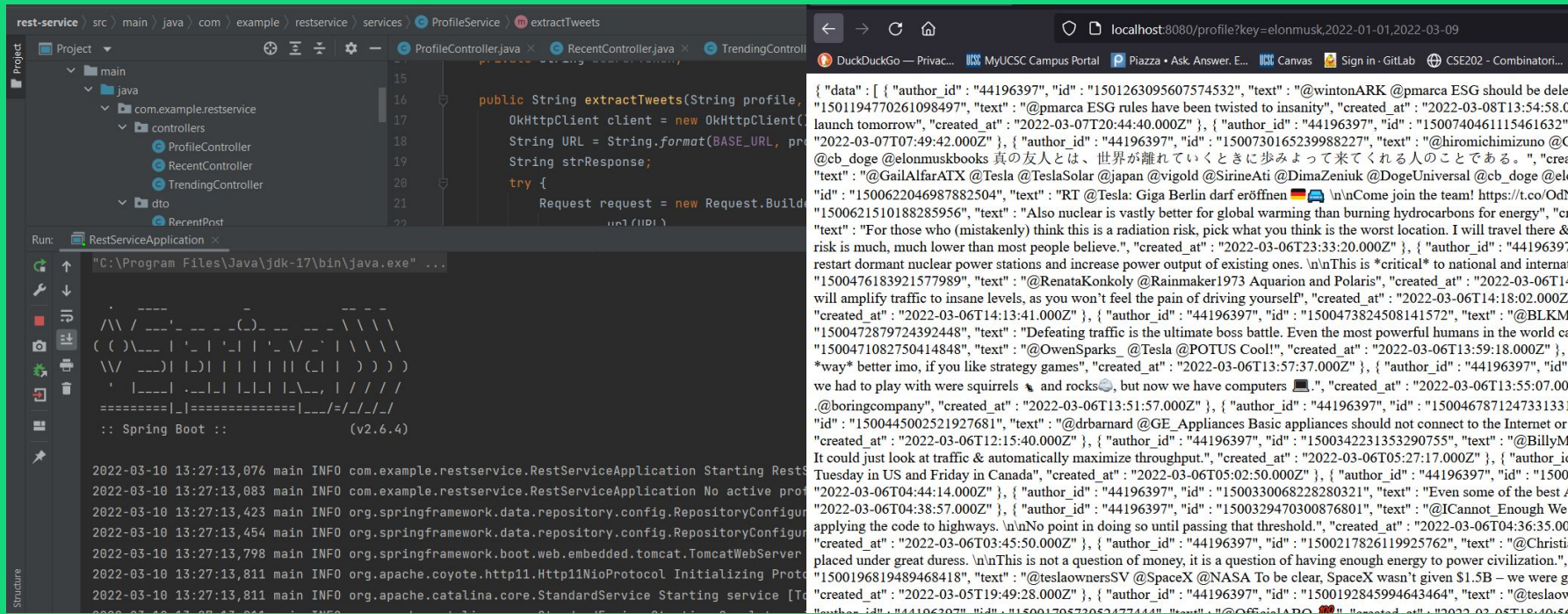
{4}
  text : Zero is a destination we're all headed for....just
        faster for people long BTC https://t.co/xXCZJOHks3
  created_at : 2021-02-16T00:12:39.000Z
  id : 1361468498971553795
  author_id : 509690453

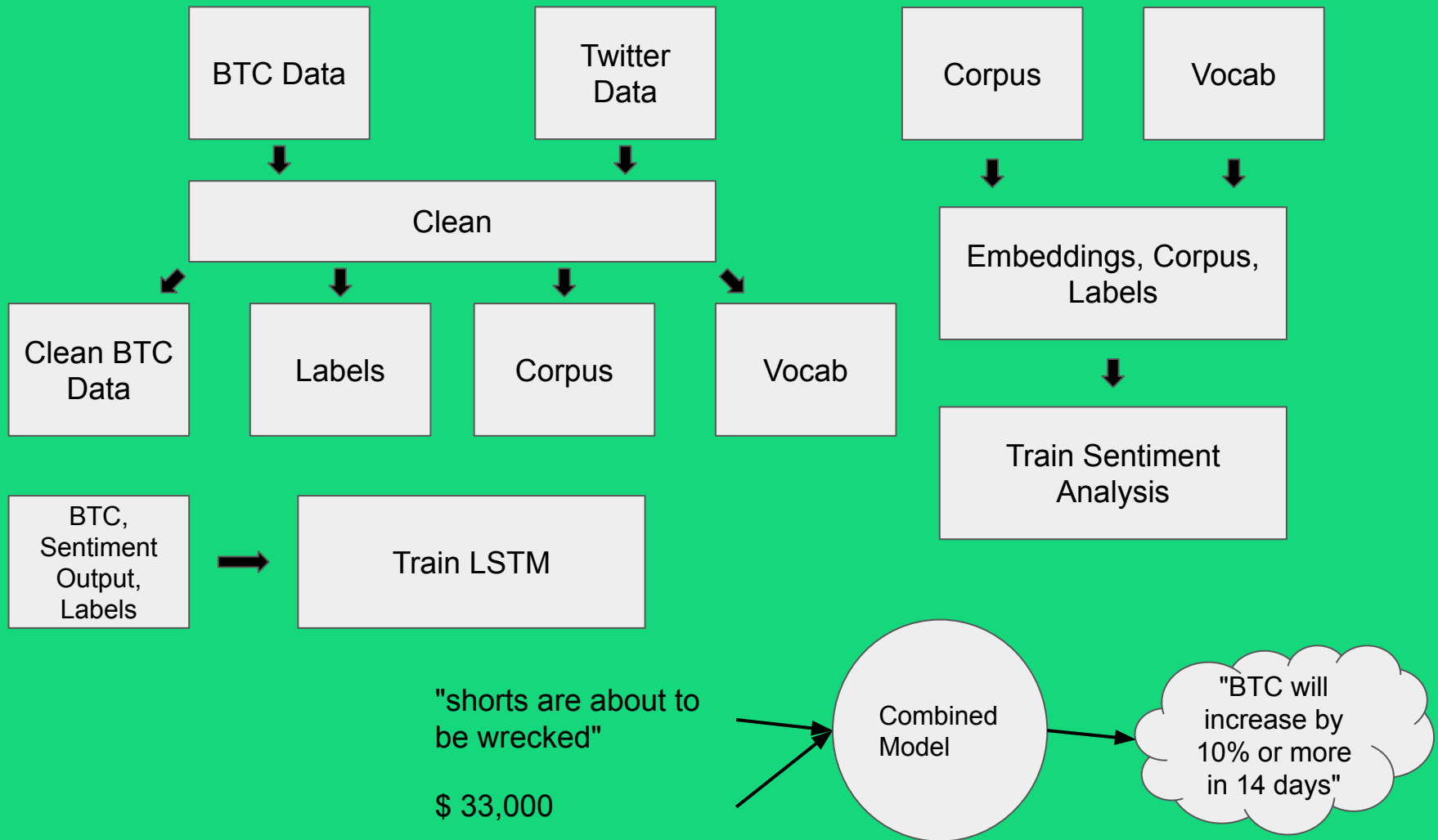
{4}
  text : @RNR_0 Is leverage trading Super Saiyan 3?
  created_at : 2021-02-15T14:42:01.000Z
  id : 1361324895401955334
  author_id : 509690453
```

Twitter Accounts Used

- @depression2019
 - ~ 35K followers
- @ThinkingBitmex
 - ~ 60K followers
- @profplum99
 - ~ 75K followers
- @lightcrypto
 - ~ 140K followers
- @ThinkingUSD
 - ~ 200K followers
- @CryptoKaleo
 - ~ 470K followers
- @zhusu
 - ~ 500K followers
- @SBF_FTX
 - ~ 600K followers
- @cobie
 - ~ 630K followers
- @elonmusk
 - ~ 73M followers

Data Collection in Action





Sentiment Model Results

- Word embeddings created using Word2Vec from Genism
- Both SVMs and CNNs were tried for sentiment analysis
- SVM performed much better than CNN
- Tried embeddings of size: 50, 100, 150, 250, 300, 350, 400, 450, and 500
- Embeddings of size 50 and 350 performed best for the 6 labels
- Tried reducing to binary classification to improve accuracy
- For binary classification SVM performed best when trained with embeddings of size 100, 250, and 400

Embedding Size	50	350
SVM_6	14.9%	14.9%

Embedding Size	100	250	450
SVM_2	38.3%	38.3%	40.4%

Bitcoin Only Model Results

- Trained both SVMs and LSTMs only with bitcoin data
- Bitcoin only SVM still performed poorly on 6 labels although was a bit more accurate (21.3% compared to 14.9%)
- Bitcoin only SVM for 2 labels was surprisingly worse than Twitter only SVM for 2 labels (14.9% compared to 40.4%)
- Bitcoin only LSTM for 6 labels wasn't accurate, but improved on Twitter only SVM for 6 labels (23.6% compared to 14.9%)
- Bitcoin only LSTM for 2 labels was a major improvement attaining an accuracy of 70.6% in the months of November and December of 2021

Results of Training Models with Individual Twitter Accounts

Twitter Account	2 labels	6 labels
elonmusk	25.5%	42.6%
cobie	25.5%	12.8%
CryptoKaleo	34.0%	17.0%
depression2019	34.0%	10.6%
lightcrypto	19.1%	2.1%
profplum99	30.0%	12.8%
SBF_FTX	48.9%	21.3%
ThinkingBitmex	36.2%	4.3%
ThinkingUSD	14.5%	0.0%
zhusu	19.1%	6.4%

Full Model Results

- All of the below models were trained with an LSTM

	Training Accuracy	Testing Accuracy
2 labels 100 embeddings	69.79%	70.6%
2 labels 250 embeddings	72.79%	70.6%
2 labels 450 embeddings	60.1%	70.6%
6 labels 50 embeddings	49% 46.4%	53% 58.8%
6 labels 350 embeddings	42.3%	29.4%