# 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



## 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)

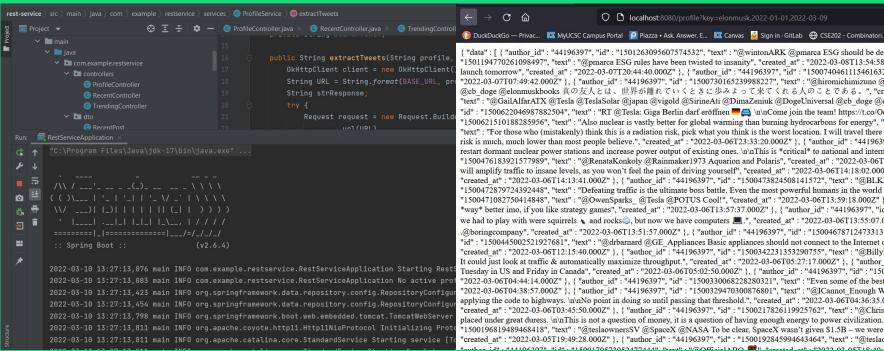
Date	Open	High	Low	Close	Volume	Market Cap
Feb 08, 2022	\$ 43,838	\$ 45,405	\$ 42,712	\$ 44,082	\$73.99B	\$830.88B
Feb 09, 2022	\$ 44,017	\$ 44,749	\$ 43,160	\$ 44,284	\$68.03B	\$832.28B
Feb 10, 2022	\$ 44,376	\$ 45,649	\$ 43,255	\$ 43,524	\$ 65.55B	\$839.93B
Feb 11, 2022	\$ 43,511	\$ 43,896	\$ 42,067	\$ 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**

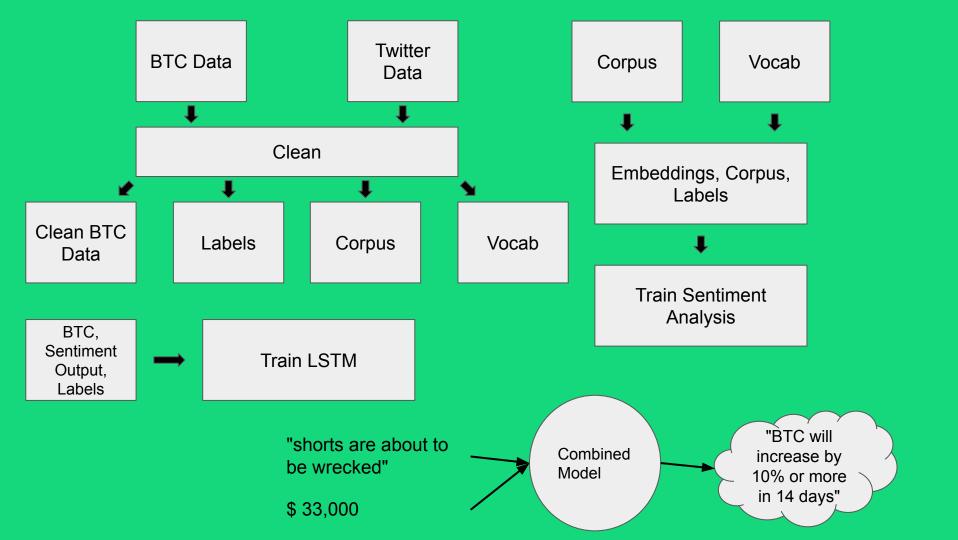
- @depression2019
  - ~ 35K followers
- @ThinkingBitmex
  - ~ 60K followers
- @profplum99
  - ~ 75K followers
- @lightcrypto
  - ~ 140K followers
- @ThinkingUSD
  - ~ 200K followers
- @CryptoKaleo
  - ~ 470K followers
- @zhusu
  - ~ 500K followers
- @SBF\_FTX
  - ~ 600K followers
- @cobie
  - ~ 630K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

#### Data Collection in Action



{ "data" : [ { "author id" : "44196397", "id" : "1501263095607574532", "text" : "@wintonARK @pmarca ESG should be dele "1501194770261098497", "text": "@pmarca ESG rules have been twisted to insanity", "created at": "2022-03-08T13:54:58.0 launch tomorrow", "created at": "2022-03-07T20:44:40.000Z" }, { "author id": "44196397", "id": "1500740461115461632" "2022-03-07T07:49:42.000Z" }, { "author id": "44196397", "id": "1500730165239988227", "text": "@hiromichimizuno @C @cb doge @elonmuskbooks 真の友人とは、世界が離れていくときに歩みよって来てくれる人のことである。", "crea "text" : "@GailAlfarATX @Tesla @TeslaSolar @japan @vigold @SirineAti @DimaZeniuk @DogeUniversal @cb doge @elo "id" : "1500622046987882504", "text" : "RT @Tesla: Giga Berlin darf eröffnen 💳 🔼 \n\nCome join the team! https://t.co/OdN "1500621510188285956", "text": "Also nuclear is vastly better for global warming than burning hydrocarbons for energy", "cr "text": "For those who (mistakenly) think this is a radiation risk, pick what you think is the worst location. I will travel there & risk is much, much lower than most people believe.", "created at": "2022-03-06T23:33:20.000Z" }, { "author id": "44196397 restart dormant nuclear power stations and increase power output of existing ones. \n\nThis is \*critical\* to national and internal "1500476183921577989", "text": "@RenataKonkoly @Rainmaker1973 Aquarion and Polaris", "created at": "2022-03-06T1 will amplify traffic to insane levels, as you won't feel the pain of driving yourself", "created at": "2022-03-06T14:18:02.000Z "created at": "2022-03-06T14:13:41.000Z" }, { "author id": "44196397", "id": "1500473824508141572", "text": "@BLKM "1500472879724392448", "text": "Defeating traffic is the ultimate boss battle. Even the most powerful humans in the world ca "1500471082750414848", "text": "@OwenSparks @Tesla @POTUS Cool!", "created at": "2022-03-06T13:59:18.000Z" }, \*way\* better imo, if you like strategy games", "created at": "2022-03-06T13:57:37.000Z" }, { "author id": "44196397", "id" we had to play with were squirrels & and rocks, but now we have computers E.", "created at": "2022-03-06T13:55:07.00 @boringcompany", "created at": "2022-03-06T13:51:57.000Z" }, { "author id": "44196397", "id": "150046787124733133. "id": "1500445002521927681", "text": "@drbarnard @GE Appliances Basic appliances should not connect to the Internet or "created at": "2022-03-06T12:15:40.000Z" }, { "author id": "44196397", "id": "1500342231353290755", "text": "@BillyM It could just look at traffic & automatically maximize throughput.", "created at": "2022-03-06T05:27:17.000Z" }, { "author ic Tuesday in US and Friday in Canada", "created at": "2022-03-06T05:02:50.000Z" }, { "author id": "44196397", "id": "1500 "2022-03-06T04:44:14.000Z" }, { "author id": "44196397", "id": "1500330068228280321", "text": "Even some of the best A "2022-03-06T04:38:57.000Z" }, { "author\_id": "44196397", "id": "1500329470300876801", "text": "@ICannot\_Enough We applying the code to highways. \n\nNo point in doing so until passing that threshold.", "created at": "2022-03-06T04:36:35.00 "created at": "2022-03-06T03:45:50.000Z" }, { "author id": "44196397", "id": "1500217826119925762", "text": "@Christia placed under great duress. \n\nThis is not a question of money, it is a question of having enough energy to power civilization.". "1500196819489468418", "text" : "@teslaownersSV @SpaceX @NASA To be clear, SpaceX wasn't given \$1.5B – we were g "created at": "2022-03-05T19:49:28.000Z" }, { "author id": "44196397", "id": "1500192845994643464", "text": "@teslaow then id! . !!44106207!! !!d! . !!1500170572052477444!! !!tort!! . !!@Official A DO 🐠!! !!arcated at!! . !!2022 02 05T18.40.5

O | localhost:8080/profile?key=elonmusk.2022-01-01.2022-03-09



#### 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%