Impact of Emotion on Bitcoin

Roadmap

1. Develop a dataset of Reddit comments from the **r/cryptocurrency** subreddit
2. Fine tune a transformer model on an emotion classification dataset
3. Repurpose the model to predict the emotion exhibited by Reddit comments
4. Investigate trends between certain emotions and the price of Bitcoin

1

Data

How we obtained and preprocessed the data

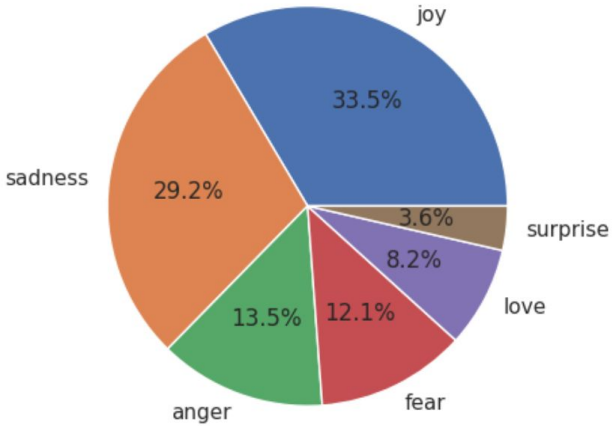
Reddit Comments

- Used PMAW, a wrapper for PushShift API
- Pulled 16 million comments from r/cryptocurrency from 1st May 2021 to 1st April 2022
- Removed bot comments, duplicates, common comments

Emotion Dataset

- This is a dataset of social media messages with six different emotions: anger, fear, joy, love, sadness, ar

Train	16000
Validation	2000
Test	2000



Bitcoin Price

- We used **pandas_dataloader**, which is a wrapper for Yahoo Finance API
- Obtained the daily prices for the date range 1st May 2021 to 1st April 2022

	High	Low	Open	Close	Volume	Adj Close
2021-05-01T...	58448	57052	57714	57828	42836427360	57828
2021-05-02...	57902	56141	57825	56631	38177405335	56631
2021-05-03...	58973	56590	56620	57200	51713139031	57200
2021-05-04...	57214	53191	57214	53333	68564706967	53333
2021-05-05...	57911	52969	53252	57424	69241316747	57424

2

Model

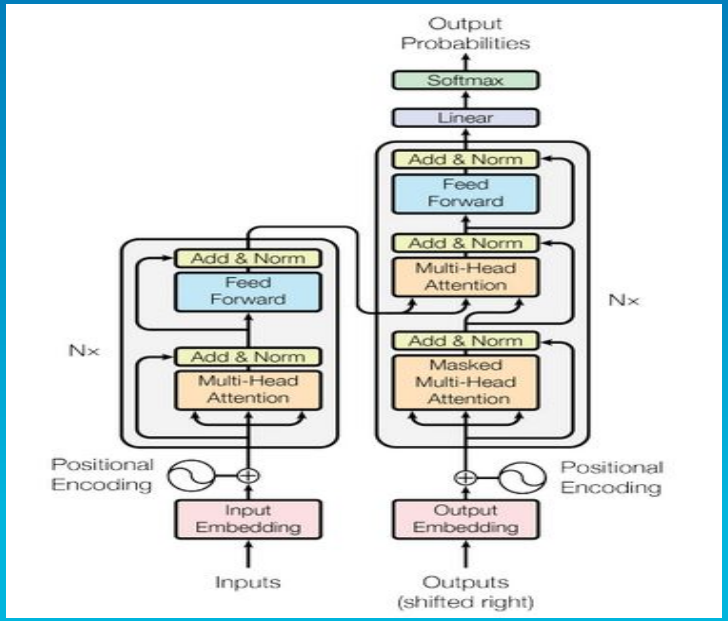
What model did we use and why?

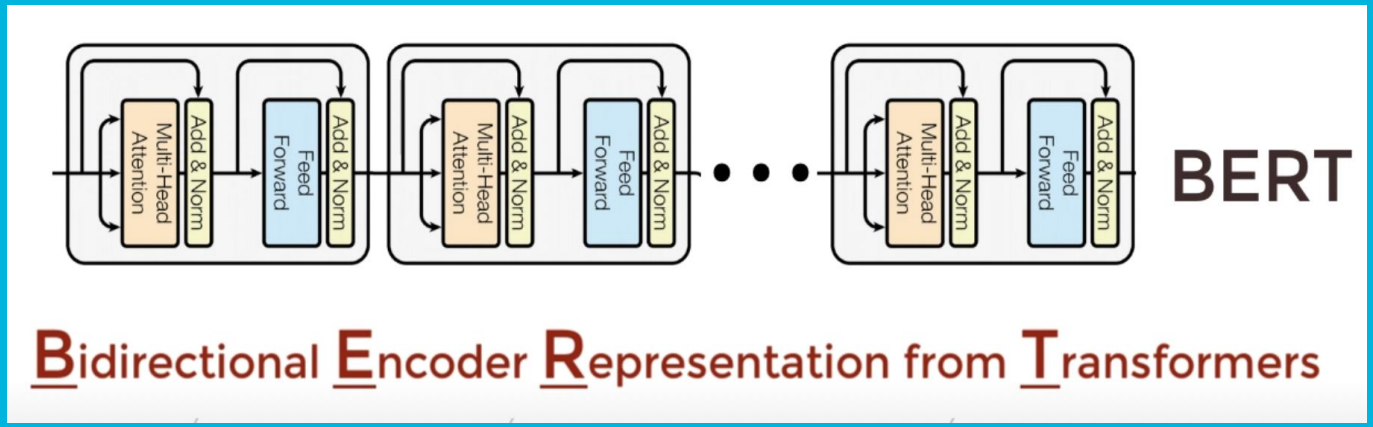
BERT

- TRANSFORMER
- BERT OBJECTIVE
- BERT PRETRAINING AND FINE TUNING

TRANSFORMER FLOW:

- ENCODER
- DECODER

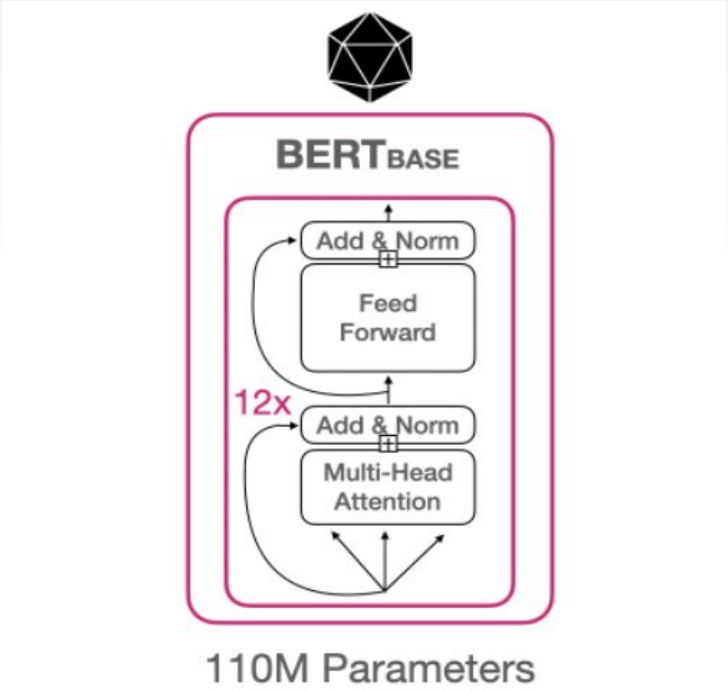




BERT OBJECTIVE:

Problem to solve:

- Language Translation
- Question answering
- Sentiment analysis
- Text summarisation



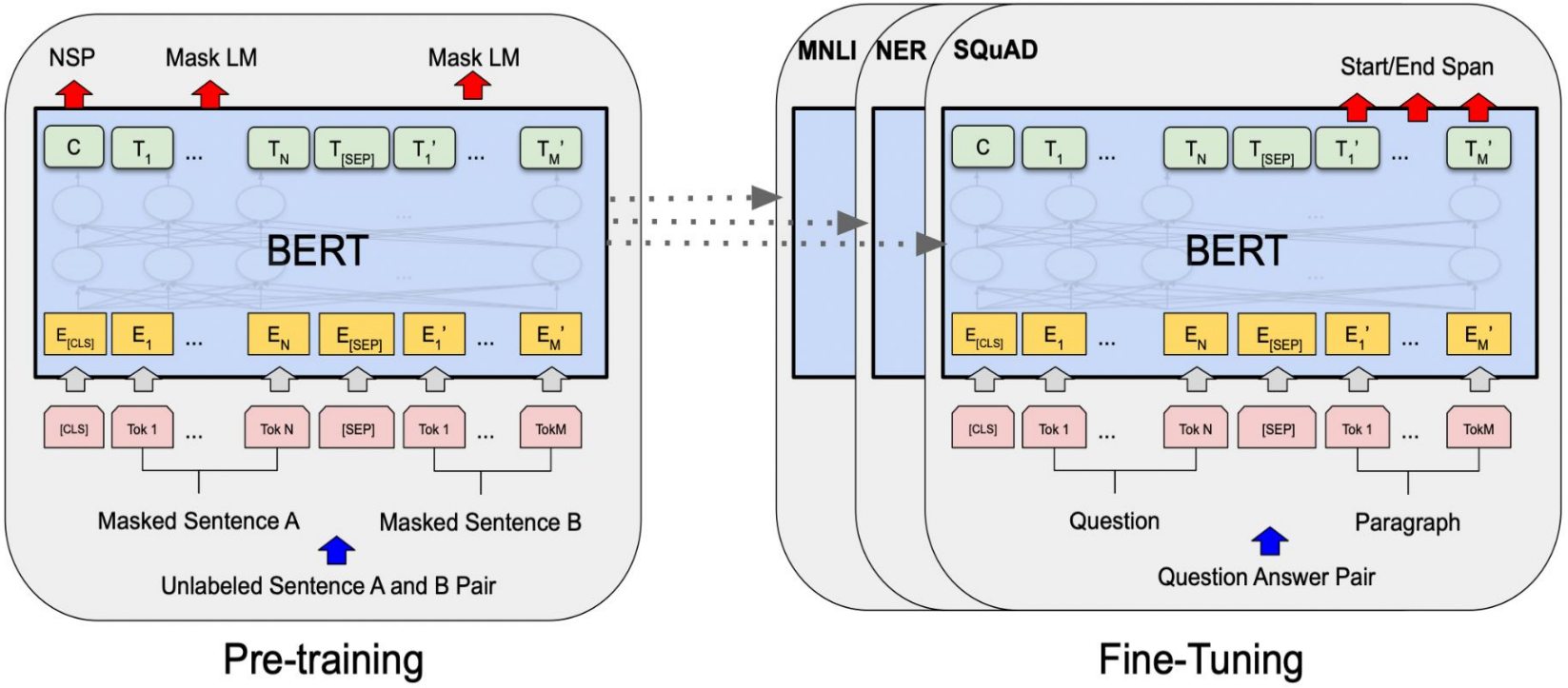
BERT Pretraining and fine tuning

1) Pretraining:

Trains on 2 unsupervised tasks simultaneously

- MLM(Masked language model): a sentence with random words filled with mask to output mask tokens
- NSP(Next sentence prediction): Takes 2 sentences as input and determines if second sentence follows the first

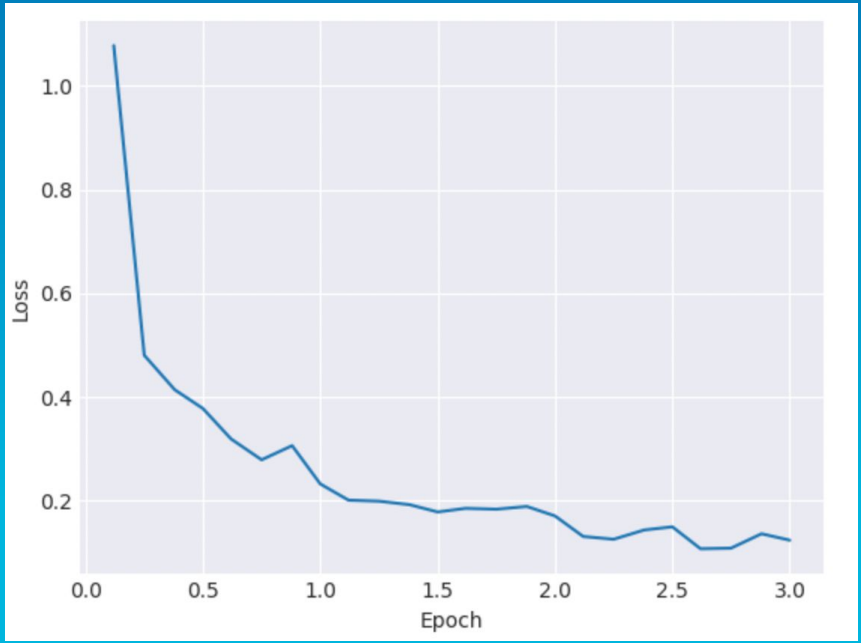
2) Fine-Tuning: Train on NLP tasks


NSP
Mask LM
Mask LM
C
T1
...
TN
T[SEP]
T1'
...
TM'
BERT
E[CLS]
E1
...
EN
E[SEP]
E1'
...
EM'
[CLS]
Tok 1
...
Tok N
[SEP]
Tok 1
...
TokM
Masked Sentence A
Masked Sentence B
Unlabeled Sentence A and B Pair
Pre-training
MNLI
NER
SQuAD
Start/End Span
Question
Paragraph
Question Answer Pair
Fine-Tuning

Training BERT

- Used the Hugging Face transformer APIs to train our model on the emotions dataset
- The dataset has three splits ;train,validations and test.
- Used the BERT tokenizer to convert our text inputs into integer encodings and boolean attention masks.
- Collate the encodings and convert them to same token length using data collator.

Hyperparameters	Values
Learning rate	2e-5
Batch size	4
num_train_epochs	3
weight_decay	0.01



Epoch	Eval F1 Score	Eval Loss
0	0.899	0.293
1	0.919	0.203
2	0.921	0.227

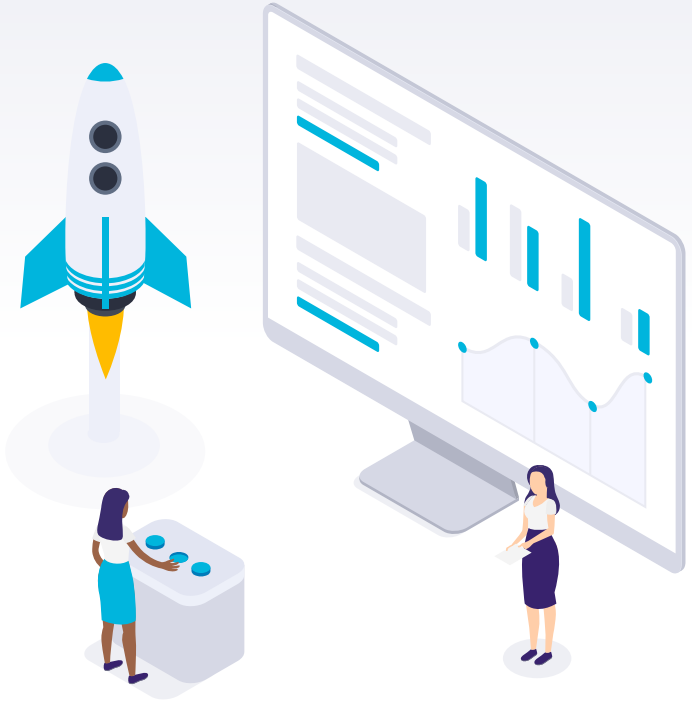
Testing BERT

	precision	recall	f1-score	support
sadness	0.95	0.97	0.96	581
joy	0.96	0.94	0.95	695
love	0.80	0.91	0.85	159
anger	0.94	0.89	0.92	275
fear	0.87	0.91	0.89	224
surprise	0.79	0.70	0.74	66
accuracy			0.93	2000
macro avg	0.89	0.89	0.88	2000
weighted avg	0.93	0.93	0.93	2000

3

Results

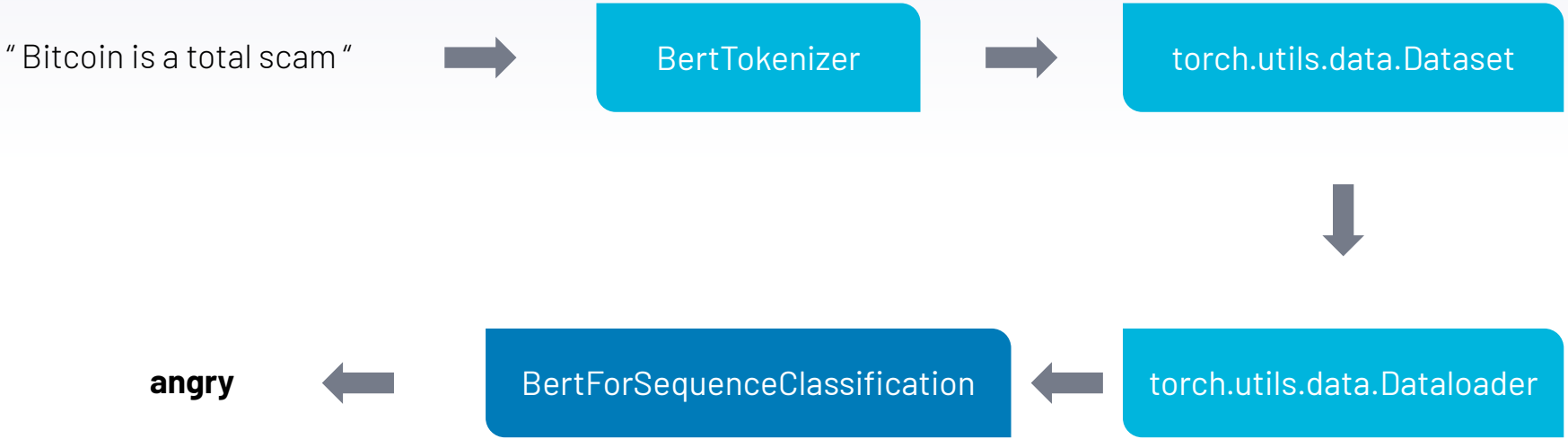
Let’s generate some predictions



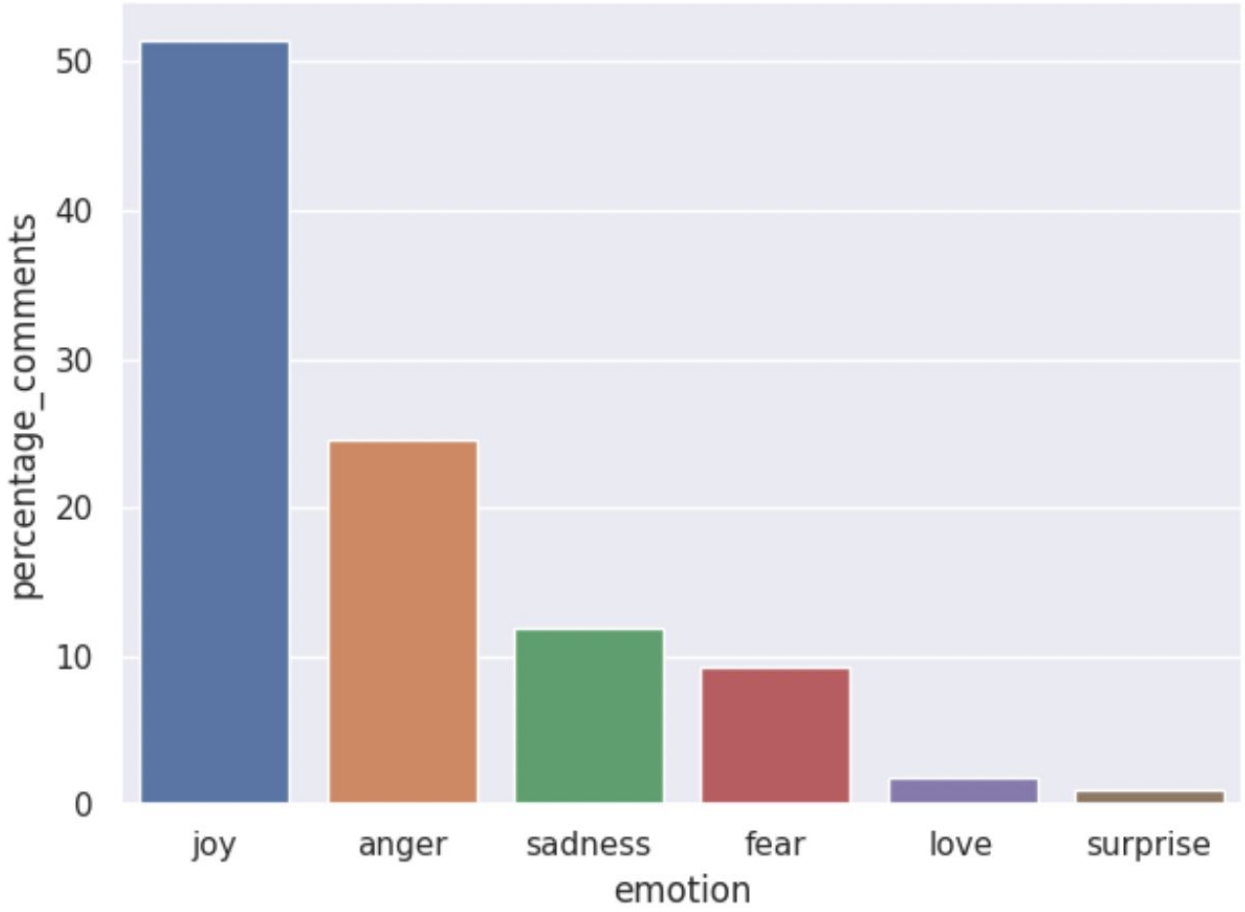
Inference

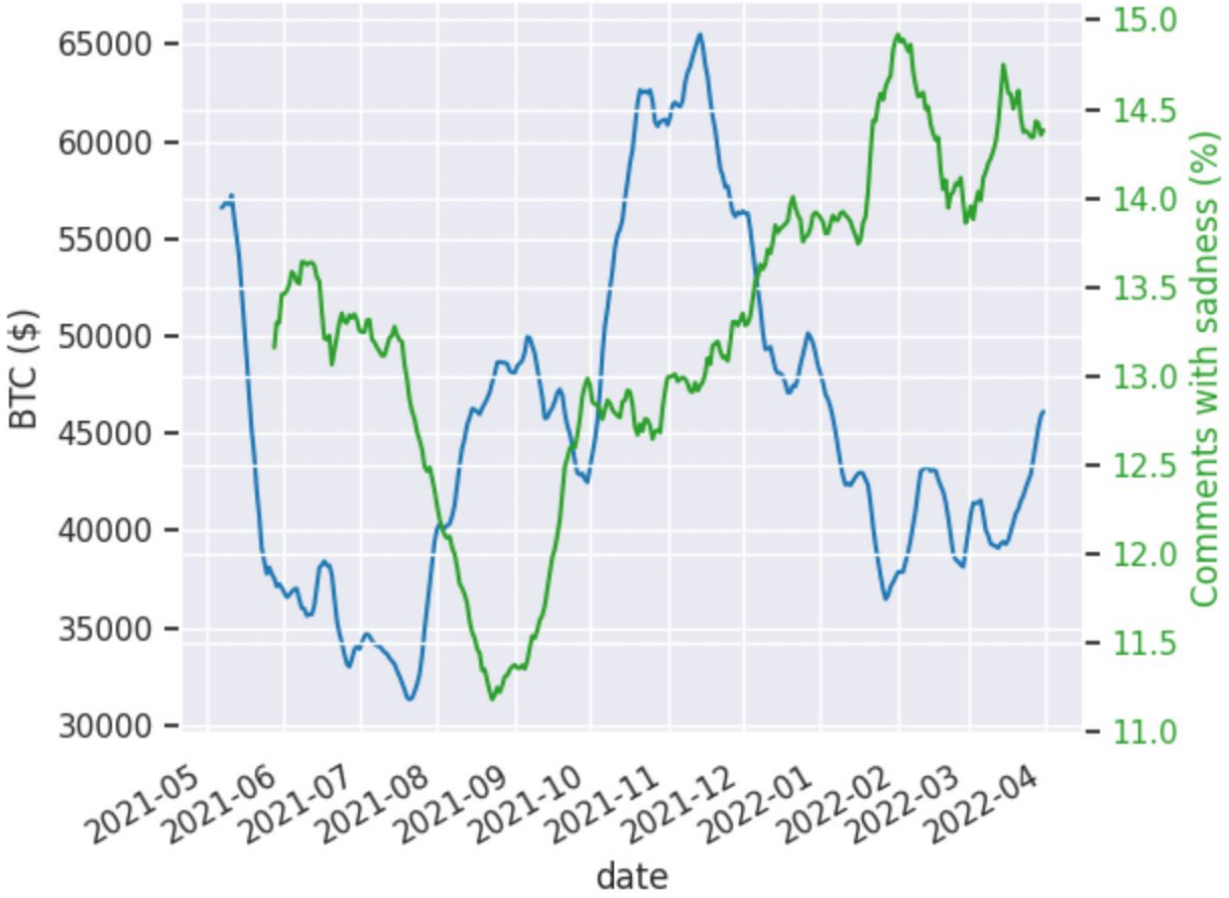
- We repurpose the model to predict emotion on the Reddit comments
- We are interested in **percentage of comments exhibiting a particular emotion on a particular day**
- We take a random subset of the comments due to computational constraints and use it to approximate this percentage

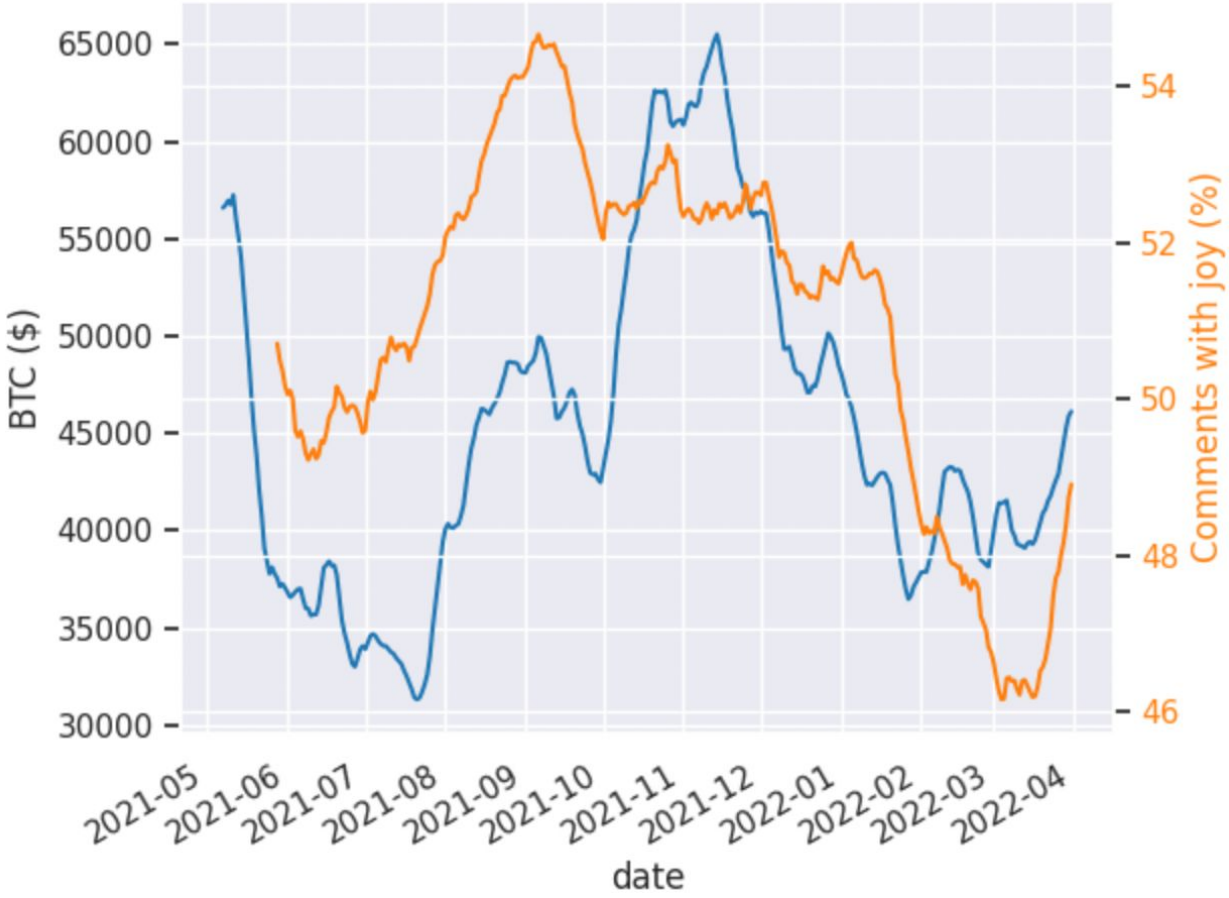
Inference

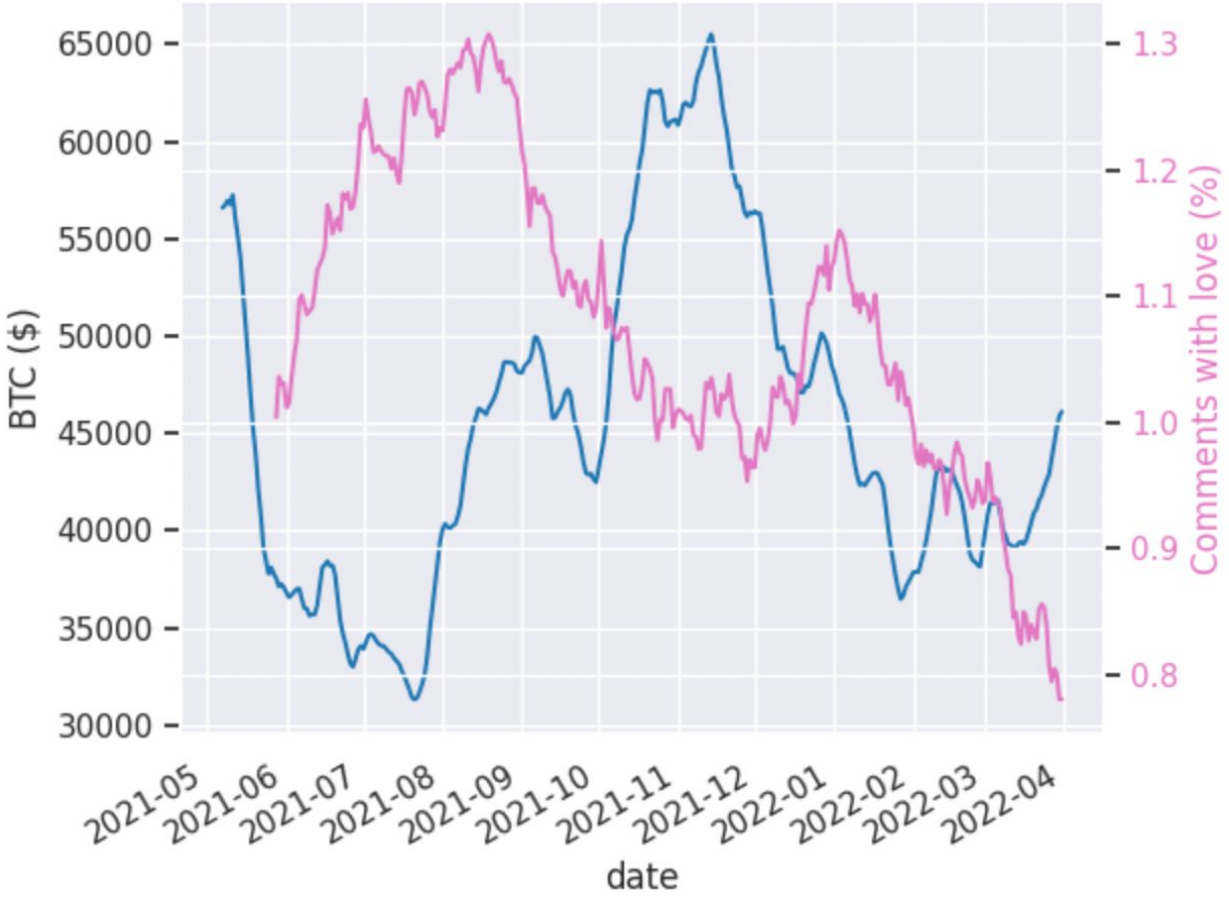

“ Bitcoin is a total scam “
BertTokenizer
torch.utils.data.Dataset
torch.utils.data.Dataloader
BertForSequenceClassification
angry

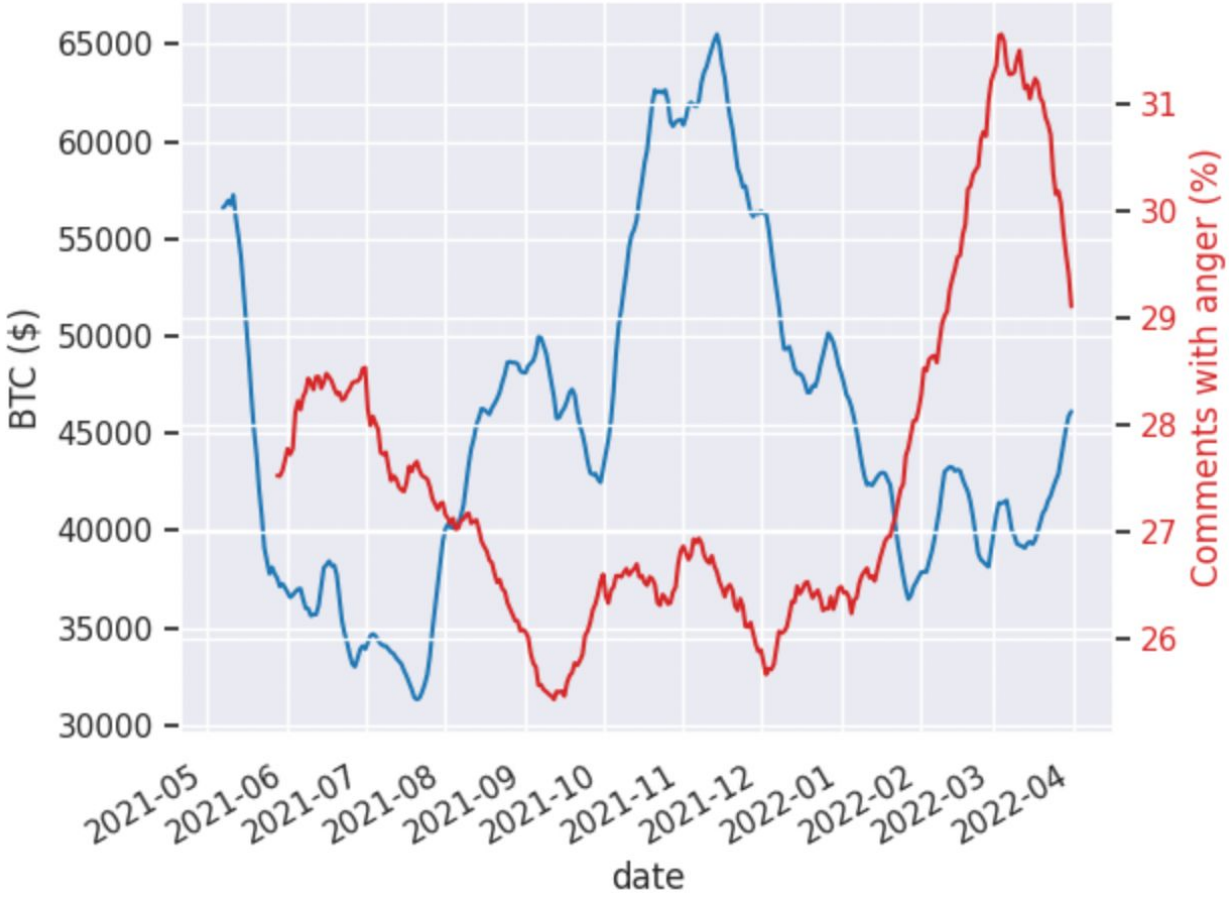
	created_utc	author	body	emotion
23803	1636293110.0...	Accompli...	When in doubt just zoom out, A...	fear
23804	1644742958.0...	archer4364	They'll be back to buy high one day	joy
23805	1621794490.0...	90DayF	When I wake up, I'd really love to...	joy
23806	1637413444.0...	Zulunatio...	A lot of people on here talk a big...	fear
23807	1633175561.0...	plurBUD...	Definitely not until the market is ...	joy
23808	1634535191.0...	MoodSog...	And now with governance it will ...	joy
23809	1640323833.0...	YamahaF...	It kind of scares me with social plat.	fear
23810	1634211630.0...	Csdsmall...	Because they refuse to give up ...	anger
23811	1627281146.0...	rakinlord22	Oh got what did he do now?	joy
23812	1624550113.0...	pc1e0	:D\n\nWould you trust your clone...	joy
23813	1621401112.00...	rideeast6...	Sell sell sell .. it is crashing !!	sadness
23814	1629576741.0...	groypley90	Timing the dip is tricky. It could e...	joy

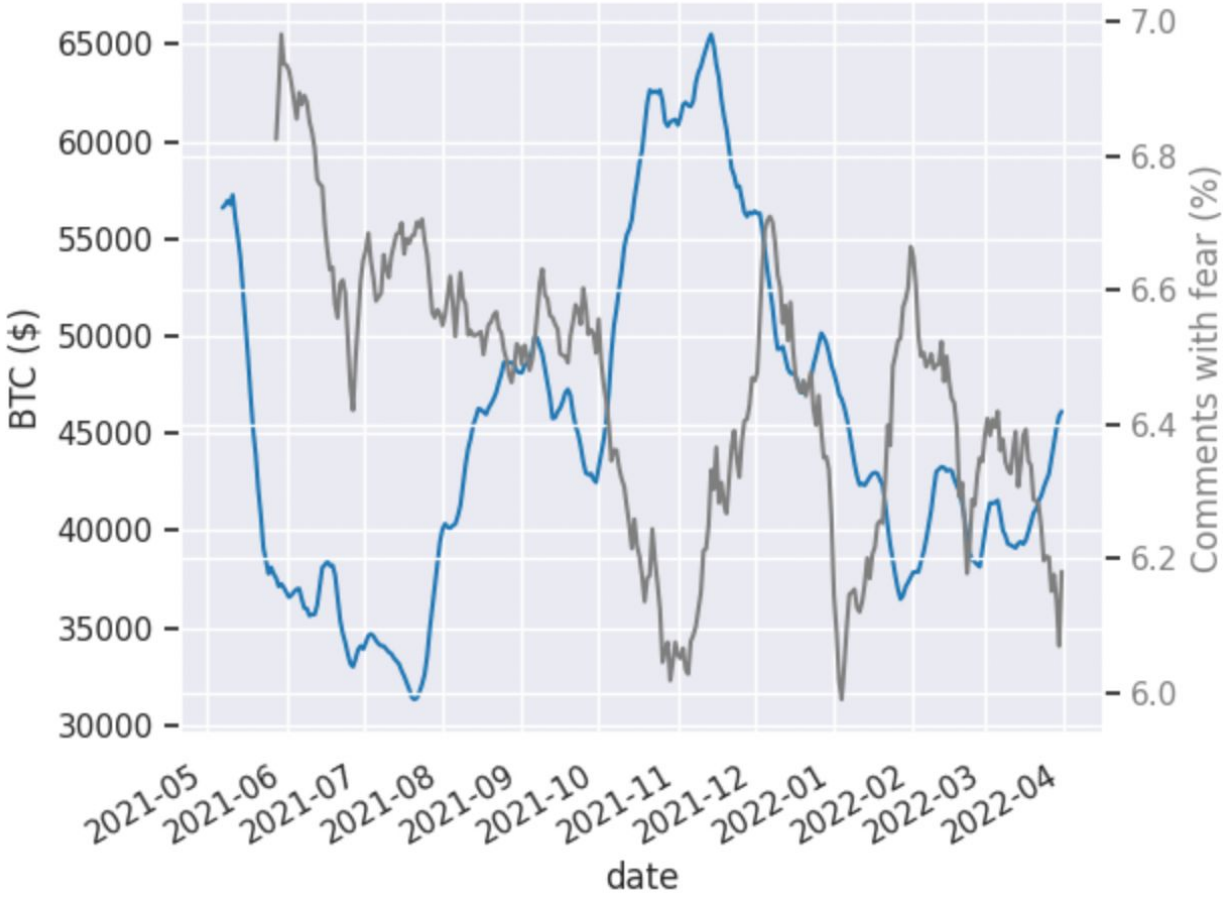

percentage_comments
50
40
30
20
10
0
joy
anger
sadness
fear
love
surprise
emotion

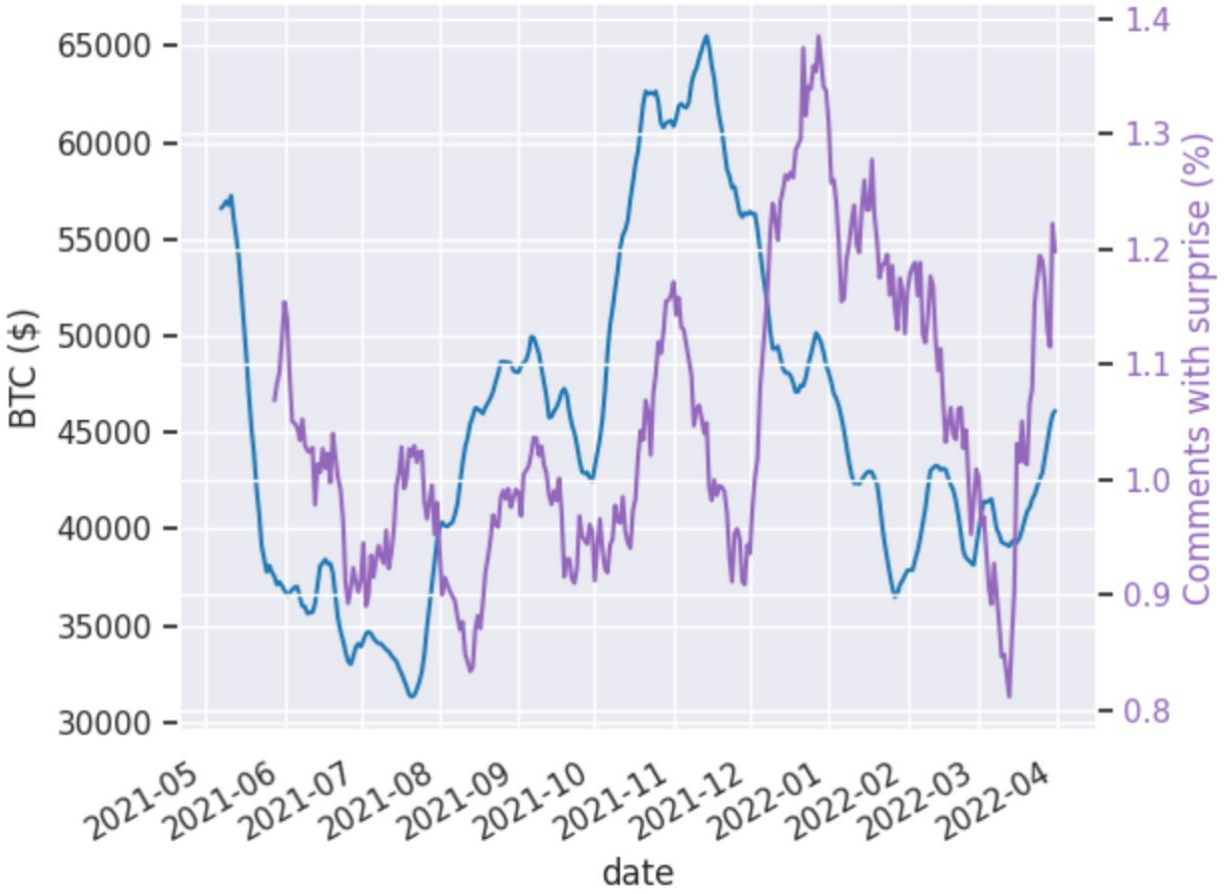

65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
15.0
14.5
14.0
13.5
13.0
12.5
12.0
11.5
11.0
Comments with sadness (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date

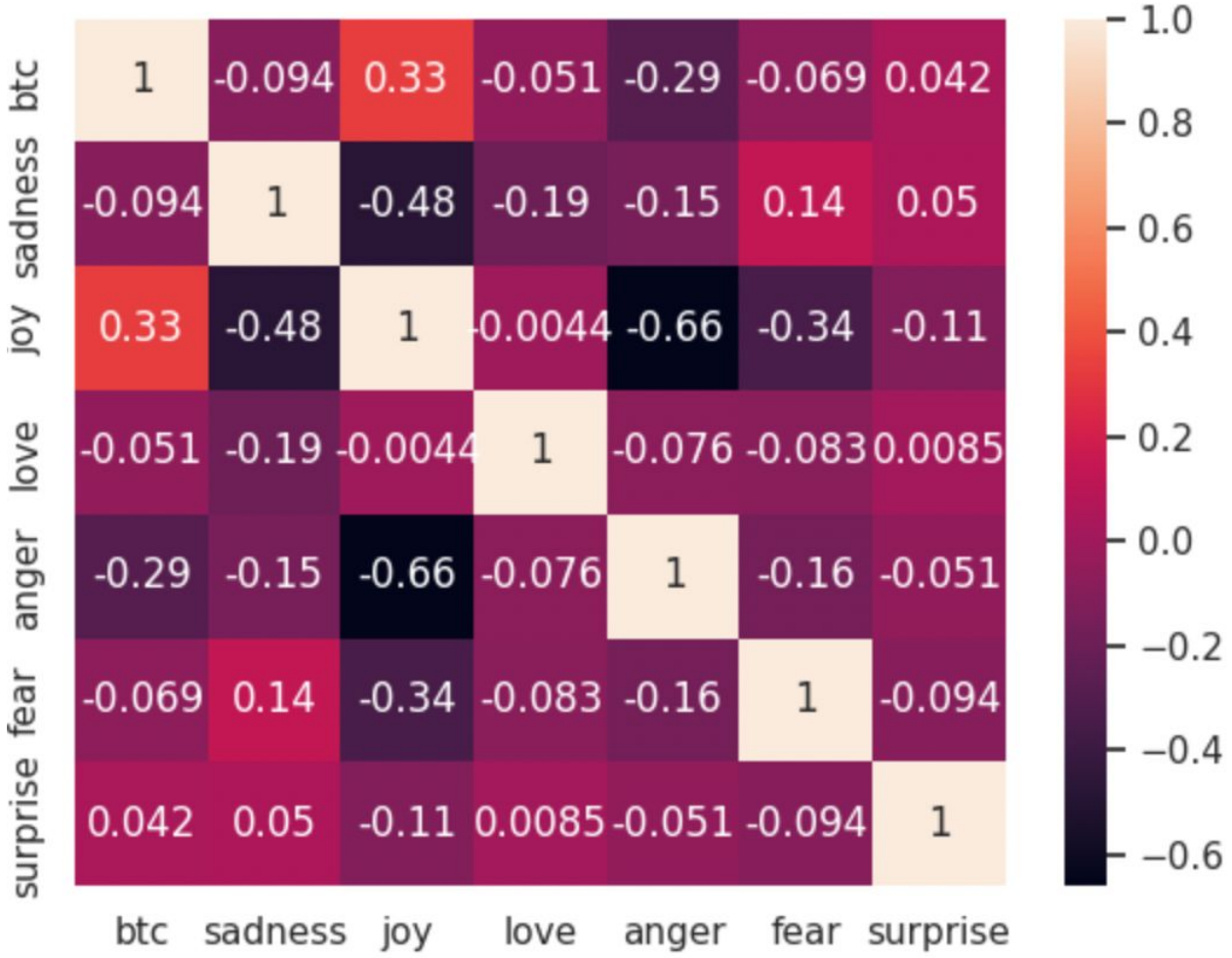

65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
54
52
50
48
46
Comments with joy (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date


65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
1.3
1.2
1.1
1.0
0.9
0.8
Comments with love (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date


65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
31
30
29
28
27
26
Comments with anger (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date


65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
7.0
6.8
6.6
6.4
6.2
6.0
Comments with fear (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date


65000
60000
55000
50000
45000
40000
35000
30000
BTC ($)
1.4
1.3
1.2
1.1
1.0
0.9
0.8
Comments with surprise (%)
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
date


btc
sadness
joy
love
anger
fear
surprise
1 -0.094 0.33 -0.051 -0.29 -0.069 0.042
-0.094 1 -0.48 -0.19 -0.15 0.14 0.05
0.33 -0.48 1 -0.0044 -0.66 -0.34 -0.11
-0.051 -0.19 -0.0044 1 -0.076 -0.083 0.0085
-0.29 -0.15 -0.66 -0.076 1 -0.16 -0.051
-0.069 0.14 -0.34 -0.083 -0.16 1 -0.094
0.042 0.05 -0.11 0.0085 -0.051 -0.094 1
btc
sadness
joy
love
anger
fear
surprise
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Conclusion

- Despite being trained on a different dataset, BERT was good enough to generate good predictions
  - This is evident by the correlation with anger and joy.
- Emotion, as opposed to logic, has a great impact on Bitcoin price and should be considered by market players.
  - This is evident by the graphs

THANKS!

Any questions?