Bitcoin sentiment analysis and its impact on price change

BY FILIP SZAFRANSKI

What is sentiment analysis?

Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in the text source.
It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as, products, individuals, topics ,organizations, and services

Twitter sentiment analysis and its importance for business analysis

Twitter Sentiment Analysis Use Cases
Twitter sentiment analysis provides many exciting opportunities for businesses. Being able to analyze tweets, and determine the sentiment of each message, adds a new dimension to social media monitoring.
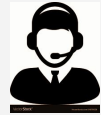
Social Media Monitoring
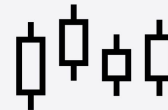
Political Campaigns

Brand Monitoring

Customer Service

Market Research

Trading Algorithms

Sentiment analysis as an additional component for trading algorithms

Trading bots mostly depend on the technical analysis and typical trading indicators and patterns. While this can be good for the stock market, it might fail with the cryptocurrencies due to the high **volatility**.

"With this study, we understood that Bots only depend upon indicators, whereas the volatility induced by Elon Musk tweets or the sentiment due to Chinese government ban on crypto was not embedded. And it requires a human to intervene now and then."

"How much can you make with a trading bot? My experience trading for a month." study by Yaswanthkumar GOTHIREDDY
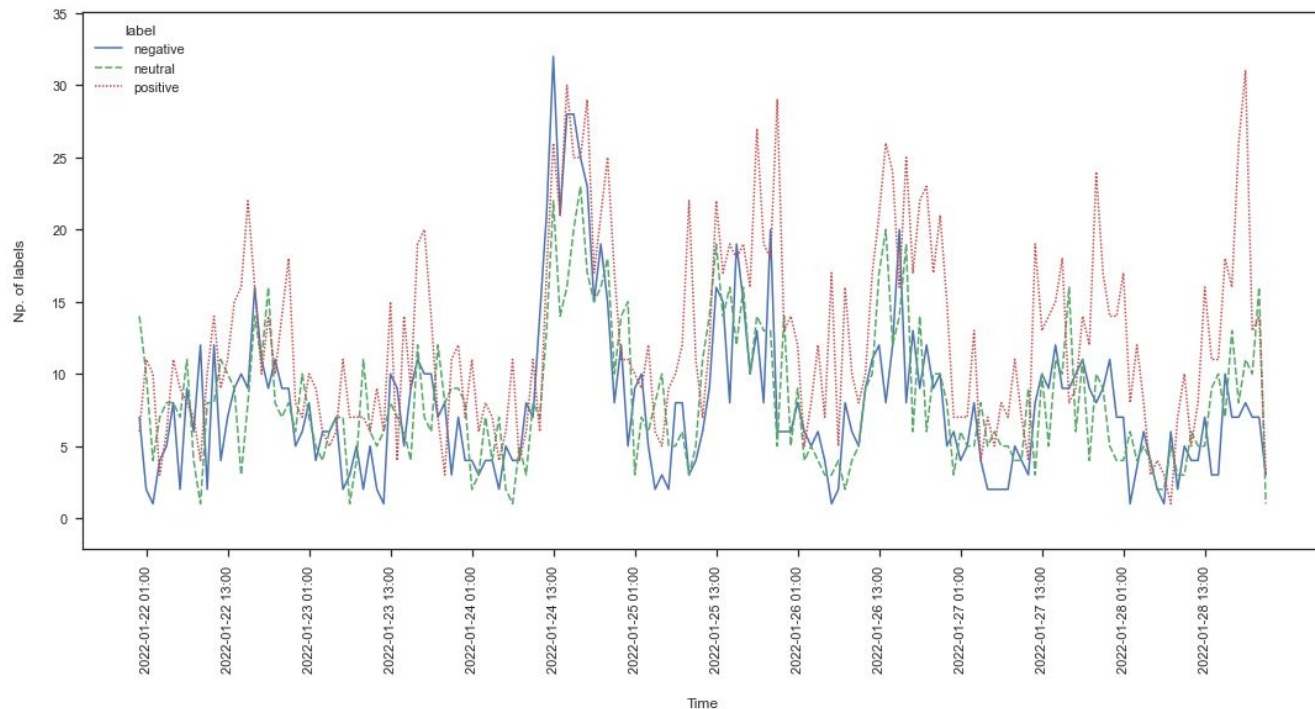
Twitter Data

Data overview: The dataset which is used to train the model contains almost 7,000 tweets about Bitcoin.

|	tweet_id	author_id	created_at	source	retweet_count	reply_count	like_count	text
0	1487185499915706371	613649581	2022-01-28 22:07:08+00:00	Twitter Web App	0	2	2	@FamiLee_Farm @0xBingBong I first dissented in...
1	1487185295757905920	19721574	2022-01-28 22:06:19+00:00	Twitter for iPhone	22	7	105	A disaster.\n\nThe sooner aid agencies worldwi...
2	1487184488719269890	21230289	2022-01-28 22:03:07+00:00	Twitter Web App	17	7	40	With the very real threat of the government tr...
3	1487184473670180870	970207298	2022-01-28 22:03:03+00:00	Twitter Web App	192	233	886	Bitcoin mining's energy use has more than trip...
4	1487184321022767115	3367334171	2022-01-28 22:02:27+00:00	Zapier.com	21	11	49	Tori Zero NFT Project Launched a Joint Coopera...

Twitter Sentiment - Bitcoin
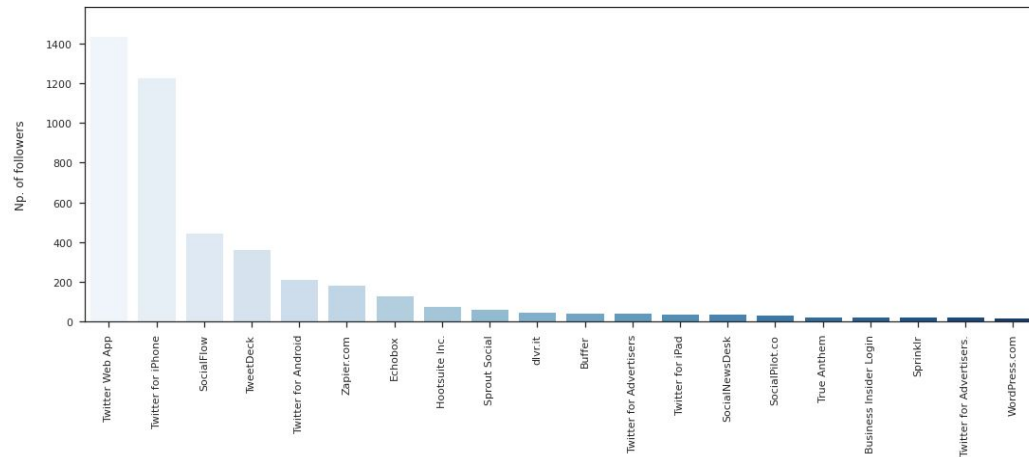
Twitter Sentiment by hour

Twitter Data

Data overview: The dataset which is used to train the model contains almost 7,000 tweets about Bitcoin.

20 Most Popular Authors
Most popular authors based on number of followers
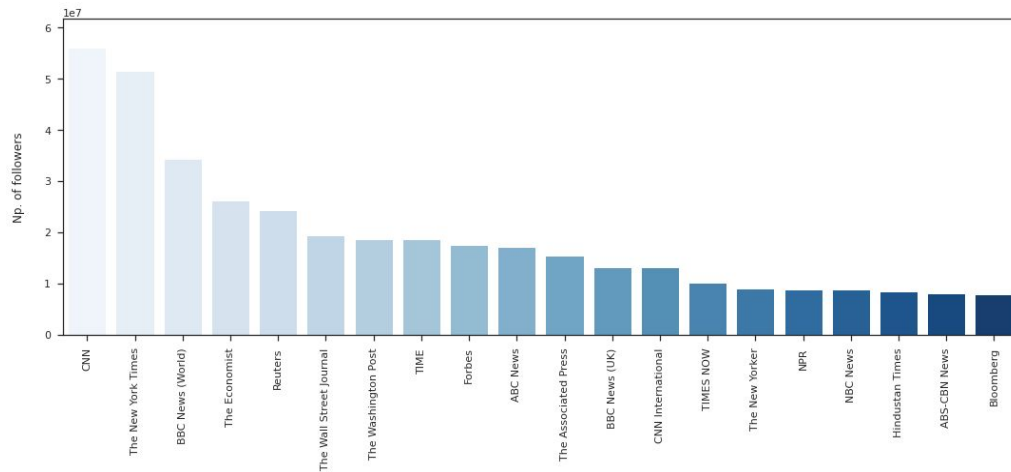
Most Frequent Source
What tools are the most used to post tweets

Data transformation - Text Vectorization

In order to make text useful for the machine learning algorithms, it has to be turned into features.
The techniques can be referred to as "*Text Vectorization*", since they all aim at one purpose: turning text into vectors, that can be then fed to machine learning models in a classical way.

	created_at	text	label
0	2022-01-28 22:00	@FamiLee_Farm @0xBingBong I first dissented in...	negative
1	2022-01-28 22:00	A disaster.\n\nThe sooner aid agencies worldwi...	negative
2	2022-01-28 22:00	With the very real threat of the government tr...	negative
3	2022-01-28 22:00	Bitcoin mining's energy use has more than trip...	neutral
4	2022-01-28 22:00	Tori Zero NFT Project Launched a Joint Coopera...	positive
...	...	...	...
4761	2022-01-22 00:00	El Salvador President Nayib Bukele, who's know...	negative
4762	2022-01-22 00:00	And here's the news from today, by @nicola_m_w...	neutral
4763	2022-01-22 00:00	Microstrategy falls 18% after SEC tells them t...	negative
4764	2022-01-22 00:00	"Cryptocurrencies are no longer an isolated ri...	negative
4765	2022-01-22 00:00	@ScottWapnerCNBC @carlquintanilla The Bitcoin ...	neutral

4766 rows × 5 columns

```
array([[5869,    23,    51, ..., 3955,     6,     1],
       [   0,     0, 5870, ..., 3956,     6,     1],
       [   0,     0,     0, ...,  208,     6,     1],
       ...,
       [   0,     0,   110, ...,   16,     1, 2452],
       [   0,   157,   687, ...,  143, 2546,     1],
       [   0,     0,     0, ...,    6,     1,     4]], dtype=int32)


array([[ 0.        ,  0.        ,  0.        , ...,  0.        ,
         0.        ,  0.        ],
       [ 0.32527   ,  1.20930004,  0.15962   , ..., -1.10020006,
        -0.91891003,  0.48025   ],
       [ 1.21739995,  1.01839995, -0.60869998, ..., -0.88479   ,
         0.58089   , -0.09434   ],
       ...,
```

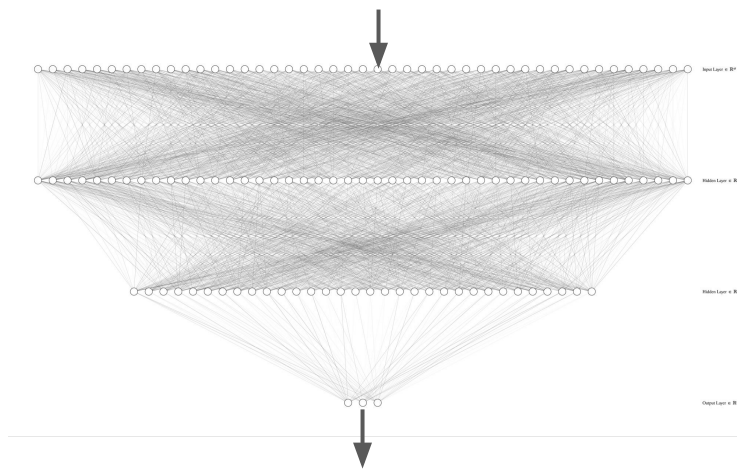*Above: sample of the twitter dataset*
Right: Dataset converted to arrays

Model architecture

3-way classification of tweets content using Neural Networks and Natural Language Processing (NLP)

Input

```
array([[5869,   23,   51, ..., 3955,    6,    1],
       [   0,    0, 5870, ..., 3956,    6,    1],
       [   0,    0,    0, ...,  208,    6,    1],
       ...,
       [   0,    0,  110, ...,   16,    1, 2452],
       [   0,  157,  687, ...,  143, 2546,    1],
       [   0,    0,    0, ...,    6,    1,    4]], dtype=int32)
```
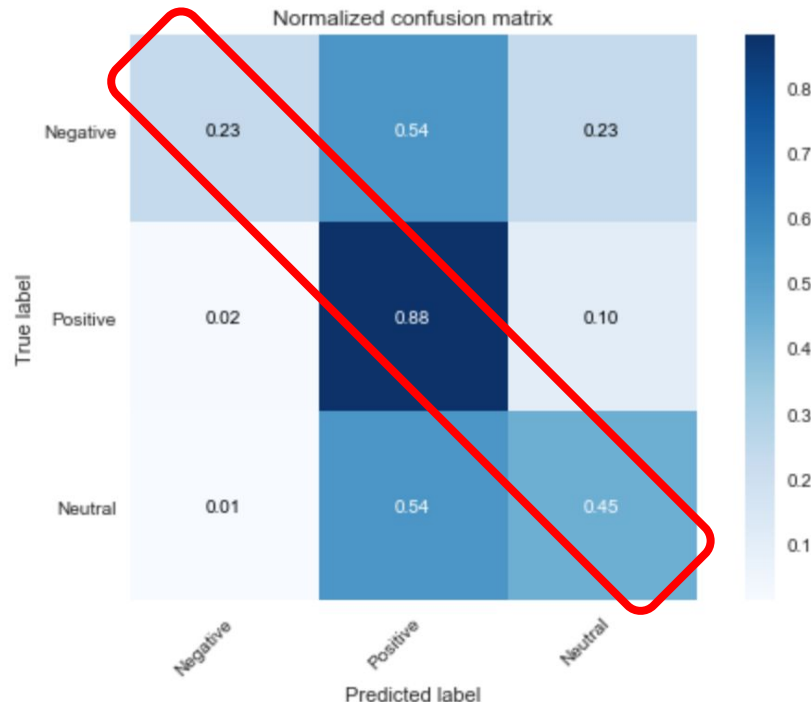

Output:
Each value in the output is interpreted as the probability of membership for each class.

```
{0: 'Negative emotion', 1: 'Neutral emotion', 2: 'Positive emotion'}
      [[0.04 0.47 0.49]] [ 2 ] Positive emotion
```

Model results and future improvements

Confusion matrix reflects the accuracy of the predictions.


Normalized confusion matrix

- In the red highlight we can see the percentages of the correct predictions. i.e. 88% for the positive emotion.
- The lower accuracy for the other classes is a results of insufficient data.
- If there was more time and data, the following further improvements to the model could be done:
  - Model accuracy for the negative and neutral labels
  - More advanced data visualizations and analysis
  - Model deployment

General model accuracy: 71%

Results

- **Negative sentiment** has impact on the trading **volume** and price **volatility**;
- Negative sentiment has has bigger impact on the price change then positive and neutral ;
- Tweets' volume does not show any significant impact on the price change (correlation = 0.2);
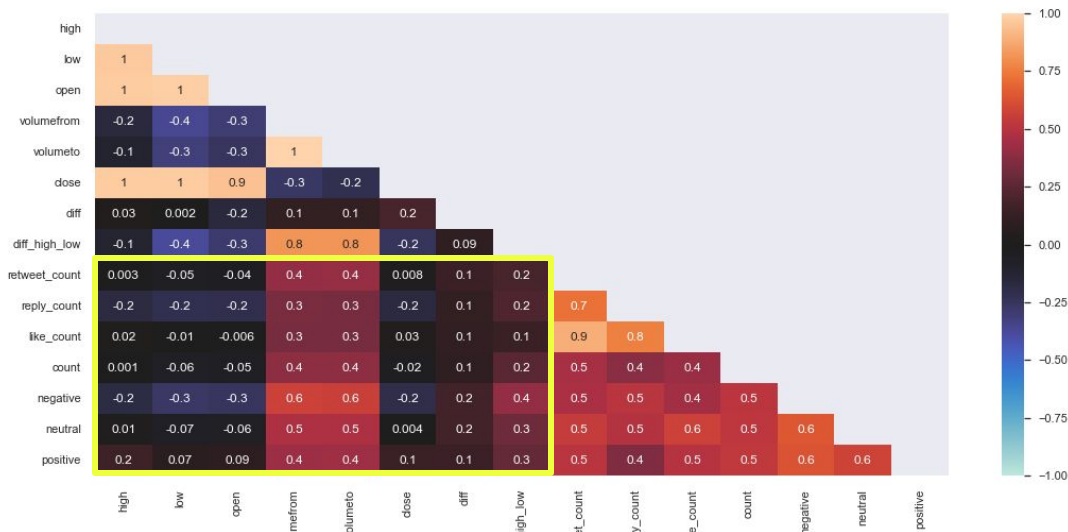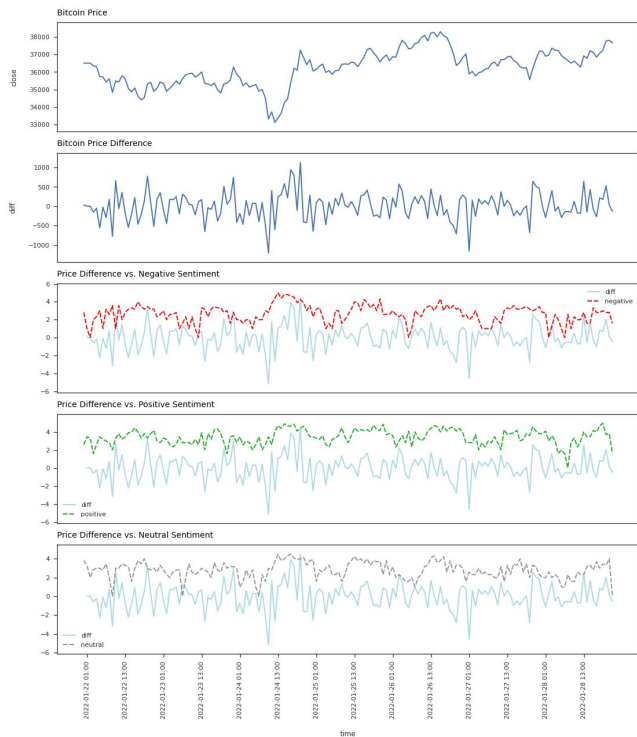
Results

- The highest correlation for price difference vs. single words has '**trying**' with **0.32** correlation coefficient;
- The highest correlation for price difference vs. bigrams has 'are, **getting**' with **0.27** correlation coefficient;
- Main tools used for posting tweets are Twitter Web App and Twitter for iPhone.

created_at	tokenized_text	bigrams
2022-01-22 00:00	[pretty, sure, handbags, @emimelker, bought, p...	[(Pretty, sure), (sure, the), (the, handbags),...
2022-01-22 01:00	[crypto, hodl, republican, party, https://t.co...	[(Crypto, has), (has, a), (a, HODL), (HODL, on...
2022-01-22 02:00	[@solelawd_, @cryptocom, hi, could, please, se...	[(@SoleLawd_, @cryptocom), (@cryptocom, Hi), (...
2022-01-22 03:00	[@cryptobuzz_blog, this, bitcoin, that's, i, m...	[(@CryptoBuzz_Blog, This), (This, is), (is, wh...
2022-01-22 04:00	[el, salvador, purchased, another, 15, million...	[(El, Salvador), (Salvador, has), (has, purcha...

	diff
find	0.282541
trying	0.324410
pasta	0.269251
all-time	0.259101
@bitcoin_phan	0.293646
@samouraiwallet	0.293646
@coldcardwallet	0.293646

	diff
(are, getting)	0.272169

Next Steps

Data can be fed into the trained model in real time through various tools i.e Twitter API or Python libraries and instantly generate insights about a brand or service.

- Gather more data
- Perform analysis of the sentiment and see how it affects the price after certain time period. i.e 1 sec, 30 sec, 1min, etc. Analyze the correlation coefficients for each period.
- Check more languages for tweets. Certain languages like time zones might be correlated with certain languages with might have an impact of price fluctuations.
- Check more sentiment sources like Reddit

Thank you!

Contact details:
- *email: szafranskifilip@gmail.com*
- *LinkedIn : https://www.linkedin.com/in/szafranskifilip/*

References:
1. https://www.linkedin.com/pulse/how-much-can-you-make-trading-bot-my-experience-month-gothiredd
2. https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/
3. https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/
4. https://www.ibm.com/cloud/learn/recurrent-neural-networks
5. https://www.investopedia.com/terms/c/correlationcoefficient.asp