Social Sentiment Analysis

TEAM MEMBERS INCLUDES:
KSENIJA CADAJEVA
CHARBEL NEHME
ANTHONY MURA
JACKY HONG

Pitch

WHY ARE WE DOING THIS?

- BUILD TRADING BOTS TO MAKE MONEY!
- TWITTER SENTIMENT ANALYSIS
- GOOGLE TRENDS ANALYSIS
- DOW JONES ANALYSIS

MODUELS/LIBRARIES/APIS

MACHINE LEARNING MODULES USED

> TIME SERIES ANALYSIS

LINEAR/MULTIPLE REGRESSION

> NATURAL LANGUAGE PROCESSING

NLTK

VADER SENTIMENT ANALYSIS

> DEEP LEARNING

SICKIT-LEARN

TENSORFLOW

KERAS

API USED

> TWITTER API

> NEWS API

> ALPACA API

TWITTER ANALYSIS BY ANTHONY

Elon Musk
@elonmusk

Good thing I never tweet anything controversial

HOW DOES HUMAN SENTIMENT AFFECT ASSET PRICES?

	text	text_compound	text_pos	text_neu	text_neg	text_sent
0	bitcoin morreu	0.0000	0.000	1.000	0.000	0
1	RT bitcoinvisions BITCOIN VISION AIRDROP PROG...	0.6633	0.280	0.720	0.000	1
2	RT Chain Chains 50000 Bitcoin Giveaway back We...	0.0000	0.000	1.000	0.000	0
3	RT PeterSchiff Attention Bitcoin HODLers Its n...	0.0000	0.000	1.000	0.000	0

MESSY DATA

date_New	close	username	name	tweet	text_compound	text_pos	text_neu	text_neg	text_sent
2018-10-22	0.001089	elonmusk	Elon Musk	vicentes Grimezsz Wanna buy Bitcoin httpstcoZ...	0.0000	0.000	1.000	0.000	0.0
2018-10-22	0.001089	elonmusk	Elon Musk, the 2nd	vicentes Grimezsz Wanna buy Bitcoin httpstcoZ...	0.0000	0.000	1.000	0.000	0.0
2020-01-10	0.018484	elonmusk	Elon Musk	bitcoinconf	0.0000	0.000	1.000	0.000	0.0
2020-01-10	0.018484	elonmusk	Elon Musk	Bitcoin not safe word	-0.3412	0.000	0.555	0.445	-1.0
2020-05-01	-0.000714	elonmusk	Elon Musk	Bitcoin How much anime Bitcoin httpstcoitqRslFNcb	0.0000	0.000	1.000	0.000	0.0
2020-05-01	-0.000714	elonmusk	Elon Musk	Bitcoin	0.0000	0.000	1.000	0.000	0.0
2020-05-15	0.039834	elonmusk	Elon Musk	jkrowling I still Bitcoins btw	0.0000	0.000	1.000	0.000	0.0
2020-05-15	0.039834	elonmusk	Elon Musk	jkrowling Pretty much although massive currenc...	0.5859	0.242	0.758	0.000	1.0
2020-11-16	-0.049578	elonmusk	Elon Musk	MaisieWilliams Toss bitcoin ur Witcher	0.0000	0.000	1.000	0.000	0.0
2020-12-20	-0.008368	elonmusk	Elon Musk	Bitcoin almost bs fiat money	0.0000	0.000	1.000	0.000	0.0
2020-12-20	-0.008368	elonmusk	Elon Musk	Bitcoin safe word	0.4404	0.592	0.408	0.000	1.0
2021-02-19	-0.083864	elonmusk	Elon Musk	business To clear I not investor I engineer I ...	-0.2500	0.139	0.697	0.164	-1.0
2021-02-19	-0.083864	elonmusk	Elon Musk	business Teslas action not directly reflective...	-0.1548	0.107	0.759	0.134	-1.0
2021-03-12	0.065461	elonmusk	Elon Musk	BTC Bitcoin anagram TBCThe Boring Company What...	-0.3182	0.000	0.753	0.247	-1.0
2021-03-24	0.020164	elonmusk	Elon Musk	Pay Bitcoin capability available outside US la...	-0.1027	0.000	0.833	0.167	-1.0
2021-03-24	0.020164	elonmusk	Elon Musk	Tesla using internal amp open source software ...	0.0258	0.052	0.948	0.000	0.0
2021-03-24	0.020164	elonmusk	Elon Musk	You buy Tesla Bitcoin	0.0000	0.000	1.000	0.000	0.0
2021-04-27	-0.031186	elonmusk	Elon Musk	stoolpresidente No not I not sold Bitcoin Tesl...	-0.2960	0.000	0.885	0.115	-1.0

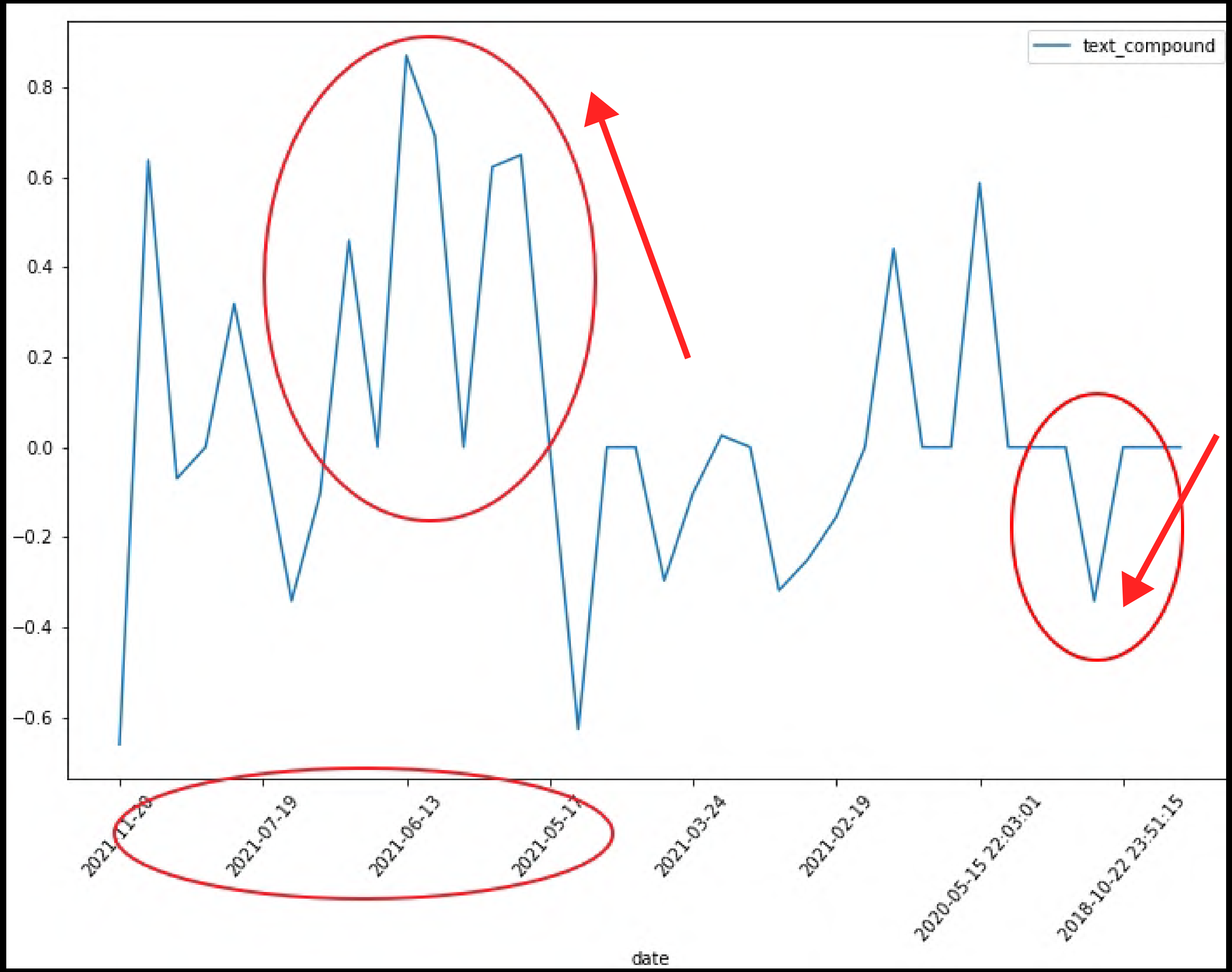
CLEANING TIME

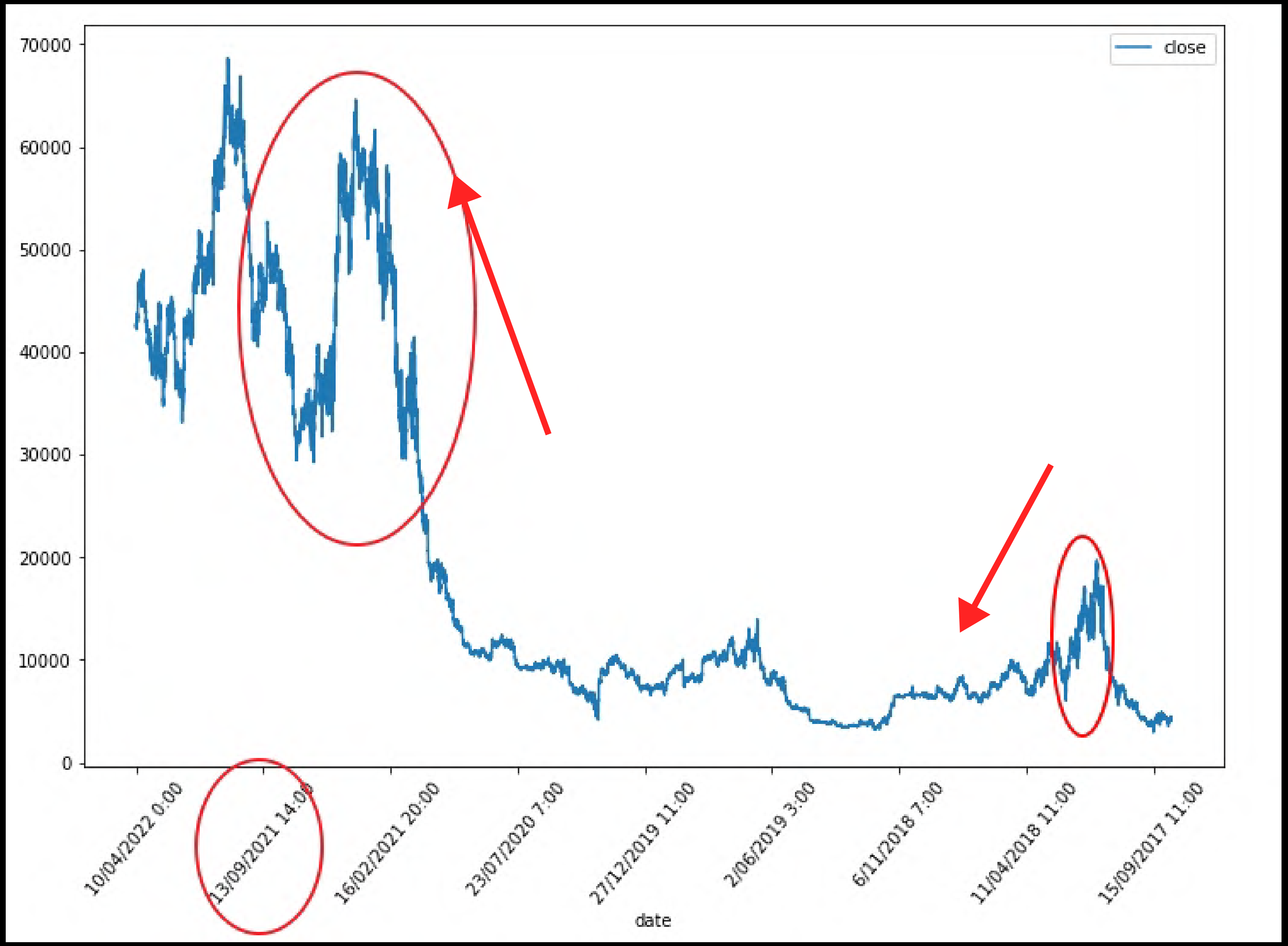
- REMOVE EMOJIS
- REMOVE PUNCTUATION
- REMOVE HYPERLINKS
- REMOVE REPEATING CHARACTERS
- DROP UNWANTED COLUMNS
- TRANSFORMING TWEET ID TO DATE
- SET INDEX AS DATE WITHOUT TIME

MACHINE LEARNING

- SAVE SENTIMENT ANALYSIS DATAFRAME AS CSV
- USE CSV FILES AND STOCK/CRYPTO PRICE DATA TO TRAIN MACHINE LEARNING MODEL
- COMBINE DATAFRAMES AND TRAIN THE MODEL
- 3 DIFFERENT MODELS: BERNOULII NAIVE BAYES, SVM AND LOGISTIC REGRESSION
- SET Y VALUE AS THE ASSET CLOSE PRICE
- SET X VALUE AS THE SENTIMENT VALUE

ELON MUSK BITCOIN TWEET SENTIMENT EFFECTS



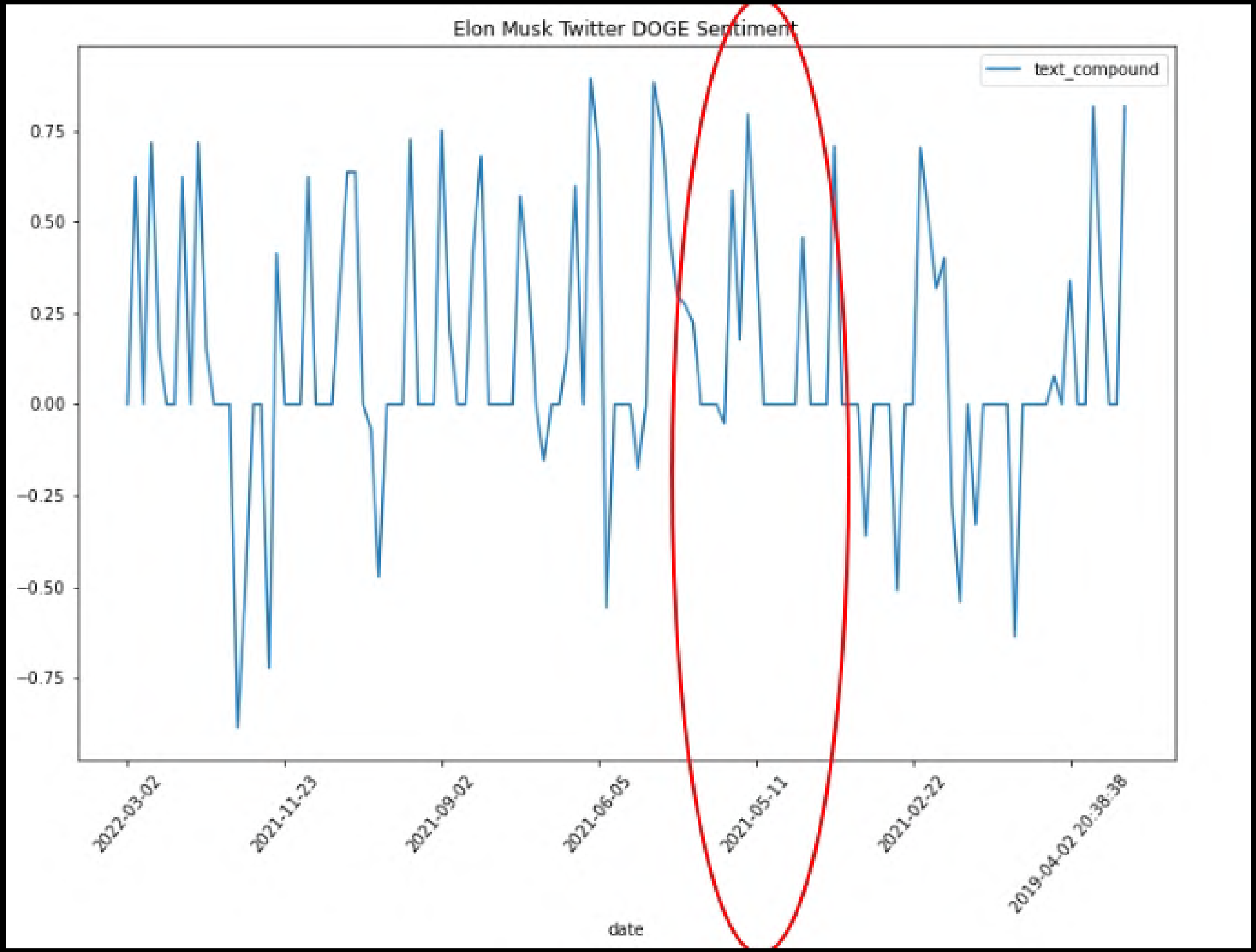


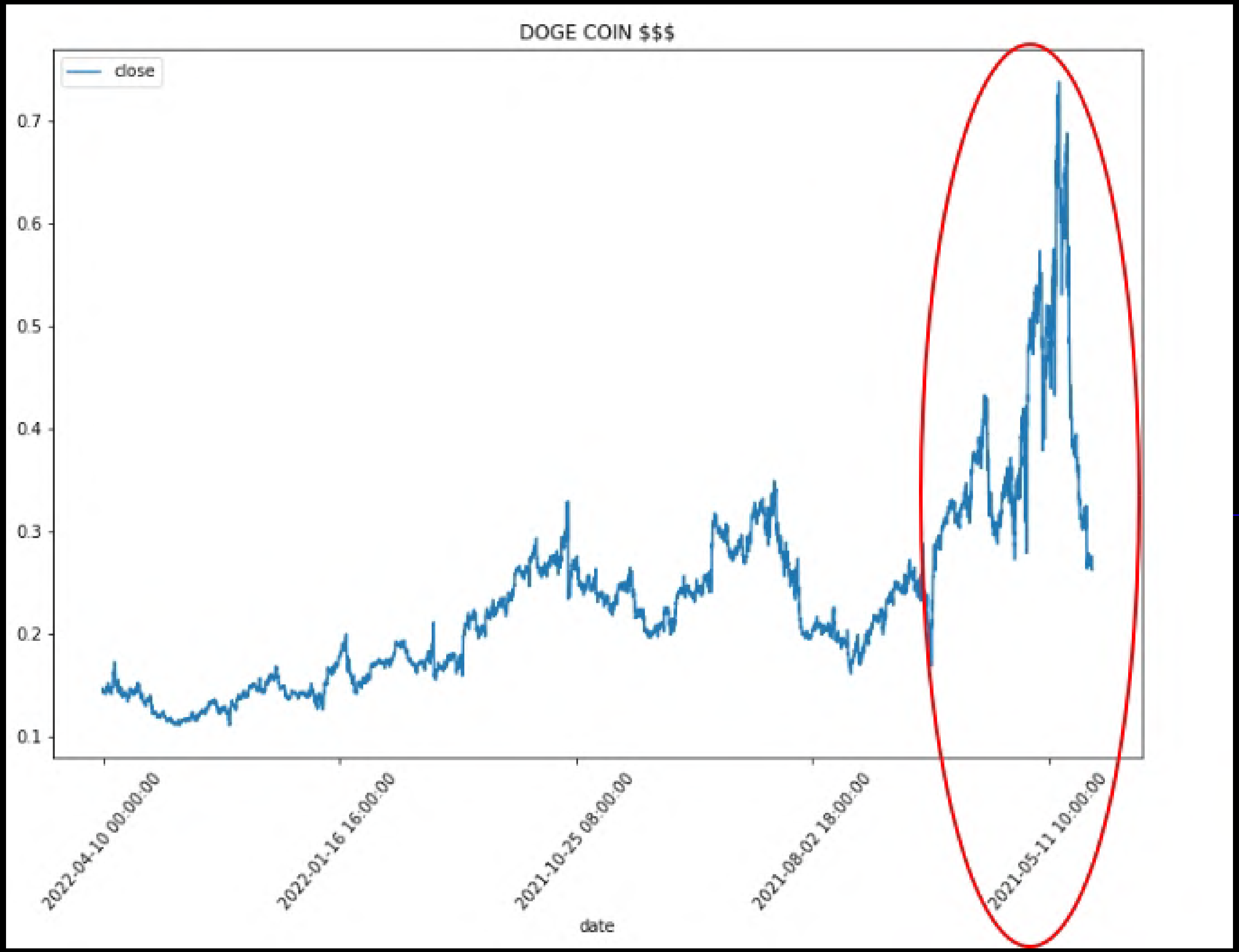
Elon Musk Bitcoin Tweet Sentiment Compound Score

Bitcoin Close Price

Pitch

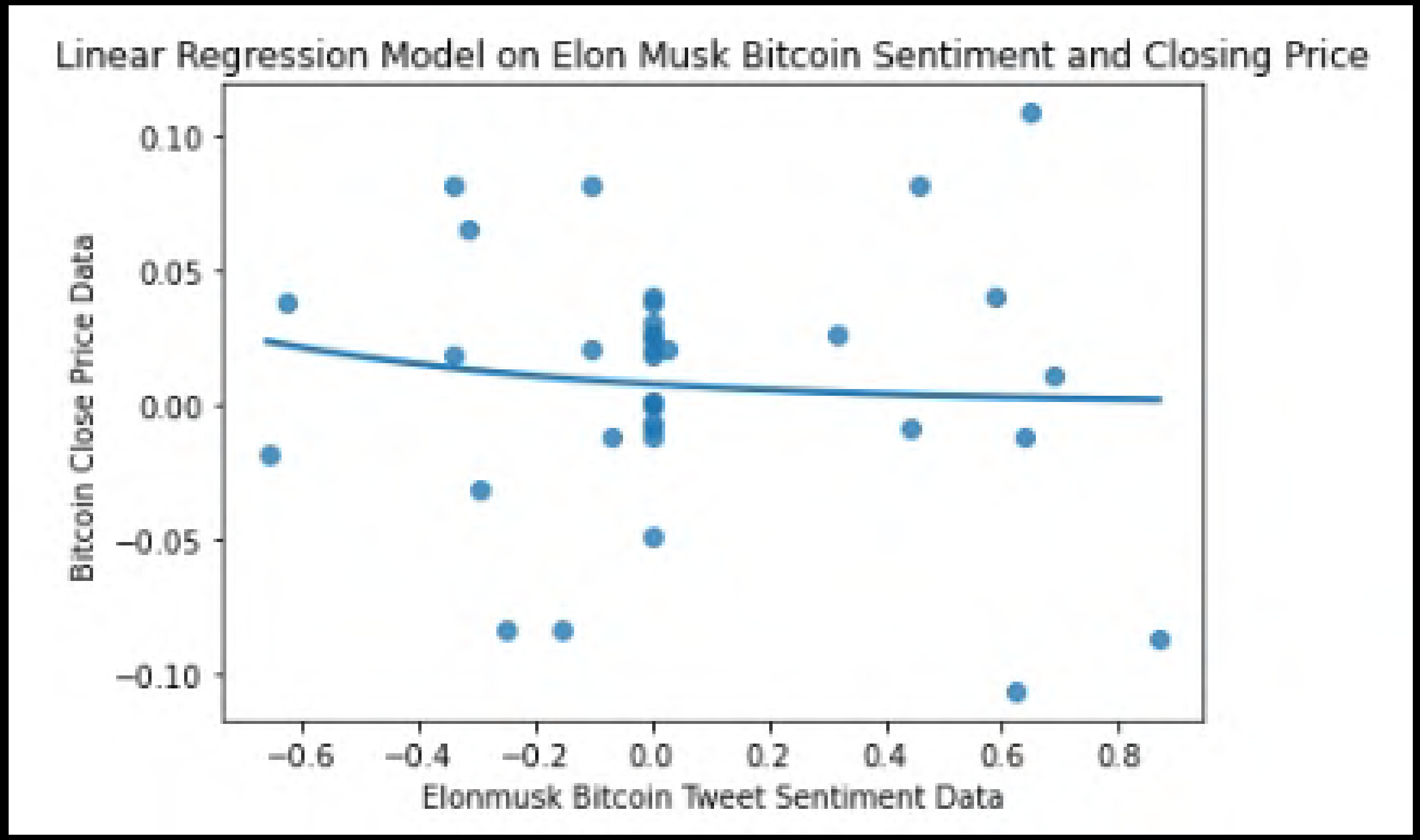
ELON MUSK DOGE TWEET SENTIMENT EFFECTS

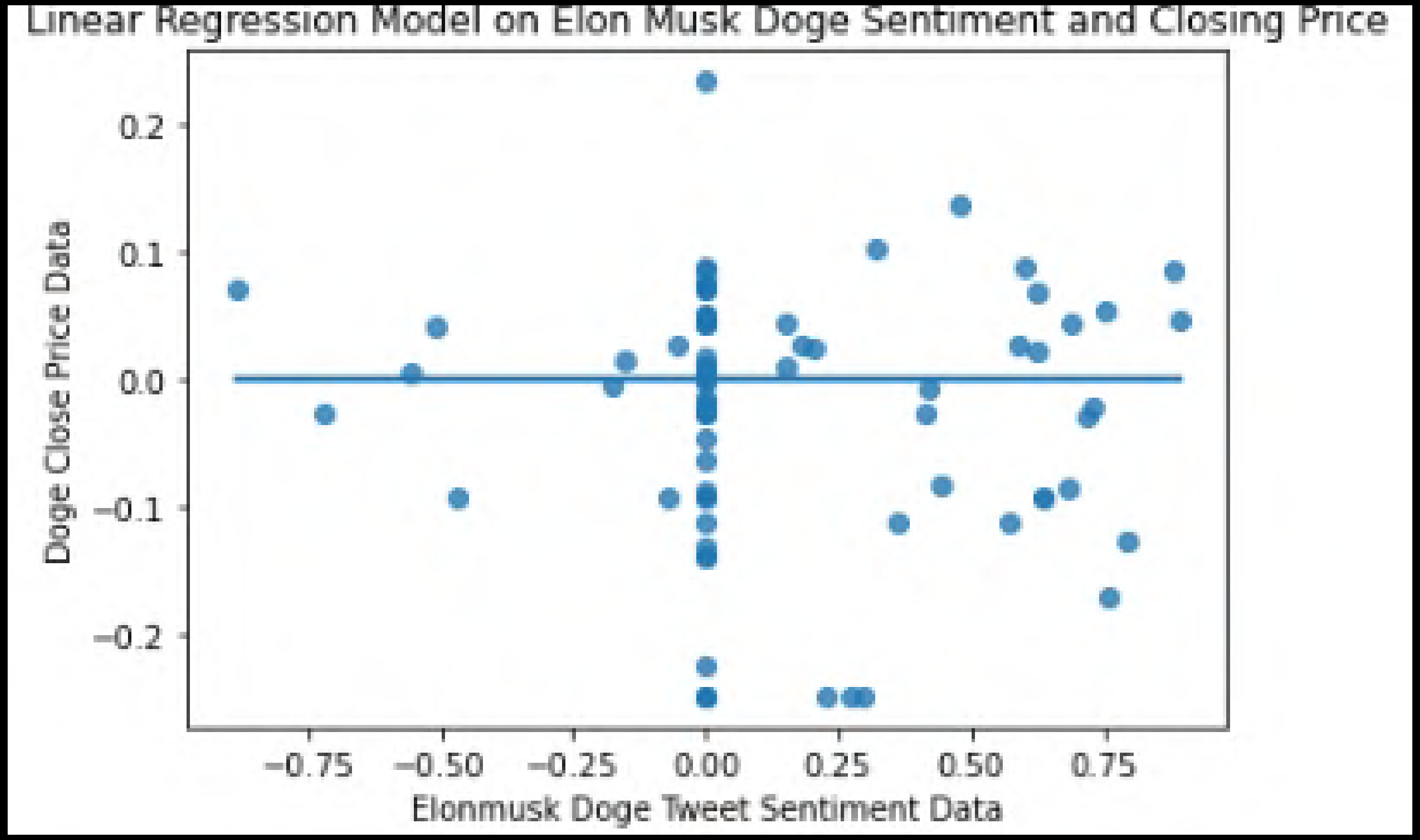




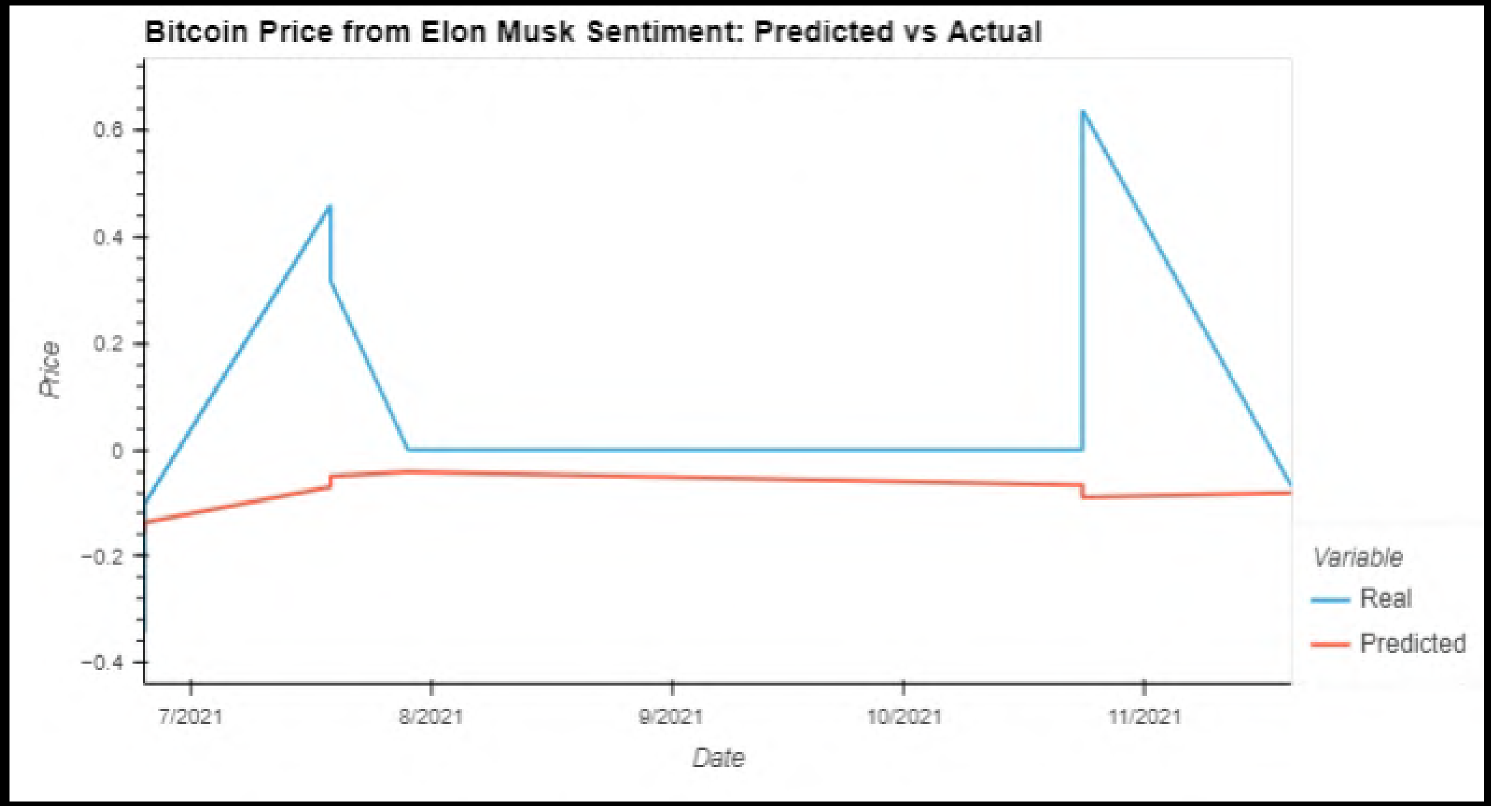
Elon Musk BTC Tweet Sentiment Compound Score

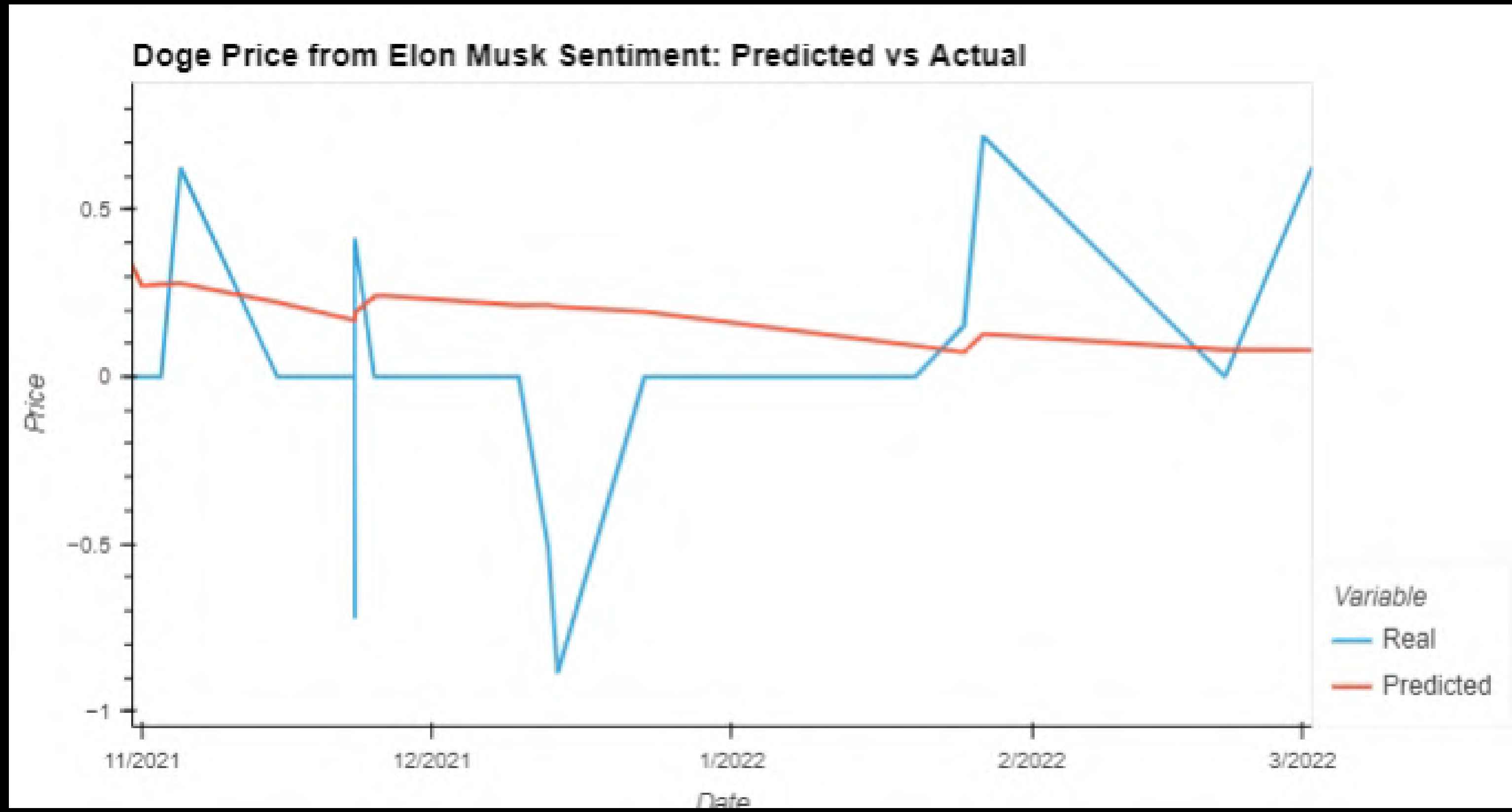
Bitcoin Close Price



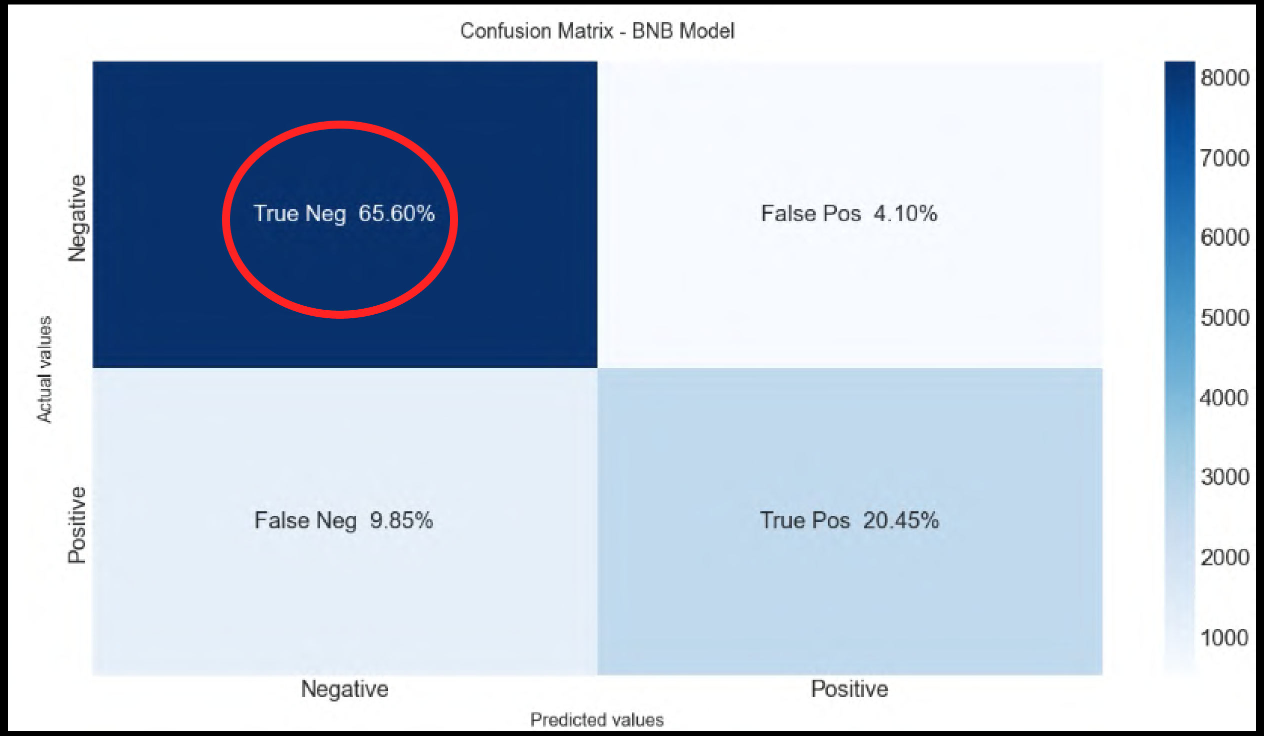


WHAT IS THE CORR?

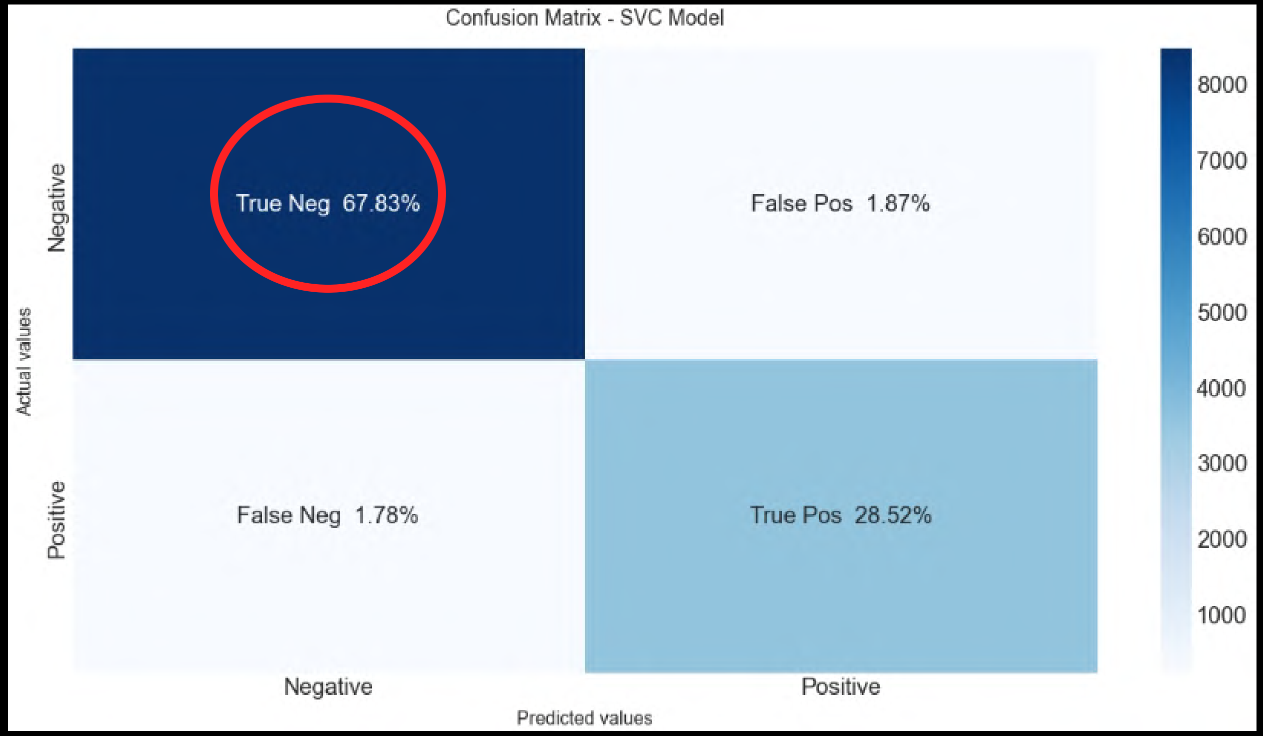




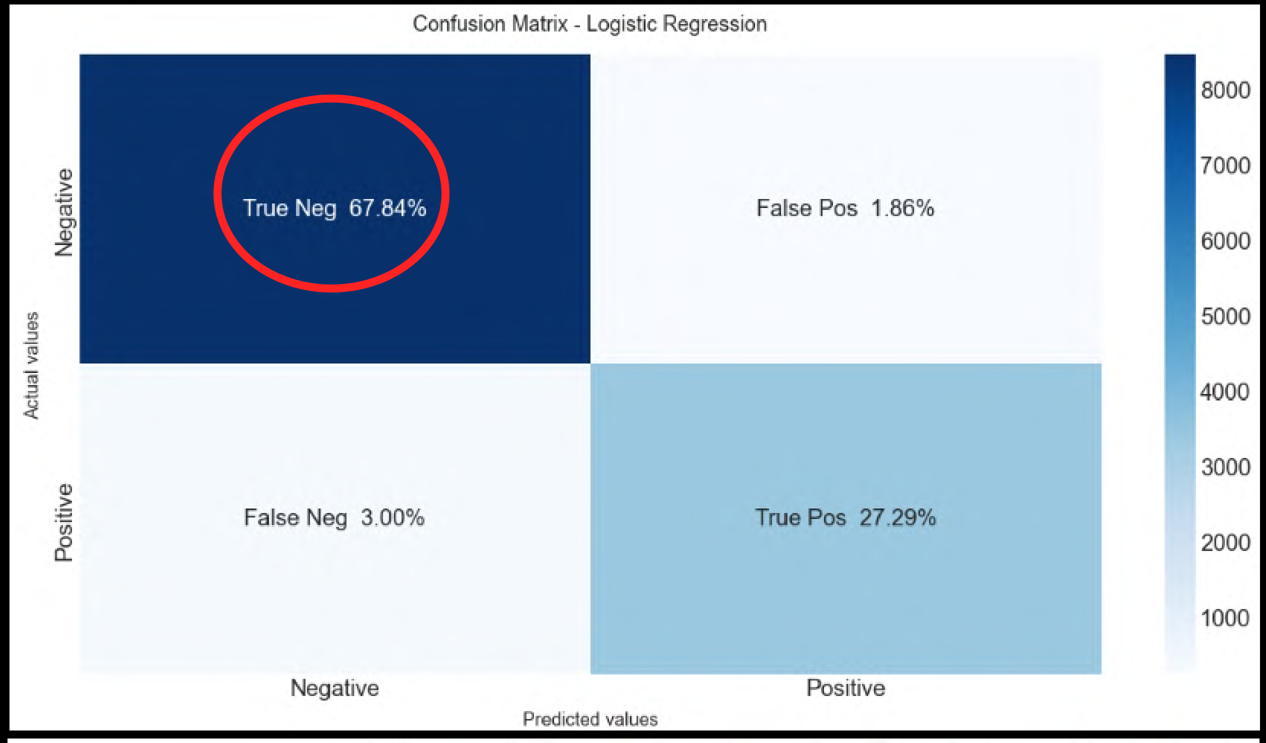
MODEL EVALUATION - DOGE TWITTER SENTIMENT



	precision	recall	f1-score	support
0	0.87	0.94	0.90	8714
1	0.83	0.67	0.75	3787
accuracy			0.86	12501
macro avg	0.85	0.81	0.82	12501
weighted avg	0.86	0.86	0.86	12501

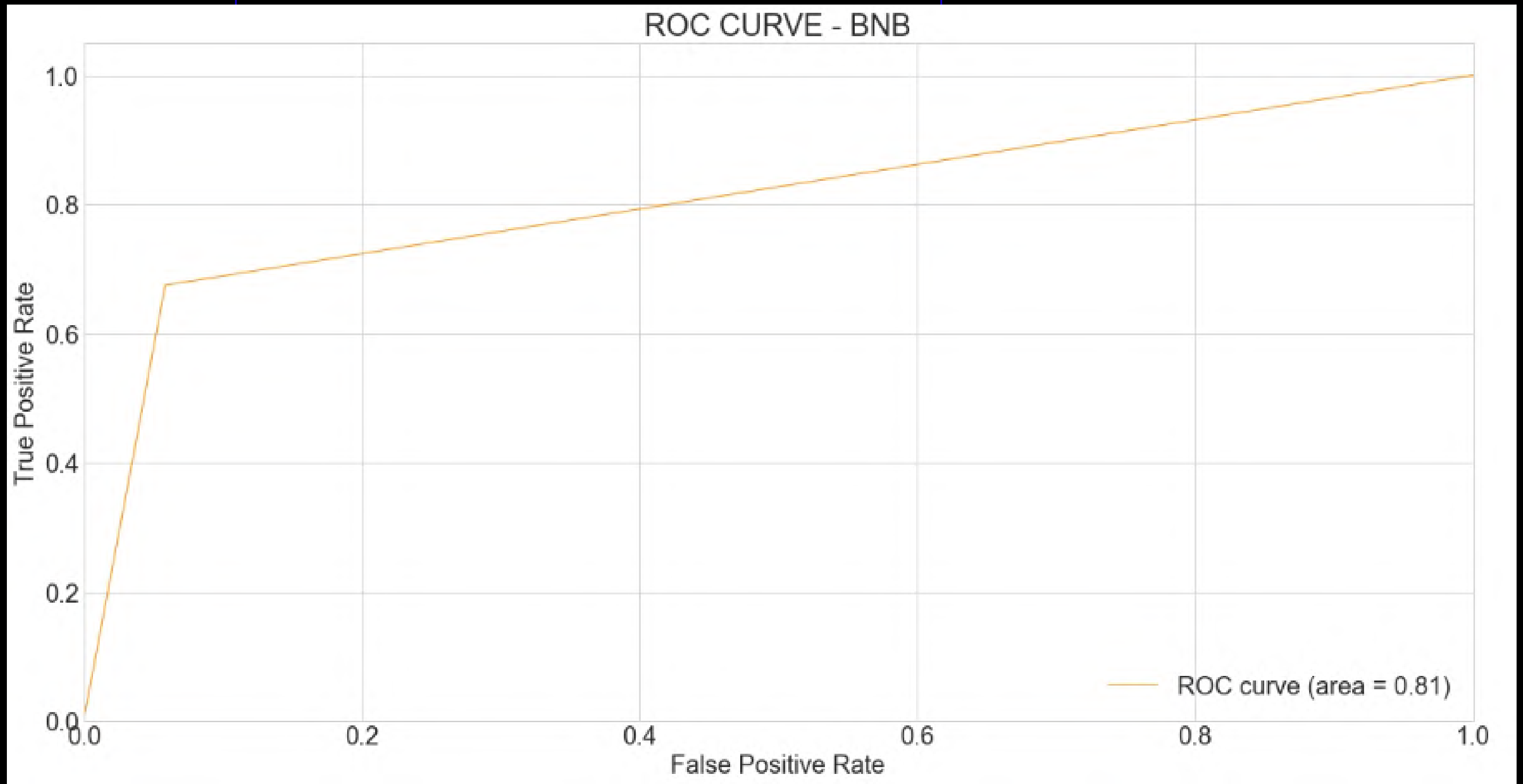


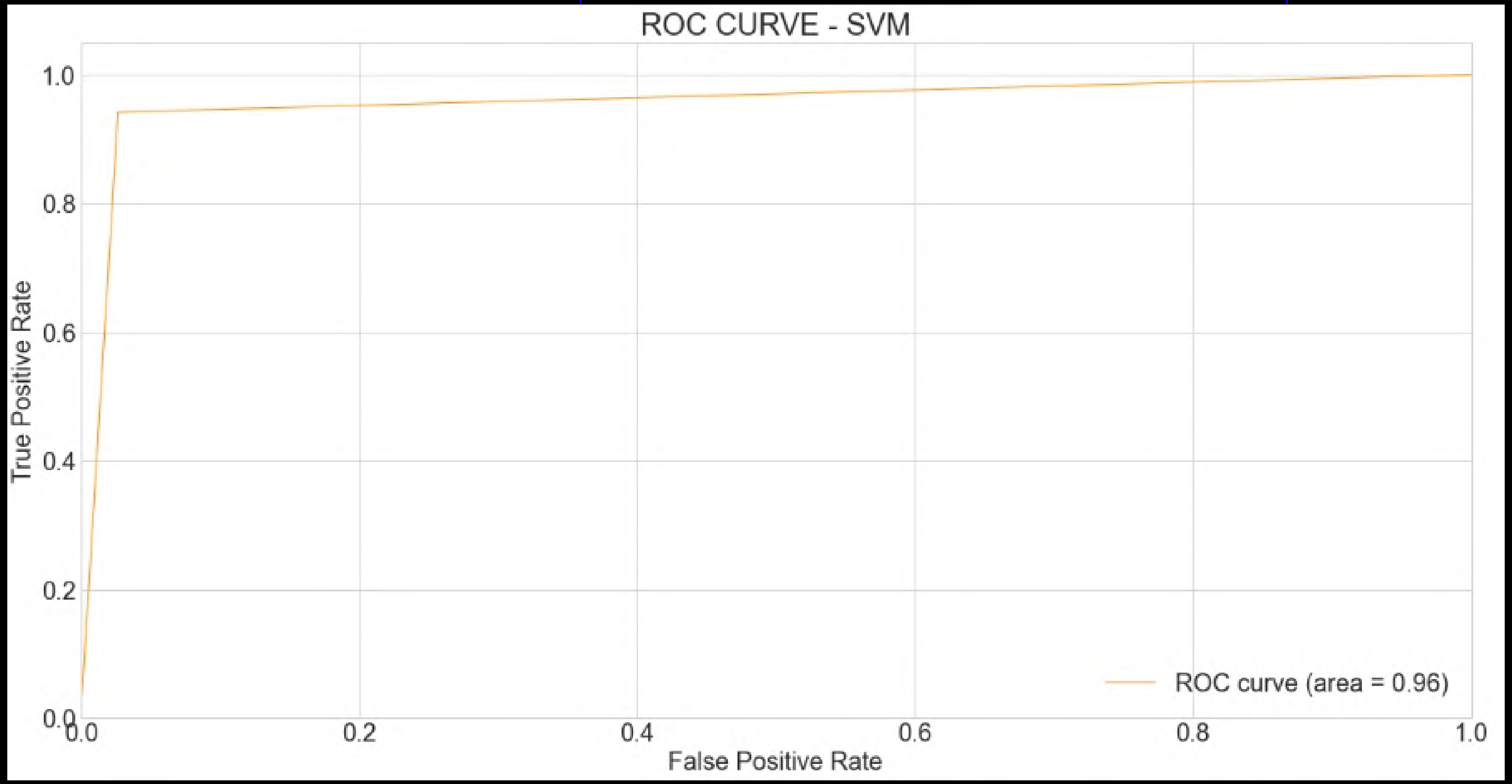
	precision	recall	f1-score	support
0	0.97	0.97	0.97	8714
1	0.94	0.94	0.94	3787
accuracy			0.96	12501
macro avg	0.96	0.96	0.96	12501
weighted avg	0.96	0.96	0.96	12501

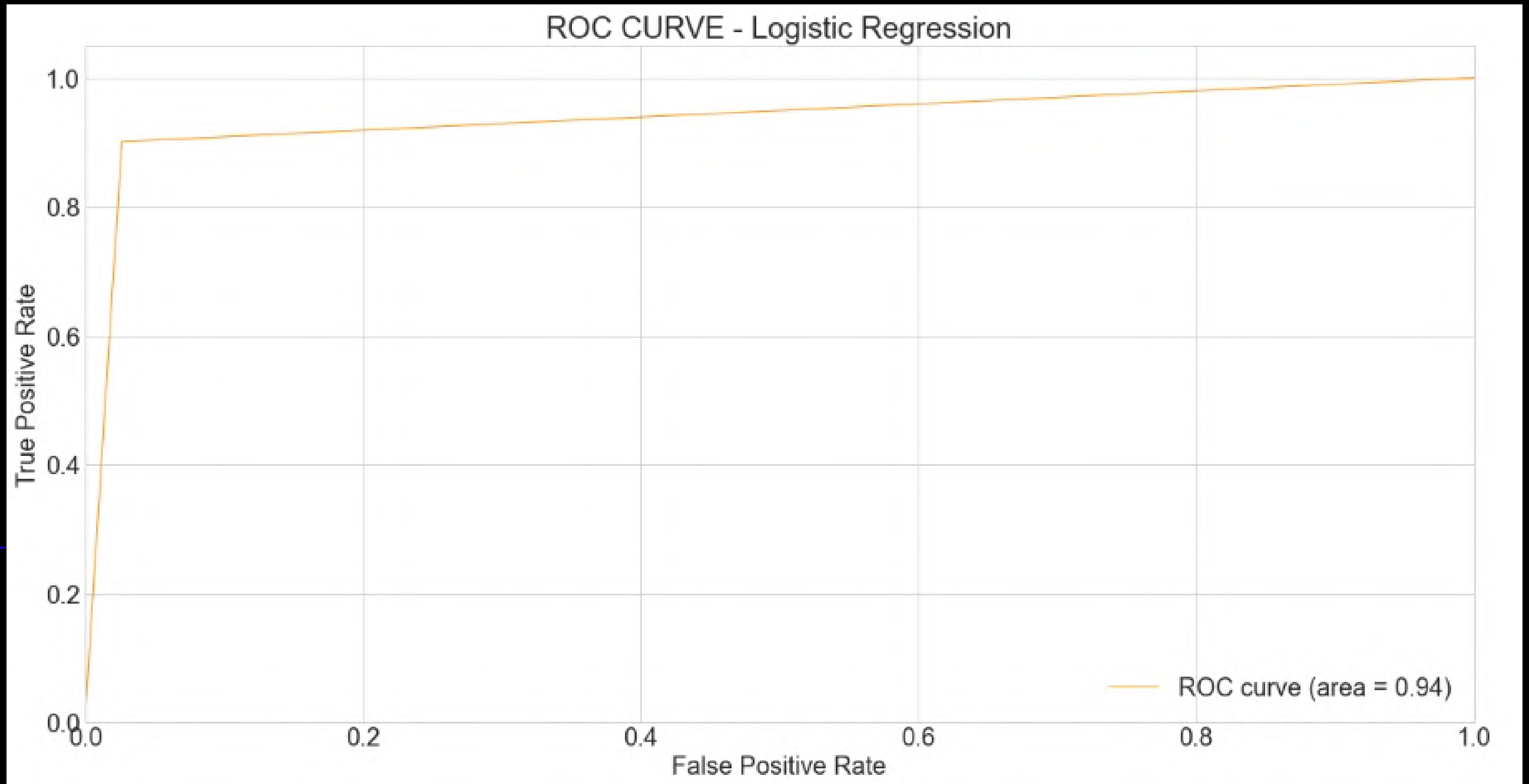


	precision	recall	f1-score	support
0	0.96	0.97	0.97	8714
1	0.94	0.90	0.92	3787
accuracy			0.95	12501
macro avg	0.95	0.94	0.94	12501
weighted avg	0.95	0.95	0.95	12501

DOGE MODEL EVALUATION







IN CONCLUSION

MACHINE LEARNING MODEL ON SENTIMENT ANALYSIS

EASY

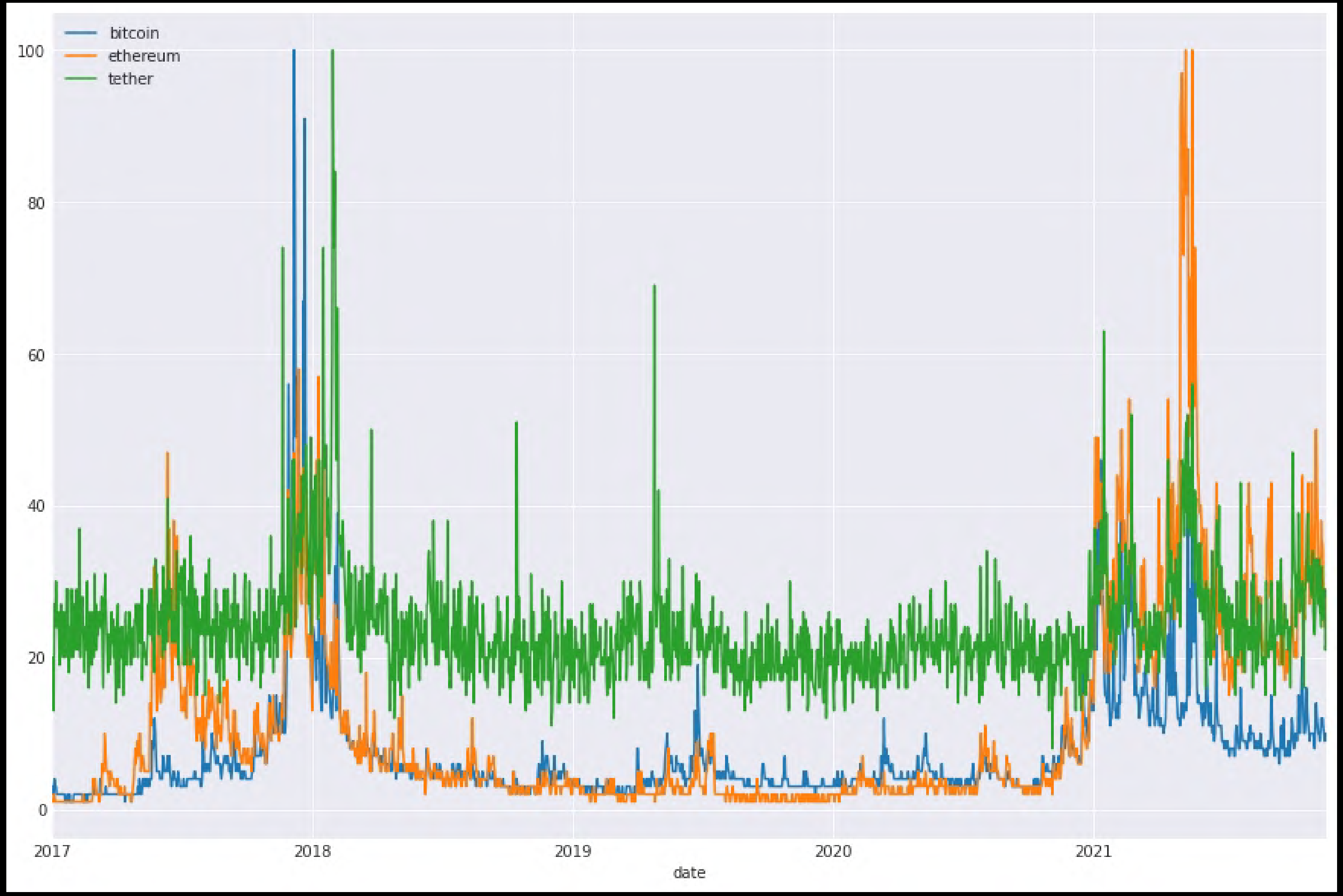
PREDICTING CLOSE PRICES WITH MACHINE LEARNING

HARD

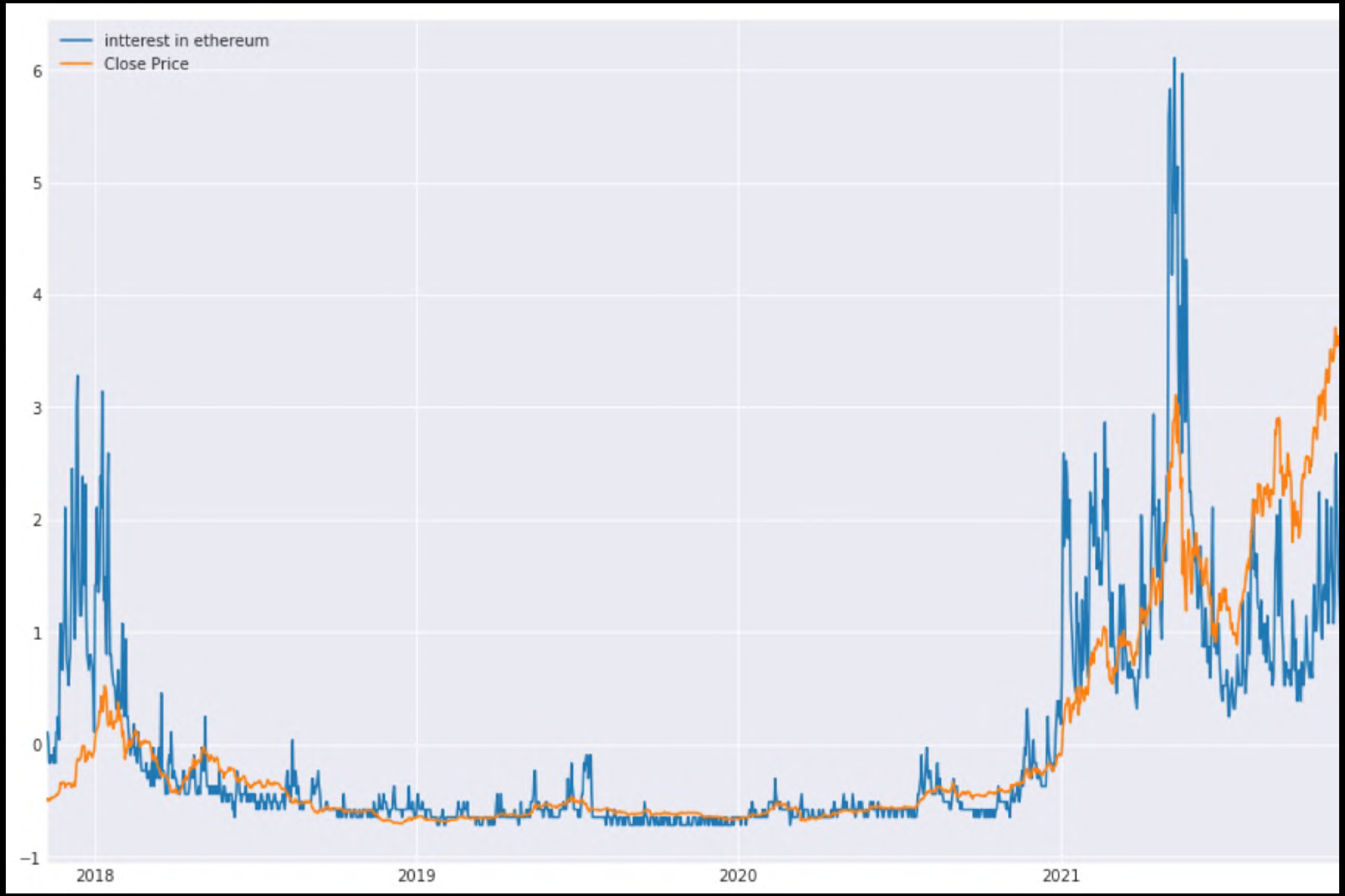
GOOGLE TRENDS ANALYSIS

DATA CLEAN UP

- THE DATA GOOGLE PROVIDES IS A SEARCH VOLUME INDEX (SVI). THE SEARCH VOLUME INDEX IS CALCULATED BY DIVIDING EACH DATA POINT BY THE TOTAL SEARCHES WITHIN A GEOGRAPHIC REGION AND TIME RANGE.
- THE NUMBERS ARE THEN SCALED BETWEEN 0 AND 100 ON A SEARCH TERM'S PROPORTION TO ALL SEARCHES ON ALL TOPICS.
- WHEN TRENDS DATA IS QUERIED FOR A PERIOD OF LONGER THAN 90 DAYS THE SVI RETURNED ARE AGGREGATED AT A WEEKLY LEVEL CAUSING A PROBLEM SO SPECIFIC CODE WAS USED.



MODEL EVALUATION



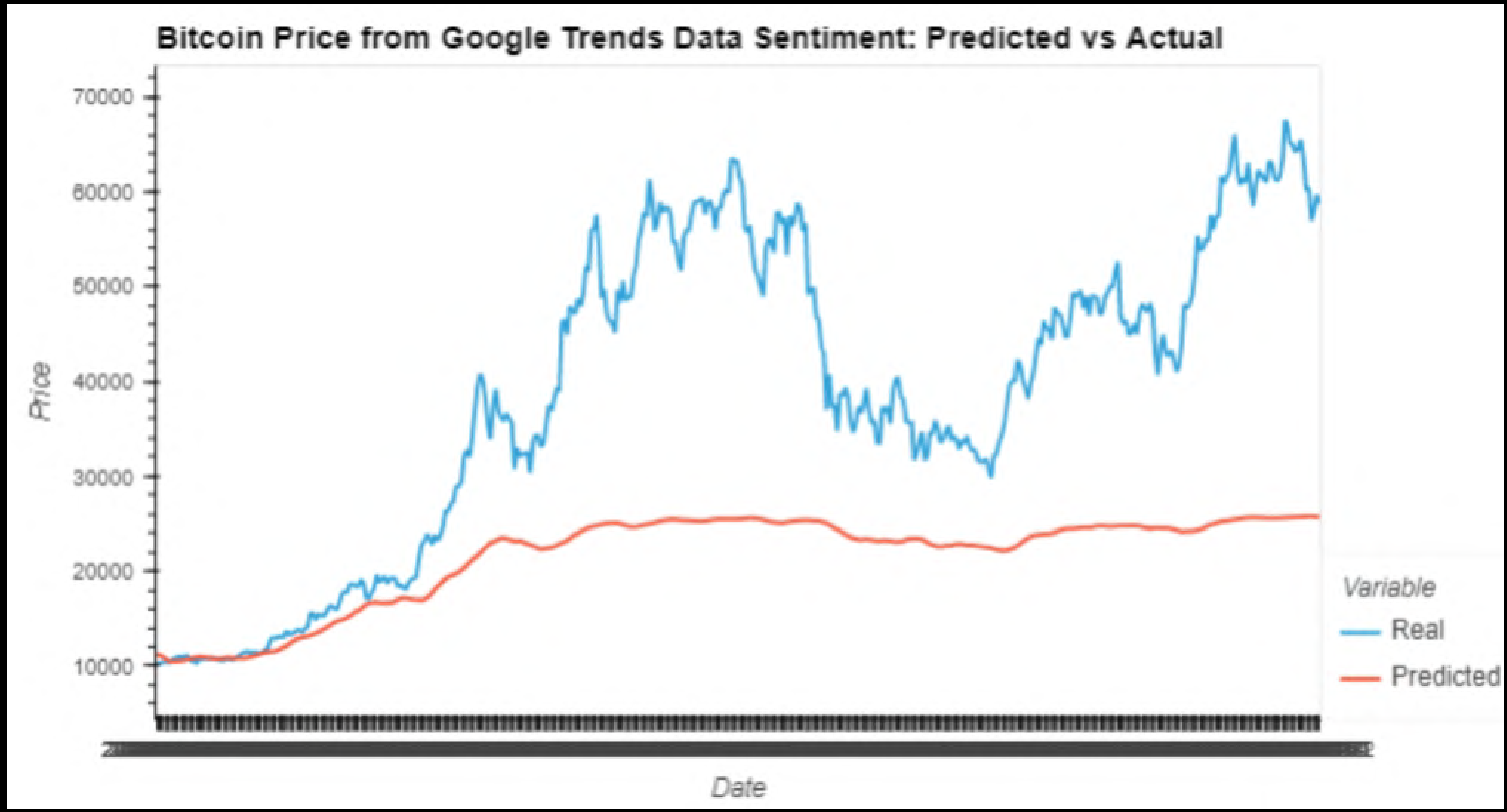
OLS Regression Results

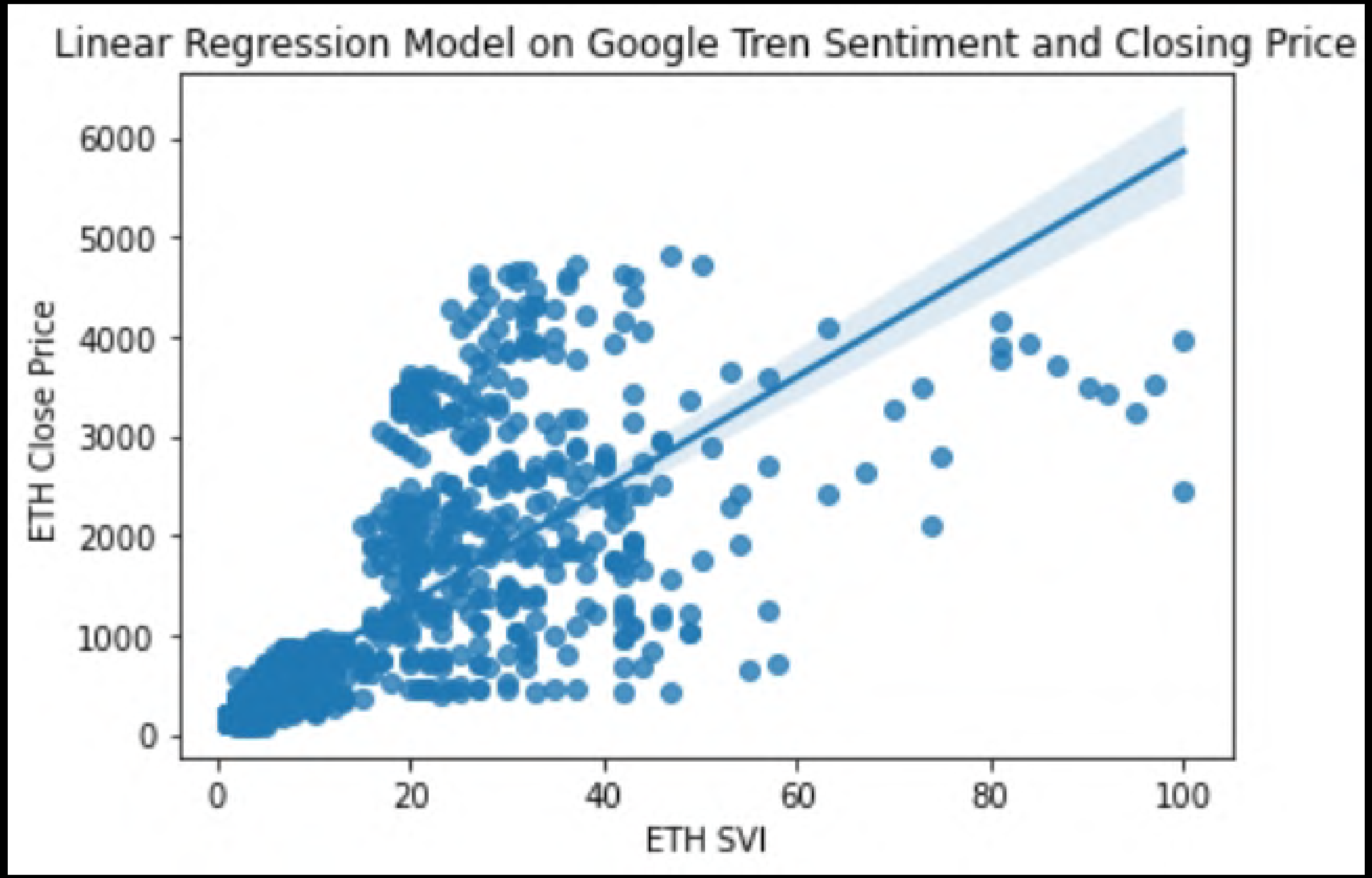
Dep. Variable:	ethereum_Close	**R-squared:**	0.598
Model:	OLS	**Adj. R-squared:**	0.597
Method:	Least Squares	**F-statistic:**	2188.
Date:	Sat, 16 Apr 2022	**Prob (F-statistic):**	1.65e-293
Time:	13:54:47	**Log-Likelihood:**	-11712.
No. Observations:	1475	**AIC:**	2.343e+04
Df Residuals:	1473	**BIC:**	2.344e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	189.6790	22.503	8.429	0.000	145.537	233.821
ethereum	56.7233	1.213	46.776	0.000	54.345	59.102

Omnibus:	390.533	**Durbin-Watson:**	0.131
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1924.692
Skew:	1.155	**Prob(JB):**	0.00
Kurtosis:	8.097	**Cond. No.**	23.6

PREDICTIONS





Pitch

DOW JONES ANALYSIS

The Dow Jones Industrial Average is a price-weighted measurement stock market index of 30 prominent companies listed on stock exchanges in the USA.

The initial goals of the analysis include:

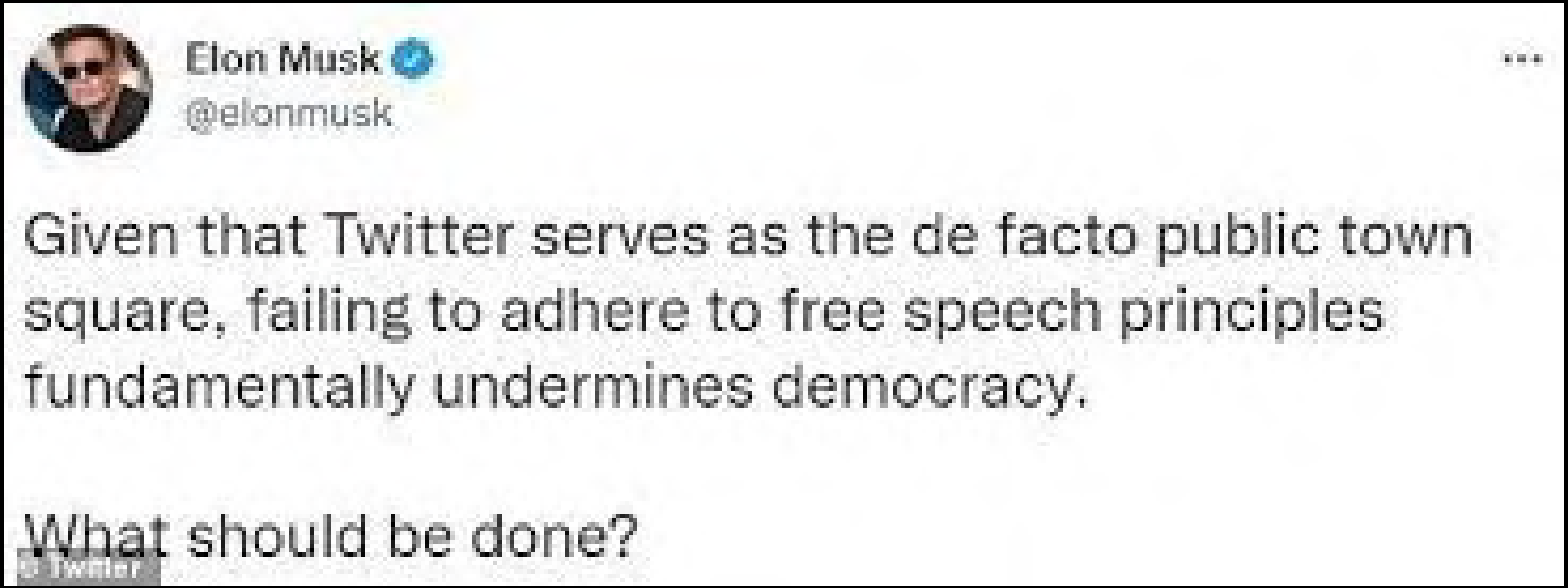
- To assess/predict the future price and the cumulative returns of DIA stock.
- Whether there was a correlation between the social media activity of current (President Joe Biden) and former world (President Donald J. Trump) leaders an effect DIA price.
- Comparative analysis of returns under both world leaders.
- Combine results with Twitter and Google Trends analysis' to create machine learning model.
- Use model to create algorithmic trading bot.

The goals achieved:

- Future price and the cumulative returns of DIA stock.
- Comparative analysis of returns under both world leaders.

Most importantly, we gained a better understanding of the time and activity required to complete the initial goals.

RELEVANCE



Elon Musk
@elonmusk
Given that Twitter serves as the de facto public town square, failing to adhere to free speech principles fundamentally undermines democracy.
What should be done?
© Twitter

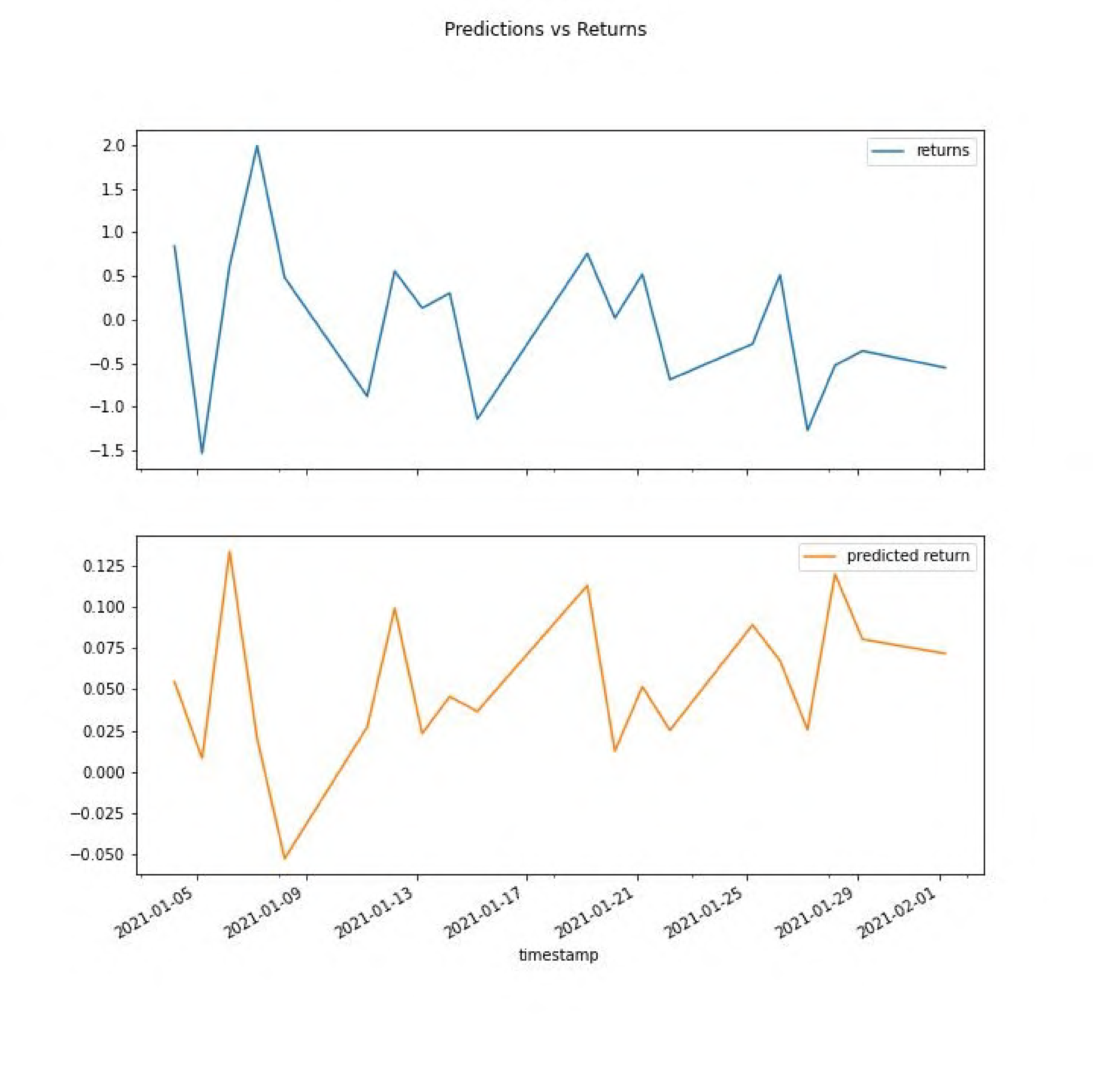
DATA CLEAN UP

Alpaca Trade API was used to retrieve stock data for the model. Data exploration and cleaning via the API was complicated by Alpacas upgrade to Trader v2 API. In my opinion, the API documentation was prepared for users with an intermediate to an advanced level of computer science. As a beginner, I overcame this initial obstacle by referring to our previous training on APIs together with Alpacas Postman collection. Alpacas postman collection is available at the following URL - https://www.postman.com/alpacamarkets/workspace/alpaca-public-workspace/documentation/17430392-af6823e1-0e8c-4387-9cef-dffbfbae836f

MODEL EVALUATION

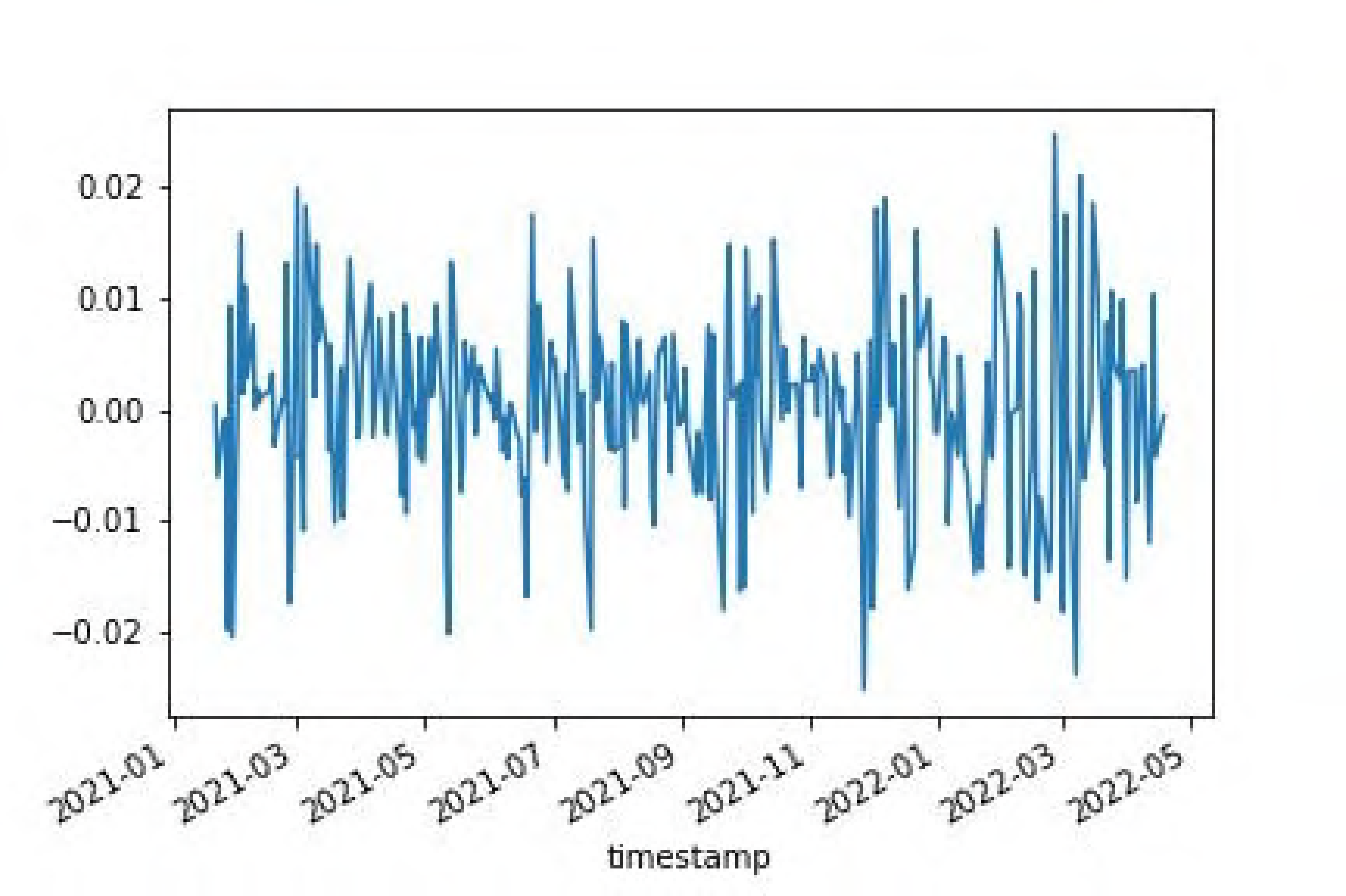
OVERALL MODEL DEVELOPMENT WAS CHALLENGED BY THE LIMITED TIME ALLOCATED TO COMPLETE THIS PROJECT. IN FUTURE PROJECTS, THIS PROBLEM MAY BE OVERCOME BY ASSIGNING A DEDICATED ROLE TO EACH GROUP MEMBER (IN ADDITION TO THE PROJECT MANAGER) FOR DATA EXPLORATION, PRODUCT MANAGEMENT ETC.

PREDICTIONS



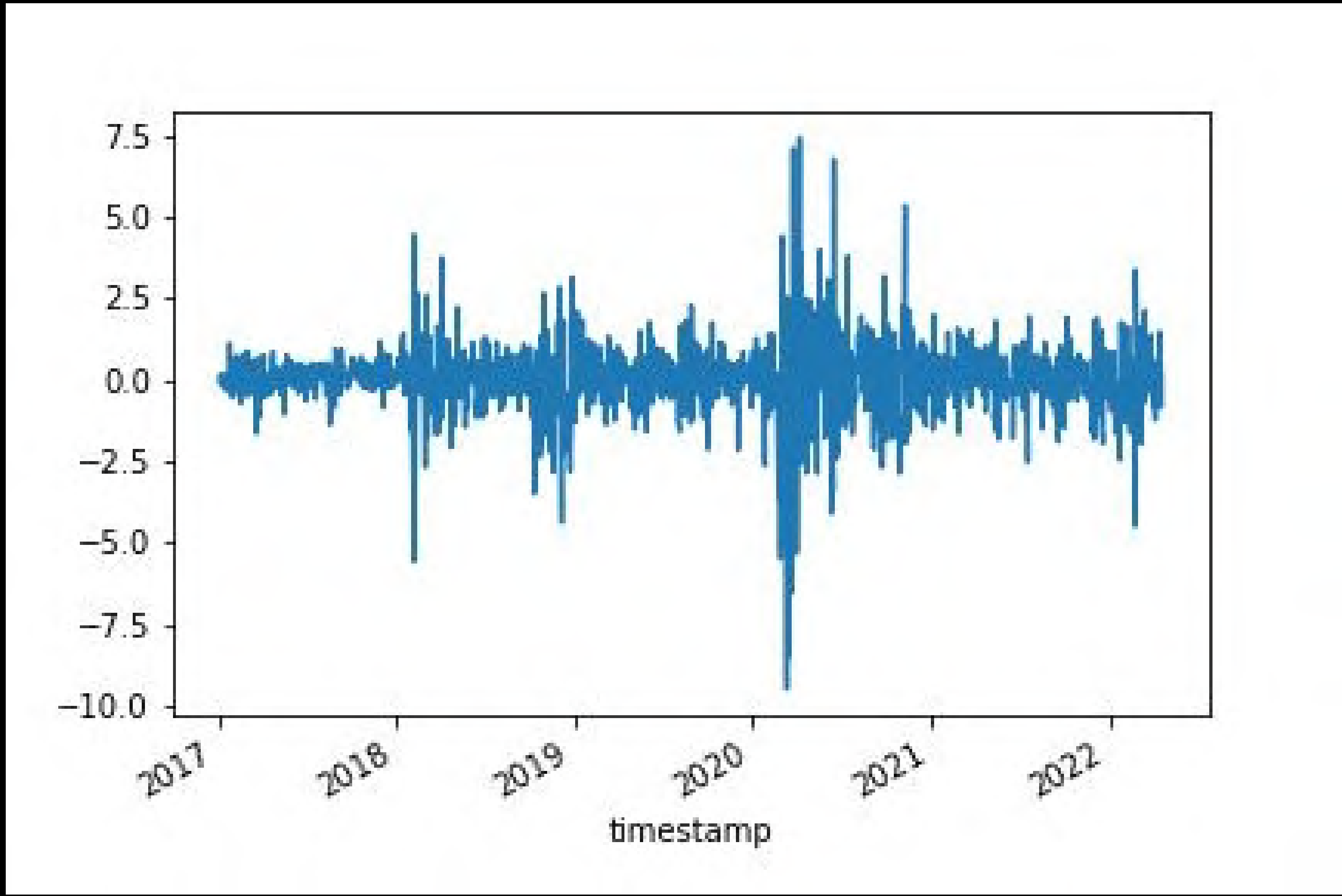
Pitch

JOE BIDEN



0.02
0.01
0.00
−0.01
−0.02
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
timestamp

DONALD J. TRUMP



OTHER

- LIVE DATA
- STOCKS
- CRYPTO
- SENTIMENT ANALYSIS


```
    'low': 41601.34,
    'open': 41623.87,
    'symbol': 'BTCUSD',
    'timestamp': 1650517260000000000,
    'trade_count': 221,
    'volume': 3.97723416,
    'vwap': 41613.0986994604})
ar({   'close': 3083.62,
    'exchange': 'CBSE',
    'high': 3085.0,
    'low': 3082.77,
    'open': 3085.0,
    'symbol': 'ETHUSD',
    'timestamp': 1650517260000000000,
    'trade_count': 308,
    'volume': 139.98522085,
    'vwap': 3083.9070820338})
ar({   'close': 41602.78,
    'exchange': 'CBSE',
    'high': 41614.65,
    'low': 41593.93,
```


Pitch

QUESTIONS

HTTPS://WWW.CSC2.NCSU.EDU/FACULTY/HEALEY/TWEET_VIZ/TWEET_APP/

CLICK FOR EXAMPLE APP

CLICK FOR EXAMPLE APP