

BIA 660 Group Project Team 2

Cryptocurrency Forecasting and Analysis

Group members: Siddhesh Shaji, Deepshika Reddy, Yogesh Awdhut Gadade

1. Introduction
2. Data Collection/Scraping
3. Data Insights
4. Data Processing
5. Models and Forecasting results
6. Conclusion
7. Future Scope and q&a

Introduction

Introduction

- ❑ About Cryptocurrency
- ❑ Trend in Cryptocurrency
- ❑ Objective: exploring cryptocurrency data sources
 - ❑ CoinmarketCap
 - ❑ LexisNexis
 - ❑ Twitter

Data Scraping

Data Scraping

- ❑ Scraping using **Selenium** and **BeautifulSoup**
 - ❑ CoinmarketCap.com
 - ❑ LexisNexis Database
- ❑ API based data scraping using snsrape
 - ❑ Twitter data
- ❑ We were able to gather around 5 years of data(structured and unstructured-text)

coinmarketcap interface

← → ↺ coinmarketcap.com/currencies/bitcoin/historical-data/ ☆ 2 ⚙️ S ⋮

 Bitcoin BTC

Price: \$56,898.19 +4.80%

Add to Main Watchlist ☆

Overview Market **Historical Data** Holders Project Info Wallets News Socials Ratings Analysis Price Estimates

↑ Share

Historical Data for Bitcoin

📅 Date Range ▼

Date	Open*	High	Low	Close**	Volume	Market Cap
Nov 28, 2021	\$54,813.02	\$57,393.84	\$53,576.74	\$57,248.46	\$28,116,886,357	\$1,081,180,416,262
Nov 27, 2021	\$53,736.43	\$55,329.26	\$53,668.36	\$54,815.08	\$30,560,857,714	\$1,035,176,249,679
Nov 26, 2021	\$58,960.29	\$59,183.48	\$53,569.76	\$53,569.76	\$41,810,748,221	\$1,011,613,083,342
Nov 25, 2021	\$57,165.42	\$59,367.97	\$57,146.69	\$57,274.68	\$34,284,016,248	\$1,081,500,188,674
Nov 24, 2021	\$57,565.85	\$57,803.07	\$55,964.22	\$56,280.43	\$36,635,566,789	\$1,062,661,698,465
Nov 23, 2021	\$56,304.56	\$57,875.52	\$55,632.76	\$57,569.07	\$37,485,803,899	\$1,086,985,075,773
Nov 22, 2021	\$58,706.85	\$59,266.36	\$55,679.84	\$56,289.29	\$35,036,121,783	\$1,062,772,703,759
Nov 21, 2021	\$59,730.51	\$60,004.43	\$58,618.93	\$58,730.48	\$26,123,447,605	\$1,108,812,311,533
Nov 20, 2021	\$58,115.08	\$59,859.88	\$57,469.73	\$59,697.20	\$30,624,264,863	\$1,127,014,430,118



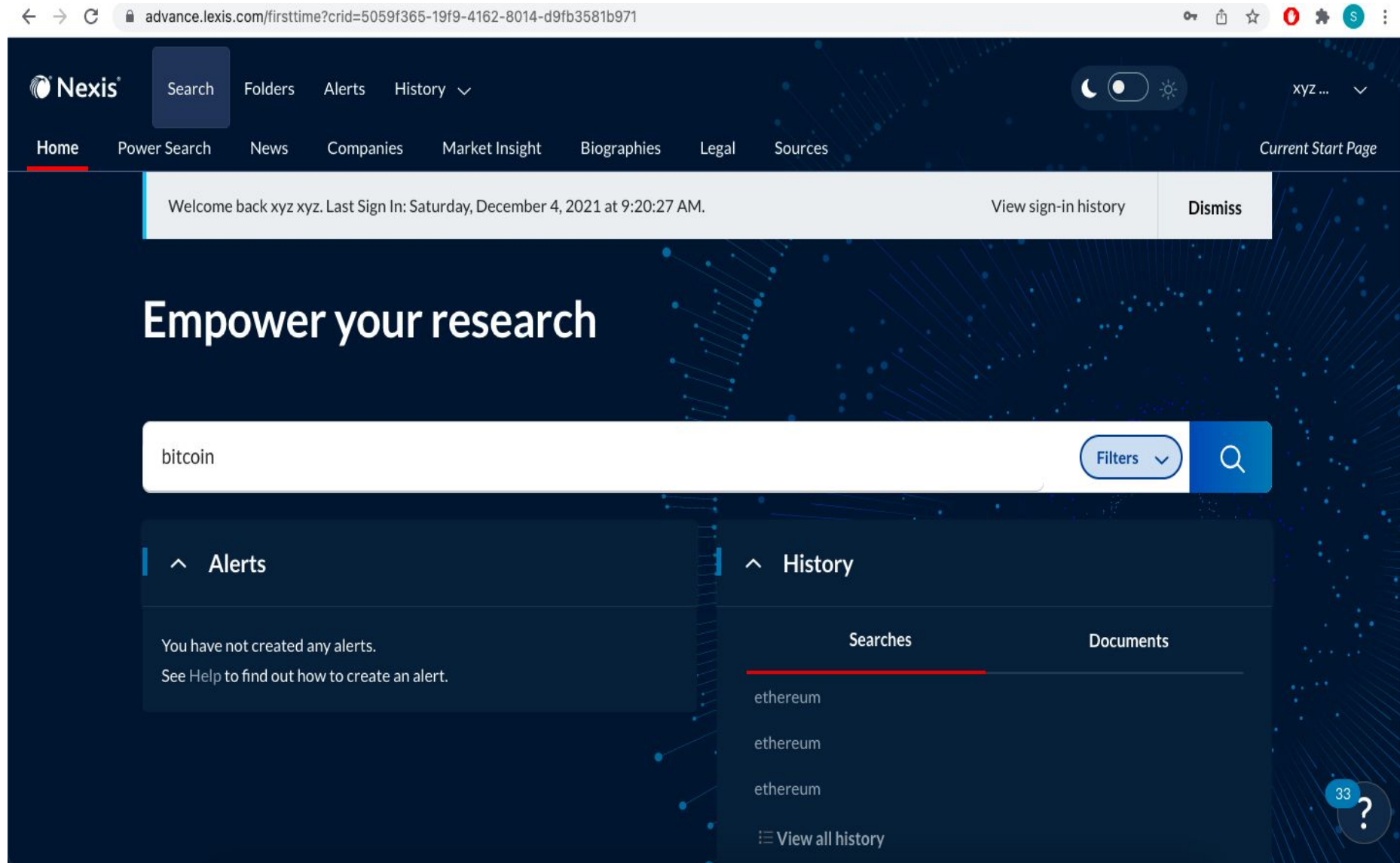
Add to Main Watchlist ☆

Load More

** Latest data in range (UTC time)



Socials



The screenshot shows the LexisNexis interface. At the top, there is a navigation bar with the LexisNexis logo, a search bar, and links for Folders, Alerts, History, and a dropdown menu. Below the navigation bar, there is a secondary navigation bar with links for Home, Power Search, News, Companies, Market Insight, Biographies, Legal, and Sources. A welcome message is displayed: "Welcome back xyz xyz. Last Sign In: Saturday, December 4, 2021 at 9:20:27 AM." with links for "View sign-in history" and "Dismiss". The main heading is "Empower your research". Below this is a search bar containing the text "bitcoin" and a "Filters" dropdown. A search button is located to the right of the search bar. Below the search bar, there are two panels: "Alerts" and "History". The "Alerts" panel shows a message: "You have not created any alerts. See Help to find out how to create an alert." The "History" panel shows a list of search results under the "Searches" tab, with the word "ethereum" listed three times. A "View all history" link is at the bottom of the history panel. A help icon with the number "33" is in the bottom right corner.

advance.lexis.com/firsttime?crid=5059f365-19f9-4162-8014-d9fb3581b971

Nexis® Search Folders Alerts History

Home Power Search News Companies Market Insight Biographies Legal Sources

Welcome back xyz xyz. Last Sign In: Saturday, December 4, 2021 at 9:20:27 AM. View sign-in history Dismiss

Empower your research

bitcoin Filters

Alerts

You have not created any alerts.
See Help to find out how to create an alert.

History

Searches Documents

ethereum
ethereum
ethereum

View all history

33 ?

LexisNexis interface

← → ↺

advance.lexis.com/search/?pdmfid=1519360&crd=80e55f8e-8b70-472a-b1d3-45f4268f12a6&pdsearchterms=bitcoin&pdstartin=hlct%3A1%3A1&pdcaseshlctsel...

Nexis

Search

Folders

Alerts

History

bitcoin

Filters

Q

xyz ...

Home > Results: (bitcoin)

Results for:bitcoin

Actions

Select Language

Disclaimer

News10,000+

Companies and FinancialGet

Biographies16

Cases0

Show more

Narrow By

Search Within Results

Include

Exclude

Enter search terms

Q

Timeline

Publication Location

Group Duplicates: Off

What's this?

Create Newsletter

Sort by: Relevance

1.

Discussing Bitcoin Businesses And Bitcoin Entrepreneurship

Dec 02, 2021

Newstex Blogs

Bitcoin Magazine

14640 words

189 hits

Preview

... mean, no wallets take an output, for example, when you're sending Bitcoin, which makes sense, but it does make it difficult to have a wallet as a Bitcoin business purely. You're not making money. And that's the case if we look at a lot of Bitcoin businesses is that they really don't have a way to monetize. ...

... build a wallet because you need your users to have it. Bitcoin ATMs might build a wallet. I know of some that have. But other than that, Bitcoin businesses, for the sake of business, it can be very difficult ...

... think you know very well, right? I think Exotica was a Bitcoin business. John Carvalho (03:09): Yeah. I think there are two different kinds of hard-to-earn money in Bitcoin. There's the kind that, like with Exotica, just hard to ...

... that actually works or even makes sense for a lot of Bitcoin projects as well. And I guess that's maybe a better term for them is a Bitcoin project rather than a Bitcoin business. ...

... Yeah. I think it's important to talk about the fact that Bitcoin literally is... If it is money, then all business is Bitcoin business. And I think ...

33

?

10. ☐

Preview

FINANZAS; Pág. 24 | 516 words | 8 hits

< 1 2 3 ... 454068 >

Twitter(using snsrape)

← Tweet



Documenting Bitcoin
@DocumentingBTC

US Bank to launch #Bitcoin ₿ custody services.

That's the 5th largest retail bank in the United States.



cnbc.com

US Bank launches bitcoin custody service as institutions race to...
The move is the latest sign that established financial players are beginning to accept cryptocurrencies as a legitimate asset class.

← Tweet



Documenting Bitcoin
@DocumentingBTC

Hungary 🇭🇺 has become home of the first Satoshi Nakamoto, creator of #Bitcoin ₿, statue in the world.

Thanks to community funding, it's a life-size bronze bust with a face of a mirror, meaning everyone sees their own face when standing in front of it.

“We are all Satoshi”

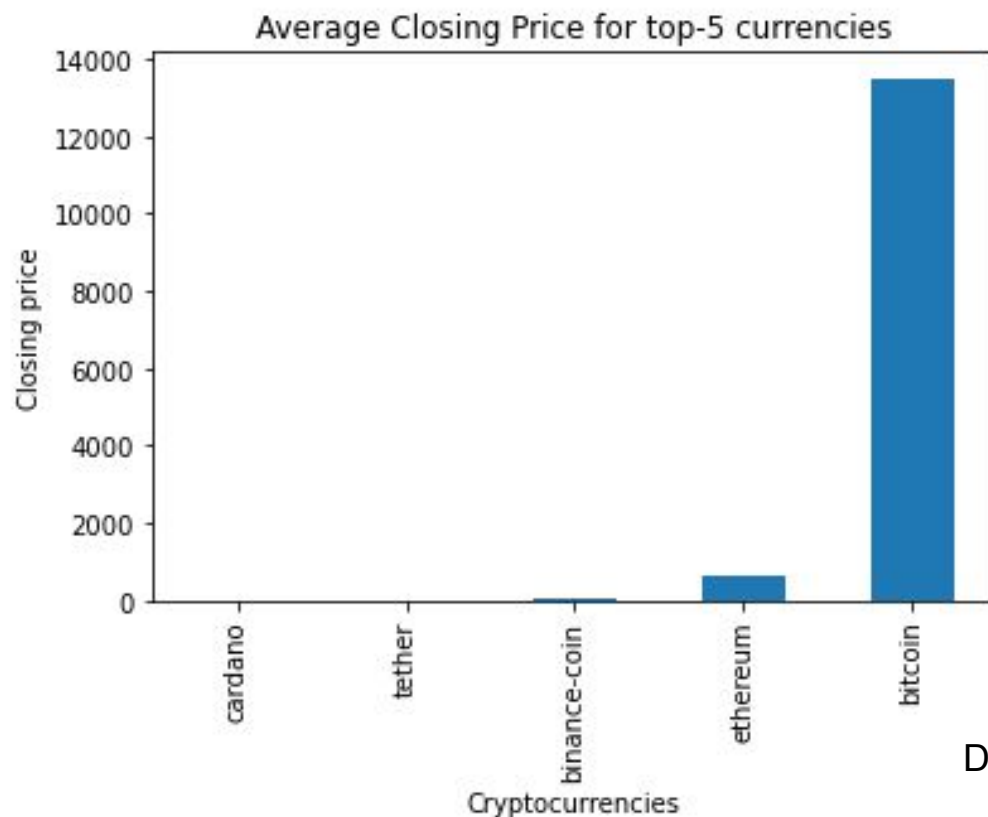


4:03 PM · Sep 16, 2021 · Twitter Web App

Data Insights

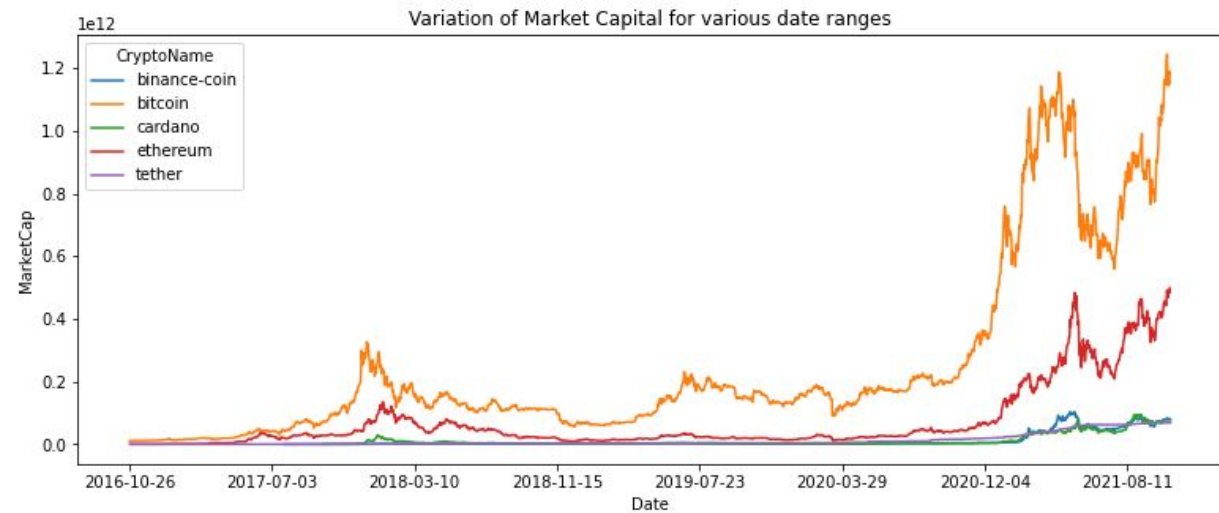
Data Insights(CoinMarketCap)

```
Out[25]: CryptoName
bitcoin      13499.099984
ethereum     640.013627
binance-coin 77.117319
tether       1.000770
cardano      0.388136
Name: Close, dtype: float64
```



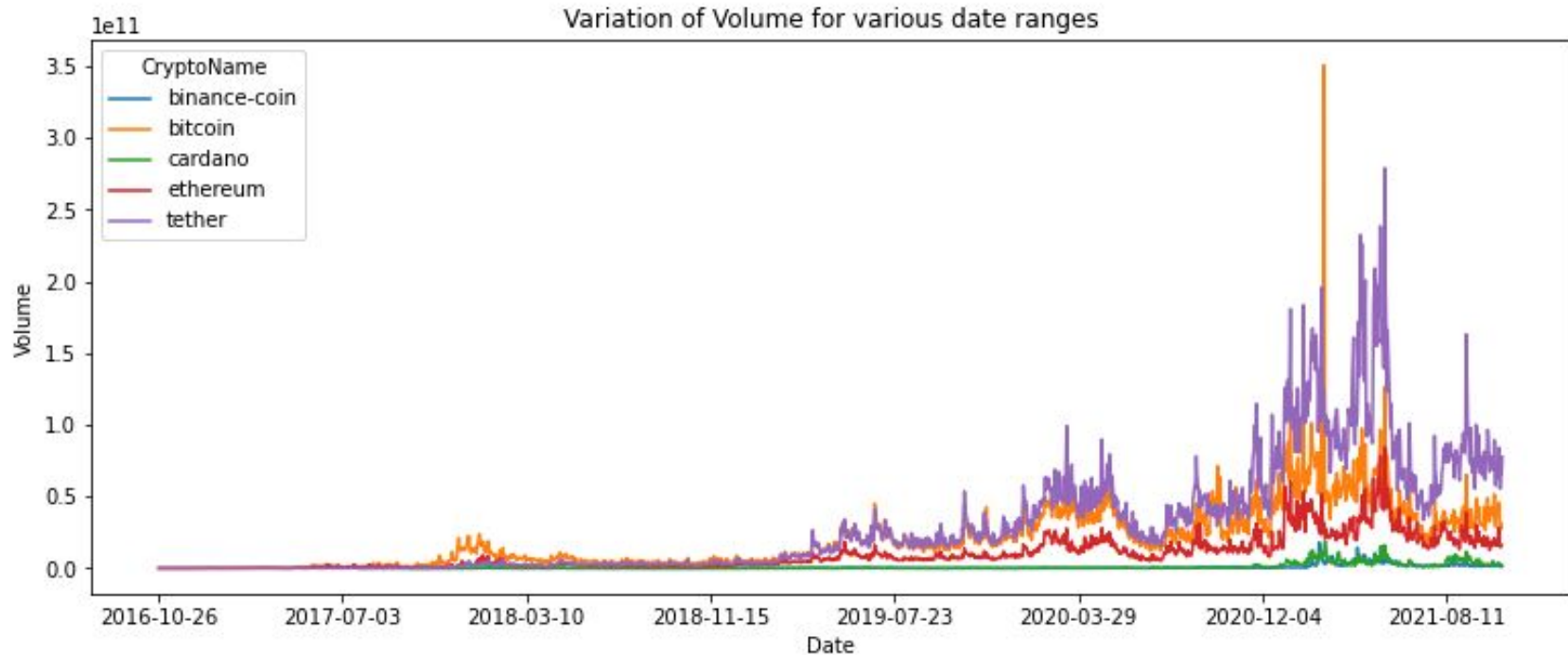
Data source: coinmarketcap.com

Observed Cryptocurrencies trend



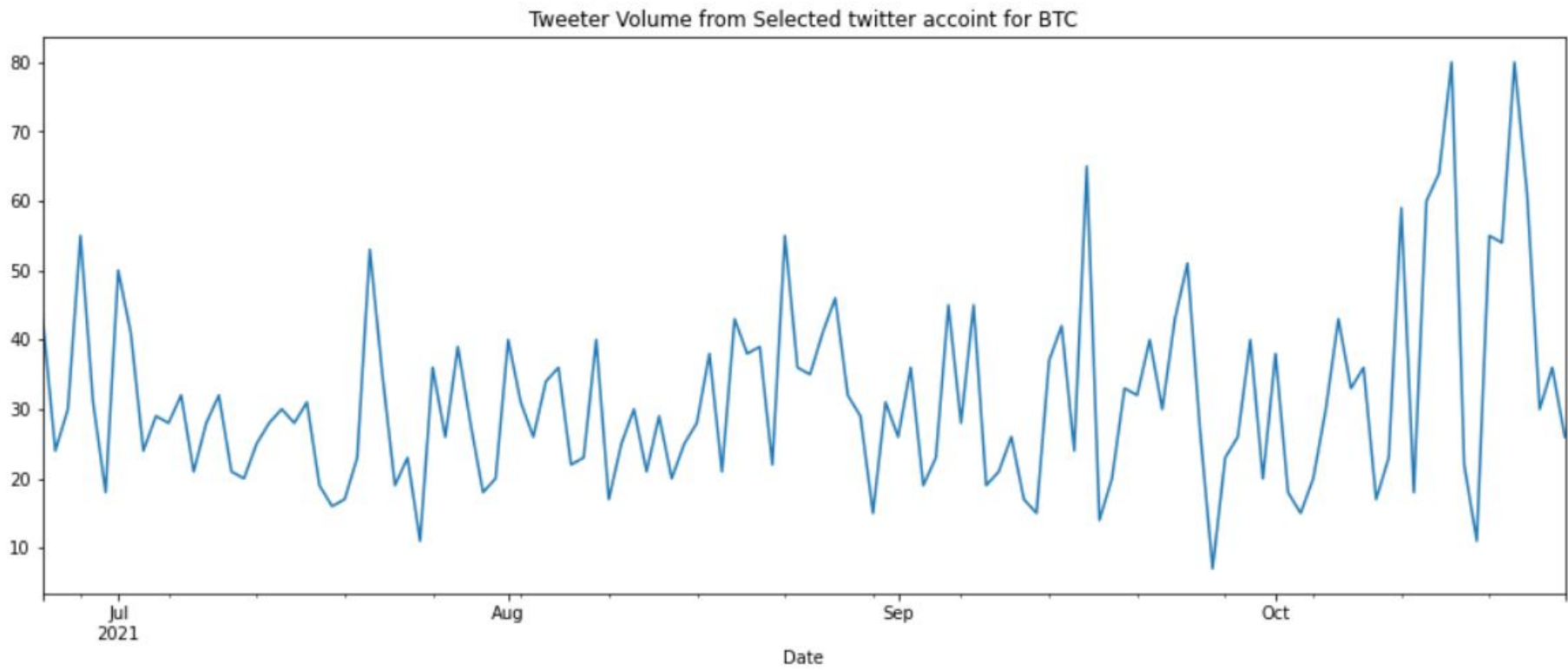
Data source: coinmarketcap.com

Data Insights(CoinMarketCap)



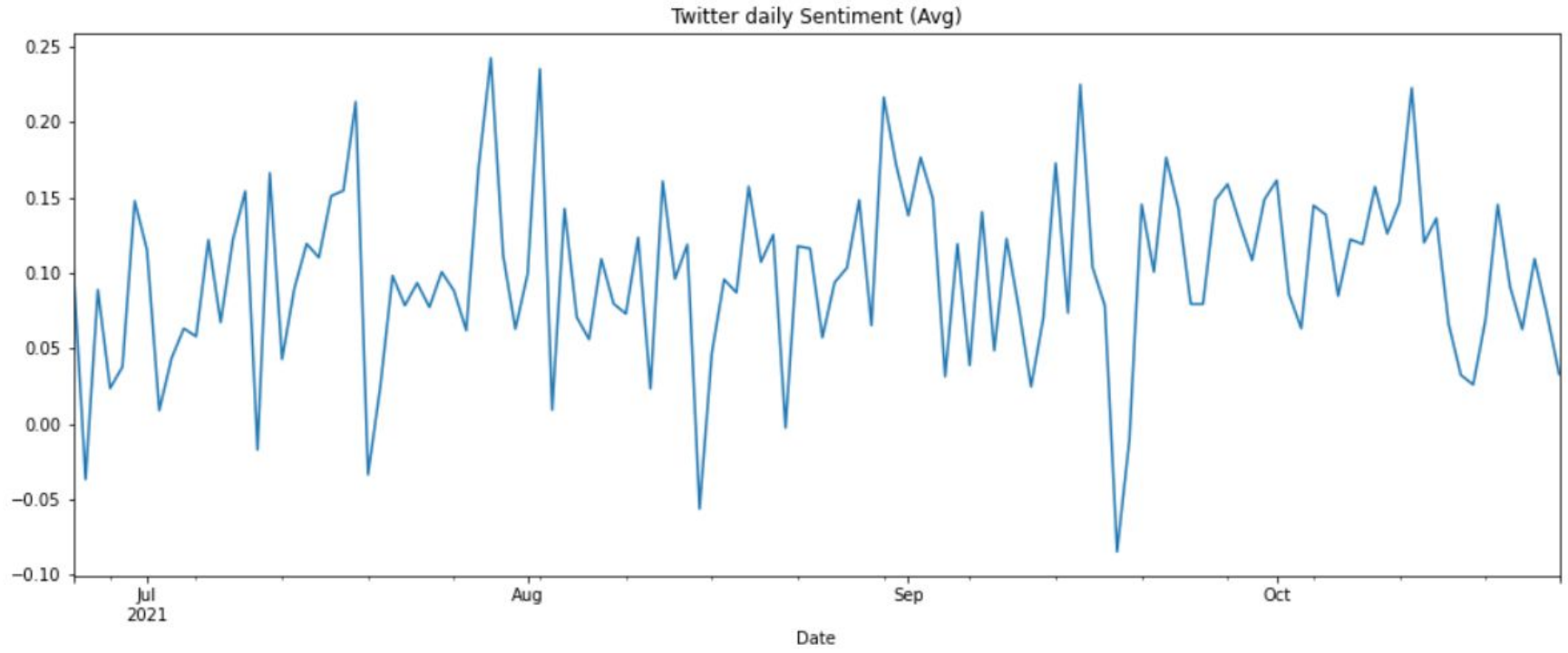
Data source: coinmarketcap.com

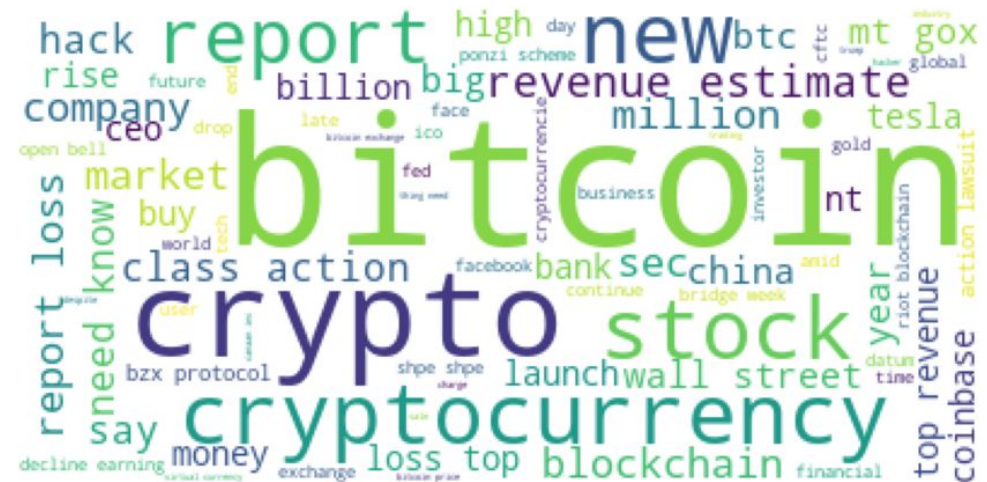
Volume of Tweets



Data source: Twitter

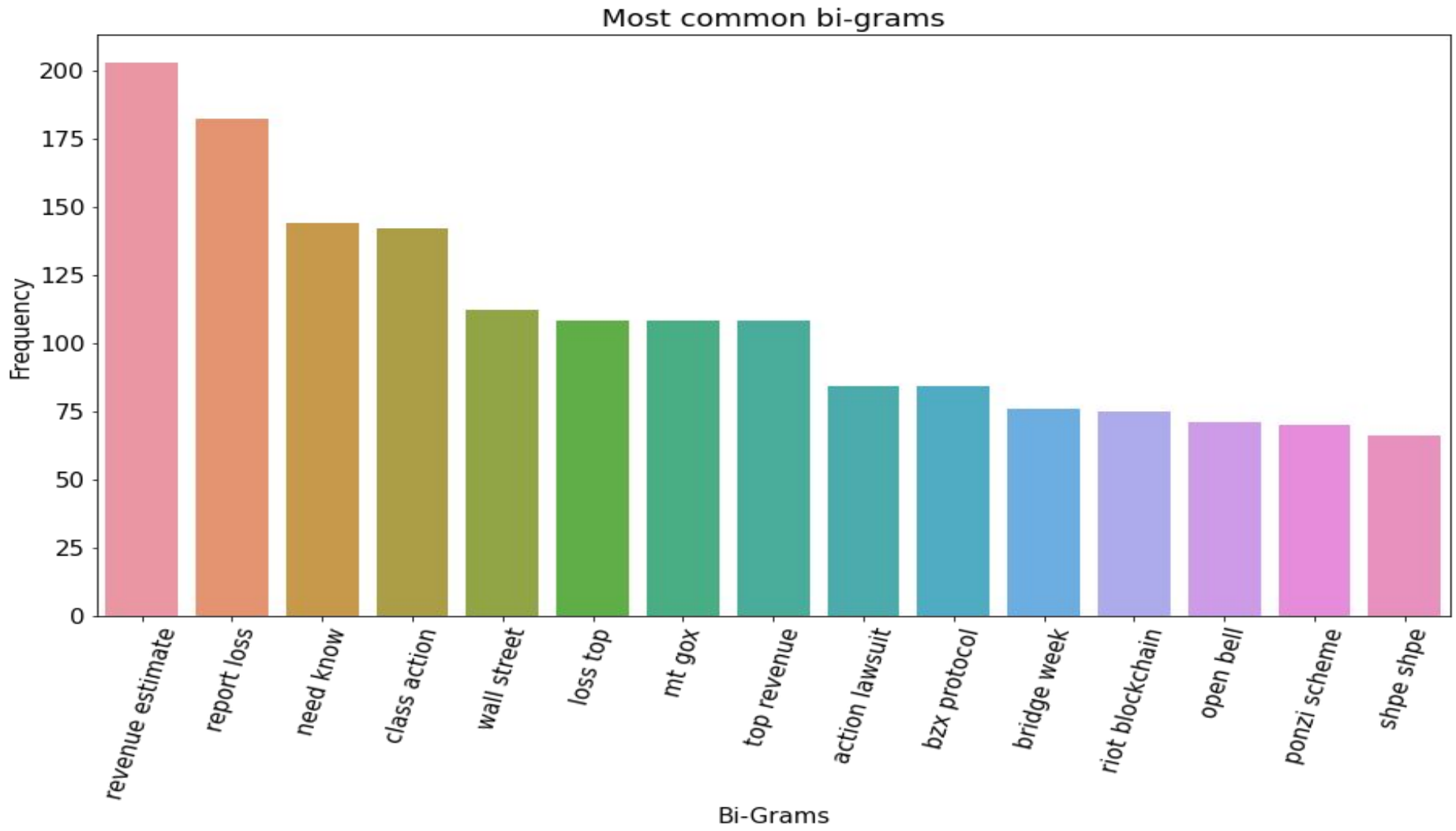
Observed Sentiment Trend





Data source: LexisNexis

Most observed bigrams



Data source: LexisNexis

Data Processing

Structured, Unstructured

Structured ones

❑ On **Structured** data (from coinmarketcap)

	Date	Open*	High	Low	closeValue	Volume	Market Cap
0	2021-10-24	61,368.34	61,505.80	59,643.35	60,930.84	27,316,183,882	1,148,743,134,468
1	2021-10-23	60,694.63	61,743.88	59,826.52	61,393.62	26,882,546,034	1,157,410,091,263
2	2021-10-22	62,237.89	63,715.02	60,122.80	60,692.26	38,434,082,775	1,144,131,483,274
3	2021-10-21	66,002.23	66,600.55	62,117.41	62,210.17	45,908,121,370	1,172,684,282,558
4	2021-10-20	64,284.59	66,930.39	63,610.67	65,992.84	40,788,955,582	1,243,927,428,207

1	pdCoinMarketData.dtypes	
	Date	object
	Open*	object
	High	object
	Low	object
	closeValue	object
	Volume	object
	Market Cap	object
	dtype:	object

❑ preprocessing using re

❑ On Unstructured data (from LexisNexis and Twitter)

	datetime	tweet_id	text	username
0	2021-10-05 14:59:05+00:00	1445403172705607683	investor says " we are at the top of the first...	DocumentingBTC
1	2021-10-05 14:57:04+00:00	1445402664049815555	to launch custody services 's the largest r...	DocumentingBTC
2	2021-10-04 16:40:41+00:00	1445066353556197384	is never down	DocumentingBTC
3	2021-10-04 15:09:26+00:00	1445043389666312195	on news with	DocumentingBTC
4	2021-10-04 14:17:15+00:00	1445030259527753740	growth is going vertical	DocumentingBTC

Date	titles
2021-10-24	NewsWatch: Big Tech stocks are the market's su...
2021-10-24	bZx Protocol (BZRX) Price Hits \$0.33 on Top Ex...
2021-10-24	bZx Protocol Price Reaches \$0.33 on Exchanges ...
2021-10-23	Inside the Courts - An Update From Skadden Sec...
2021-10-23	What's The Deal With Well Everything?!!
...	...

❑ text processing

Forecasting Results

Models, Development, Evaluation

BTC price prediction

Bitcoin

WHAT INVESTORS NEED
TO KNOW ABOUT
ALTCOINS

GUIDE TO BITCOIN

[CRYPTOCURRENCY](#) > [BITCOIN](#)

Why Bitcoin Price Predictions Are Unreliable

By [NATHAN REIFF](#) Updated July 26, 2021

Reviewed by [ERIKA RASURE](#)

Fact checked by [MARCUS REEVES](#)

It's a phenomenon familiar to anyone who follows the [cryptocurrency](#) industry. A prominent figure—the [CEO](#) of a [digital currency exchange](#), a key developer or researcher, a successful cryptocurrency investor—makes a dramatic prediction about the price of [Bitcoin](#) or the general movement of the [digital currency](#) sphere.

Many of these predictions call for major shifts away from the current climate. ("[Bitcoin will hit \\$100,000!](#)" or perhaps, "[Bitcoin will collapse entirely!](#)")

Source: <https://www.investopedia.com/tech/why-bitcoin-price-predictions-are-unreliable/>

What we are forecasting

- Closing price trend of BTC



Models Selected

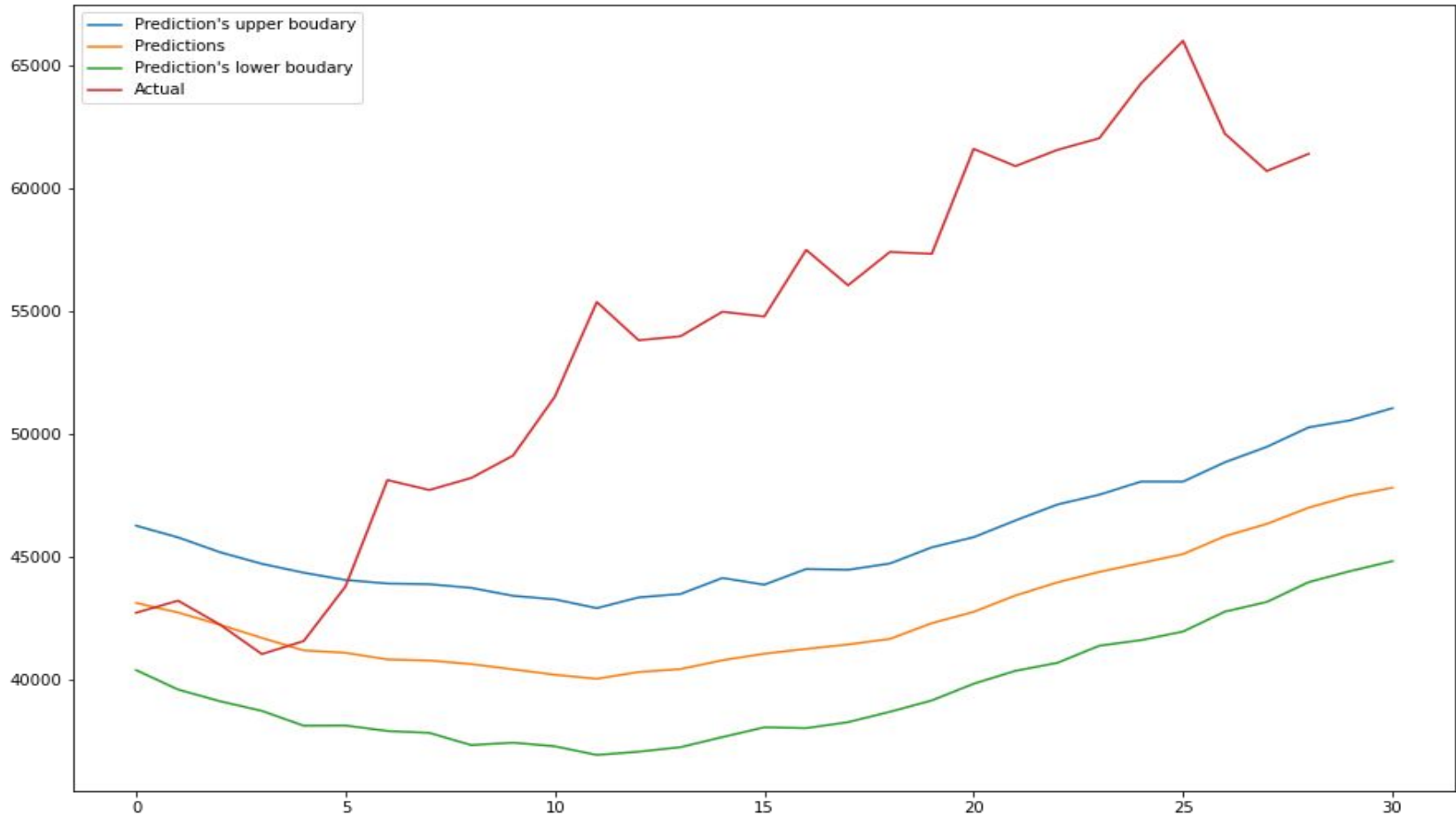
- ❑ We selected following models
 - ❑ Two time-series models: Facebook's Prophet, SARIMA
 - ❑ Text Processing -> TFIDF -> Models (Linear SVM, RF, XGBoost)
 - ❑ BERT -> Models (Linear SVM, RF, XGBoost)

Time Series forecasting

- ❑ How it works?
- ❑ Input for the models?
- ❑ Explored parameters?

Forecasting of BTC Closing price

Using 1 year of data

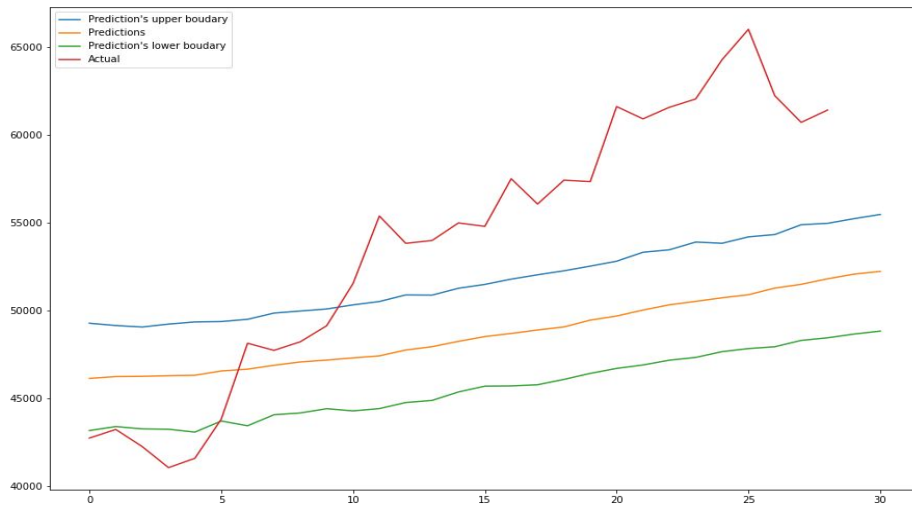


Model: FbProphet

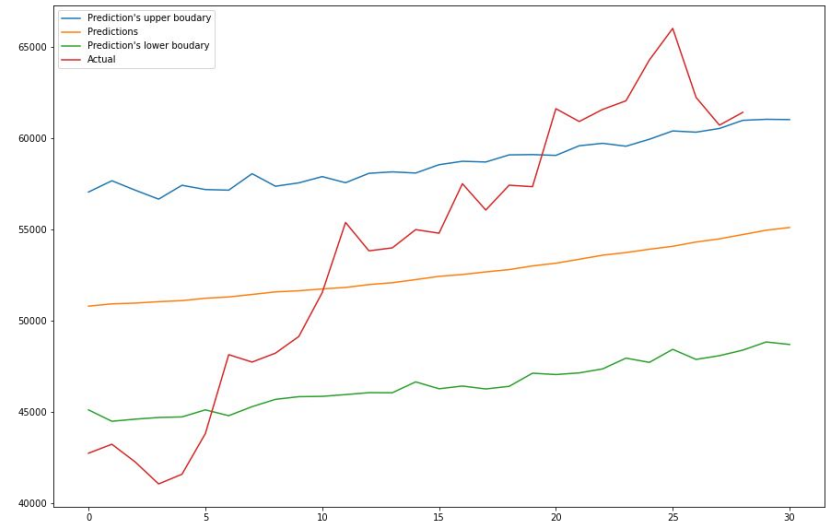
Data Source: **coinmarketcap.com**

Forecasting of BTC Closing price

Using 2 years of data



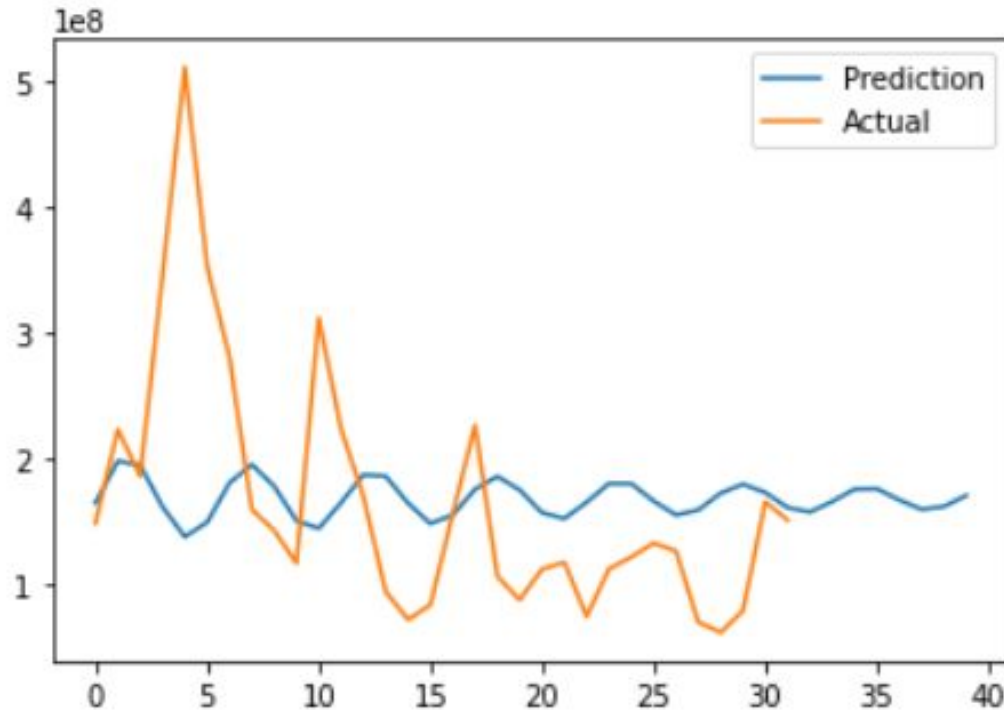
Using 5 years of data



Model: FbProphet

Data Source: coinmarketcap.com

Forecasting of BTC Closing price



Model: ARIMA

Data Source: **coinmarketcap.com**

Preparing text data

```
1 sampleText
```

```
'The messages will be "unwrapped" by sculptor Richard Wentworth, who is responsible for decorating the tree with broken plates and light bulbs. Artists who have decorated the Tate tree in previous years include Tracey Emin in 2002.'
```

```
1 preprocessDataset(sampleText)
```

```
--- Tokenized ----
['the', 'messages', 'will', 'be', 'unwrapped', 'by', 'sculptor', 'richard', 'wentworth', 'who', 'is', 'responsible', 'for', 'decorating', 'the', 'tree', 'with', 'broken', 'plates', 'and', 'light', 'bulbs', 'artists', 'who', 'have', 'decorated', 'the', 'tate', 'tree', 'in', 'previous', 'years', 'include', 'tracey', 'emin', 'in', '2002']

---- Removed Stop Words ----
['messages', 'unwrapped', 'sculptor', 'richard', 'wentworth', 'responsible', 'decorating', 'tree', 'broken', 'plates', 'light', 'bulbs', 'artists', 'decorated', 'tate', 'tree', 'previous', 'years', 'include', 'tracey', 'emin', '2002']

---- Joined as Sentence ----
messages unwrapped sculptor richard wentworth responsible decorating tree broken plates light bulbs artists decorated tate tree in previous years include tracey emin 2002

---- Numbers removed ----
messages unwrapped sculptor richard wentworth responsible decorating tree broken plates light bulbs artists decorated tate tree in previous years include tracey emin

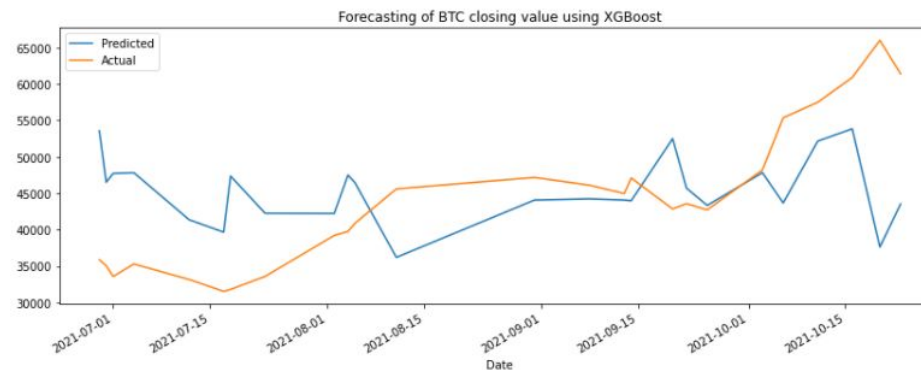
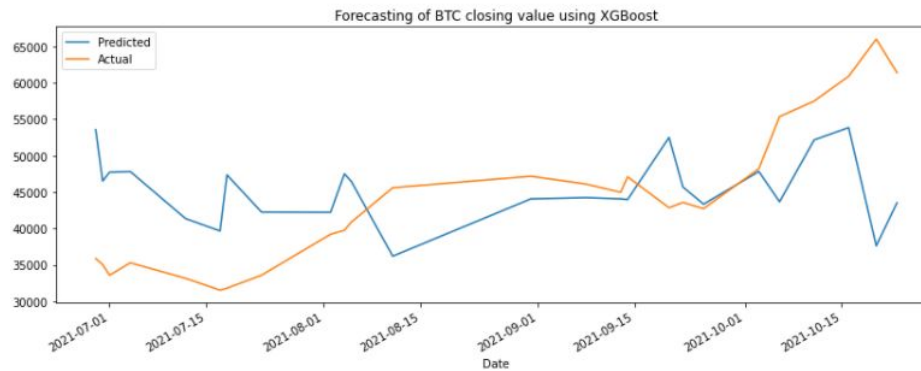
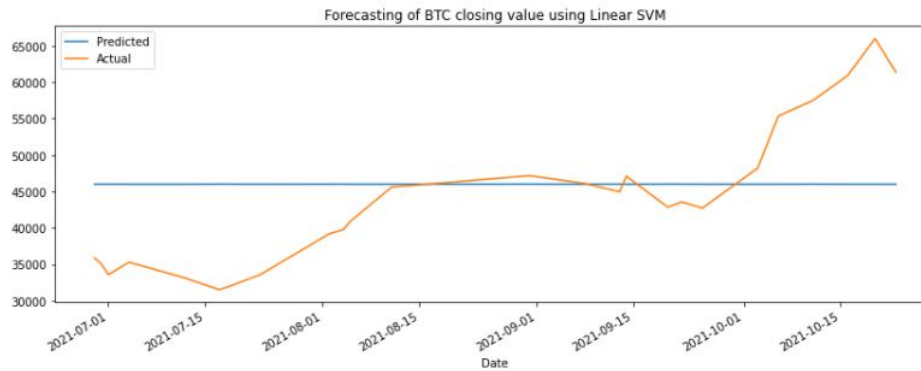
---- Stemmed ----
messag unwrap sculptor richard wentworth respons decor tree broken plate light bulb artist decor tate tree previou year includ tracey emin

---- Lemmatized ----
messag unwrap sculptor richard wentworth respons decor tree broken plate light bulb artist decor tate tree previou year includ tracey emin
```


Preparing text data-vectorization

- ❑ Converted processed text data into machine understandable numbers using TF-IDF: **Term Frequency-Inverse document frequency**
- ❑ $TF = (\# \text{ of reparations of word in a document}) / (\# \text{ of words in a document})$
- ❑ $IDF = \text{Log}(\# \text{ of doc.s }) / (\# \text{ of docs containing the word })$
- ❑ Vectorized to create matrix of all text samples

Forecasting with processed text

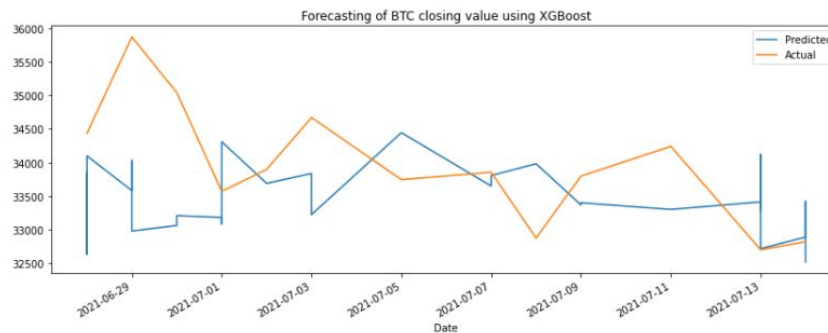
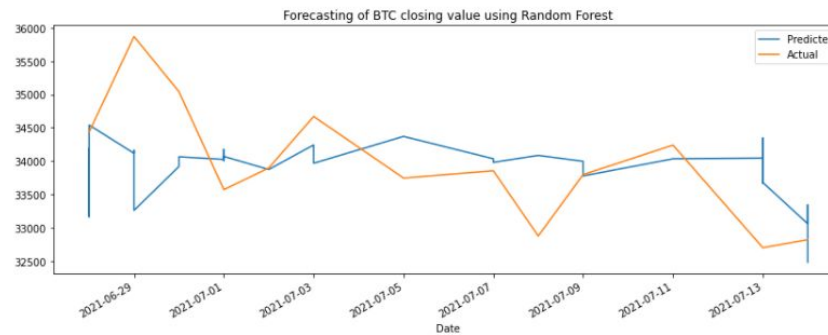
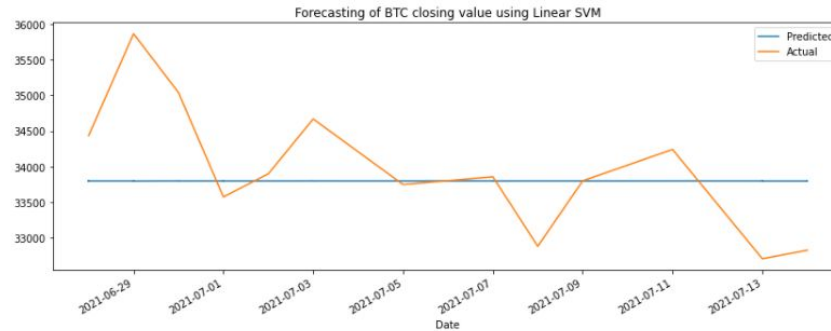


Data Source:
Twitter

Forecasting using text data

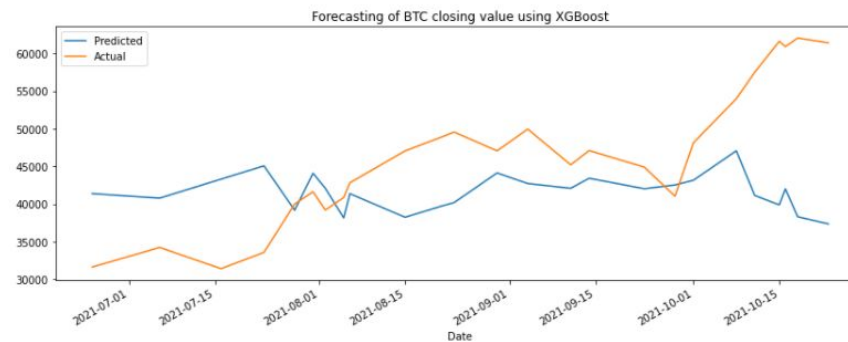
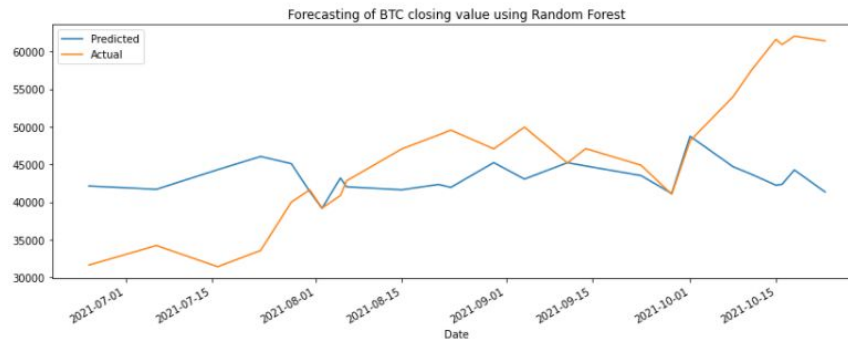
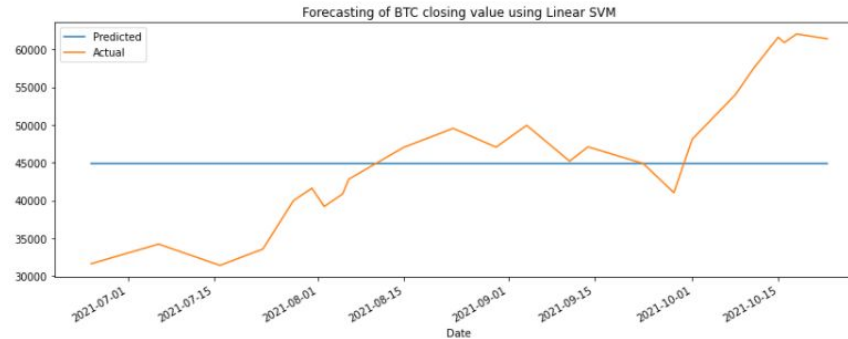
- ❑ Text processing, Vectorization and predicting the values using ML models (Linear SVM, RF, XGBoost)
- ❑ Feature Extraction using BERT (Bidirectional Encoder Representations from Transformers) and predicting the values using ML models (Linear SVM, RF, XGBoost)

forecasting btc using LexisNexis



Data Source:
LexisNexis

forecasting of btc using Tweets



Data Source:
Twitter

Forecasting results

- ❑ Data sources and results
 - ❑ Unstructured performed well
- ❑ Data size and results
- ❑ Models and results

Conclusion

- ❑ There is a scope of improvement
 - ❑ Time series models
 - ❑ Distilled-BERT

- ❑ The business application
 - ❑ To understand the crypto trend
 - ❑ forecasting
 - ❑ grasp the sentiment
 - ❑ unusual market fluctuation
 - ❑ valuable dataset for other research work

- ❑ You can find the code: [GitHub](#)

Thank You