NUS National University of Singapore

DBA5102 Group Project

Predicting Bitcoin Prices

Ankit Malhotra - A0232322X
Donghwan Kim - A0231887U
Felipe Chapa - A0179033E
Sahil Sharma - A0232063U
Widya Gani Salim - A0231857Y
Wong Cheng An - A0232039M

Scan/click to find a surprise Bitcoin visualiser! (Different for each slide)

Table of Contents

Problem Statement

Objective: Find the best data sources and models to predict Bitcoin (BTC) prices.

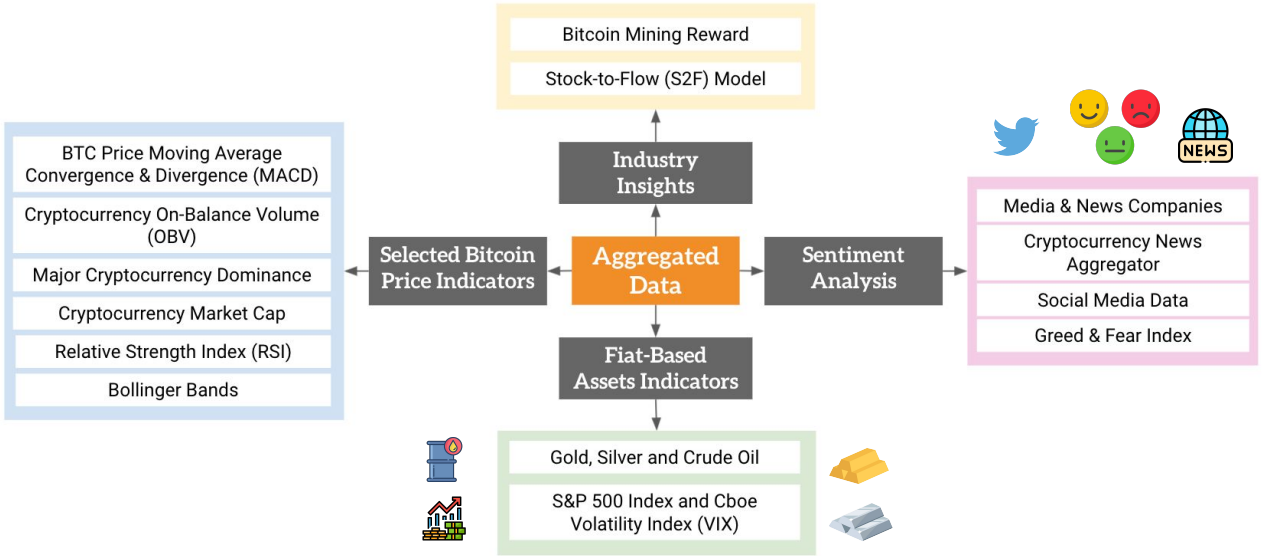
Why is BTC price so volatile?

- Lack of regulations
- Supply limitation/scarcity
- Correlation with other assets/stock market
- Lack of investment infrastructure

NUS National University of Singapore

Data Gathering & Processing

We focused on gathering data which could be categorised into 4 groups:



Industry Insights
Indicators accounting for Bitcoin's scarcity and mining reward.

Selected Bitcoin Price Indicators
Include common cryptocurrency trading indicators such as RSI, MACD, OBV, market dominance and cap.

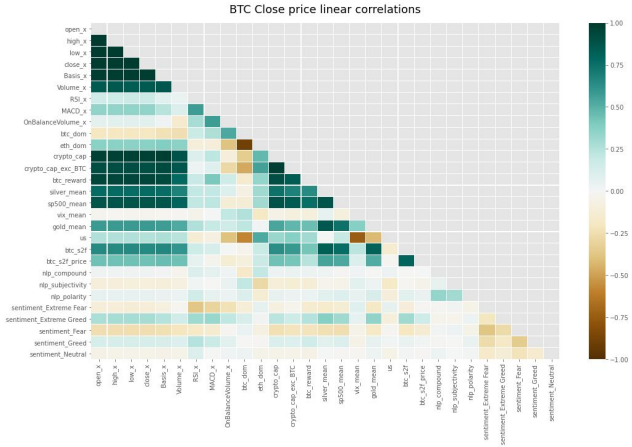
Sentiment Analysis
Indicators that capture the public's and industry sentiment.

Fiat-Based Assets Indicators
Selected other asset/market prices that could be correlated with BTC price.

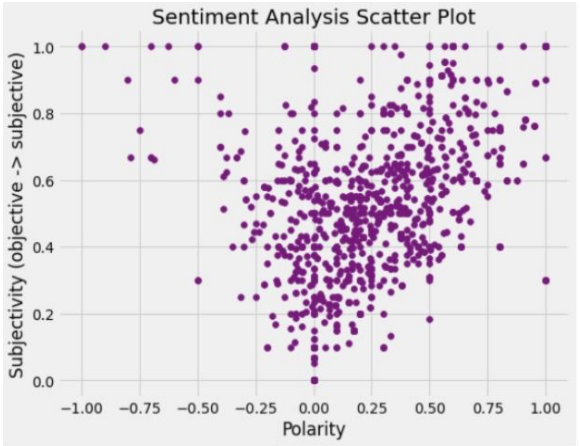
Collected 1,646,826 data points & 123,791 entries from 20 data sources.

NUS National University of Singapore

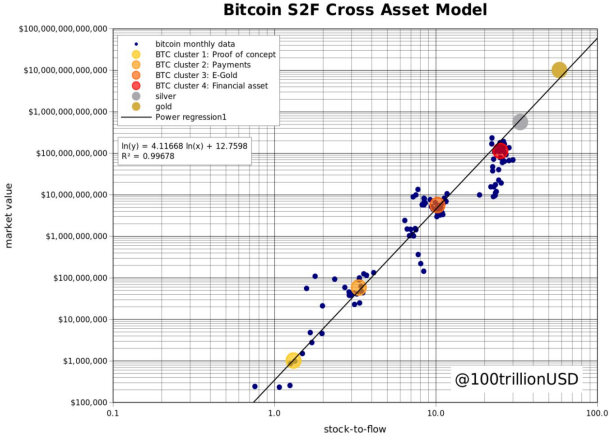
Data Gathering & Processing



Asset 1 - Correlation Heatmap between all features and BTC prices



Asset 2 - Social Media Data Sentiment Analysis Polarity



Asset 3 - Bitcoin S2F Cross Asset Model Graph[1]

Feature Engineering:

- Aggregated articles & indicators from top exchanges and cryptocurrency news aggregators
- Sentiment analysis through Natural Language Processing (NLP) on Tweets and news articles (Web crawling + Twitter API)
- Predicted BTC Stock-to-Flow (S2F)[1] index and estimated price using BTC mining data
- Turned Greed & Fear Sentiment Index into dummy variables

NUS National University of Singapore

Data Gathering & Processing

	time	open_x	high_x	low_x	close_x	Basis_x	Volume_x	RSI_x	MACD_x	OnBalanceVolume_x	...	btc_s2f
0	2021-05-04	42774.461875	42859.831250	39810.560000	39889.267500	38095.884875	2.103081e+08	30.197215	-289.288887	6.627343e+07	...	45.287975
1	2021-05-05	39888.854375	43487.158125	39674.550000	43073.273750	37899.245281	1.928531e+08	36.635091	-152.584587	2.591266e+08	...	52.213717
2	2021-05-06	43080.058750	43772.912500	41432.267500	42298.459375	37729.223031	1.949187e+08	35.091982	-100.557042	6.420790e+07	...	48.223643
3	2021-05-13	37186.293750	38518.738750	34338.900000	37256.128750	37777.979125	2.641556e+08	26.192213	-5			
4	2021-05-17	34817.921875	34964.095000	31562.319375	32654.088750	37037.300406	2.778292e+08	20.371791	-17			
...	...	...	...	...	...	...	...	...				
697	2021-04-26	33734.797500	37362.988750	33531.115000	37123.436875	39213.652875	1.851308e+08	31.445327	-10			
698	2021-04-27	37121.261250	38117.250000	36603.150000	37822.023125	39181.138250	1.231269e+08	33.176546	-9			
699	2021-04-28	37821.973125	38780.800000	36995.347500	37699.113750	39069.627938	1.382809e+08	32.887363	-8			

Selected columns:

```
Index(['time', 'open_x', 'high_x', 'low_x', 'close_x', 'Basis_x', 'Volume_x',
       'RSI_x', 'MACD_x', 'OnBalanceVolume_x', 'btc_dom', 'eth_dom',
       'crypto_cap', 'crypto_cap_exc_BTC', 'btc_reward', 'silver_mean',
       'sp500_mean', 'vix_mean', 'gold_mean', 'us', 'btc_s2f', 'btc_s2f_price',
       'nlp_compound', 'nlp_subjectivity', 'nlp_polarity',
       'sentiment_Extreme Fear', 'sentiment_Extreme Greed', 'sentiment_Fear',
       'sentiment_Greed', 'sentiment_Neutral'],
      dtype='object')
```

From 1.6 million data points & 123,791 rows from 20 data sources collected

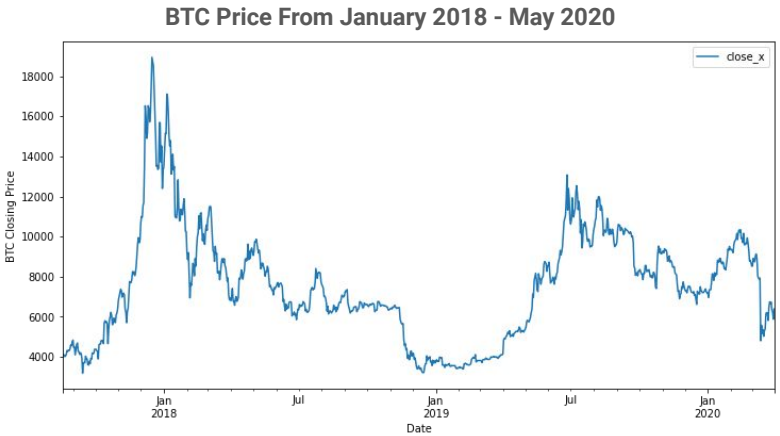
Clean aggregated data has **41,403 data points**, **1,488 entries** & **30 selected features**.

NUS National University of Singapore

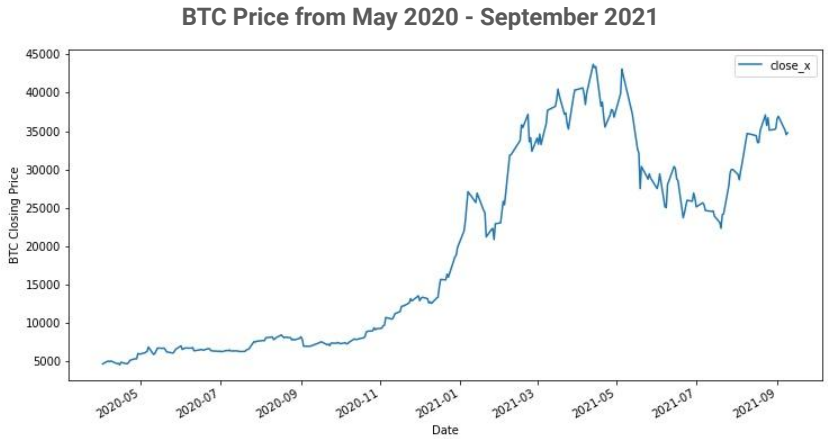
Prediction Strategy

Prediction Horizon: January 2018 - September 2021

Cryptocurrency Winter: 2018 Crash



Cryptocurrency Boom and Crash: 2020 Halving



DBA5102 Group Project: Predicting Bitcoin Prices

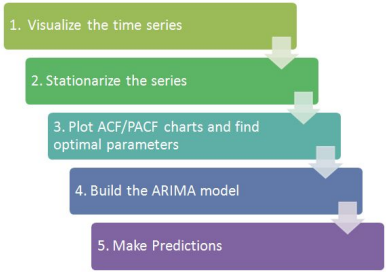
NUS National University of Singapore

Prediction Strategy

Prediction Frequency: Monthly (1M) for SARIMA & Daily (1D) for LSTM

Model Selection

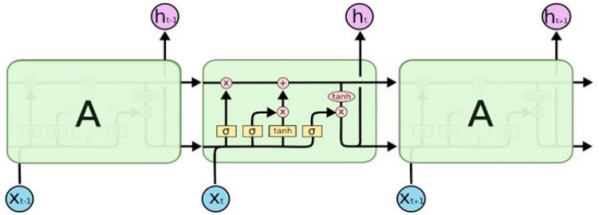
Seasonal Autoregressive Integrated Moving Average (SARIMA)



Asset 4 - SARIMA Thought Process

- Traditional basic methodology for financial time series.
- Requires the linear variation in the stock prices to remain stationary[4].

Long-Short Term Memory (LSTM)



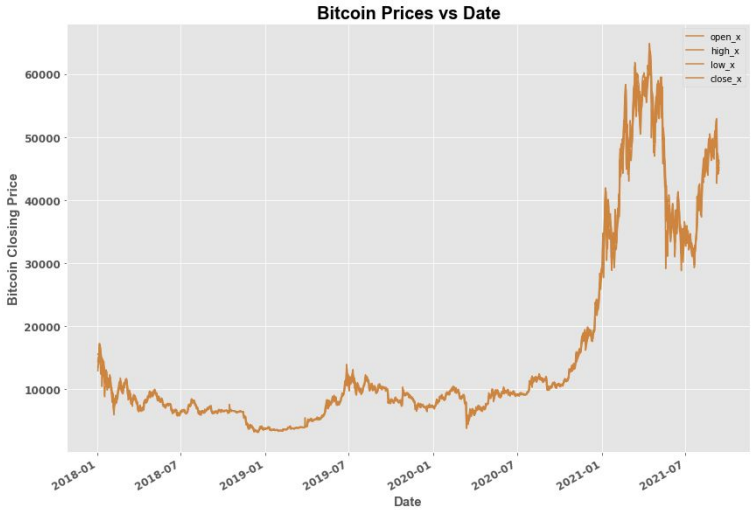
Asset 5 - Internal Structure of LSTM[2]

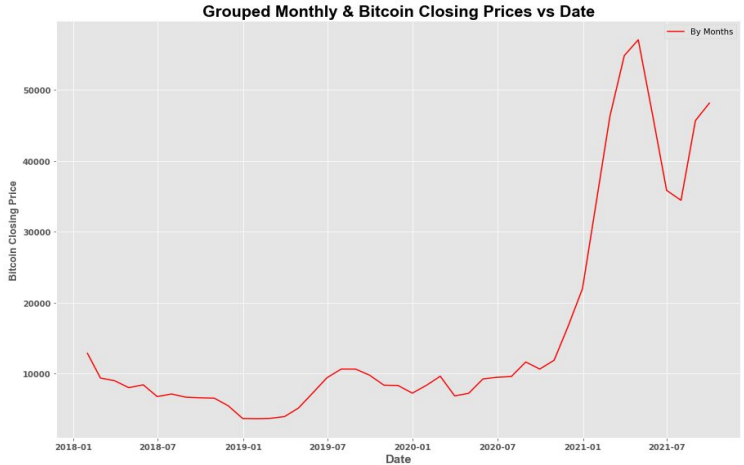
- Suitable for time series prediction, commonly used to predict stock prices[3].
- Able to resolve the problem of long-term dependence with internal gate mechanisms[2].

NUS National University of Singapore

Time Series Modelling

Feature Used	Time Frame	Data Used	Prediction Horizon
Bitcoin Closing Price	Monthly	Past 4 years	5 months (Apr ’21- Sep ’21)



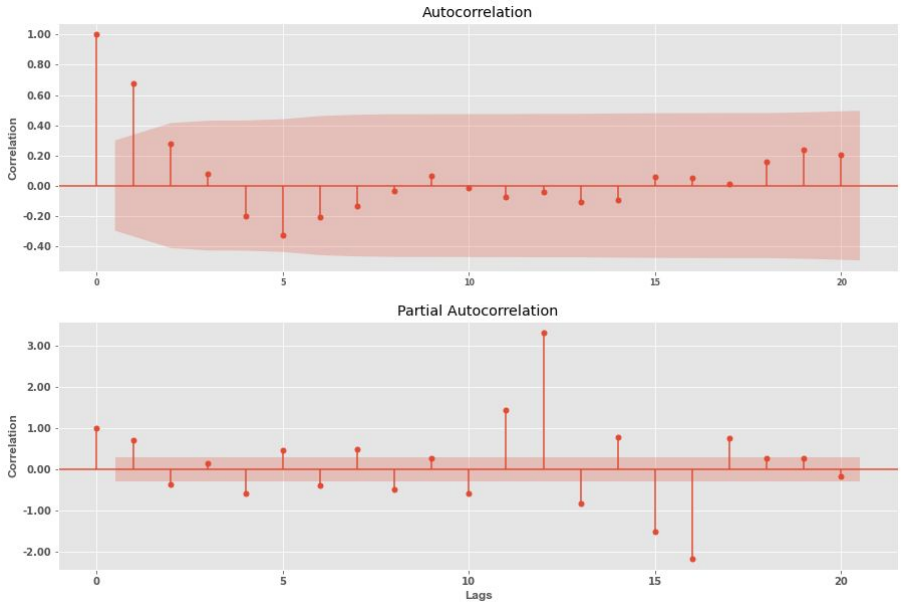


DBA5102 Group Project: Predicting Bitcoin Prices

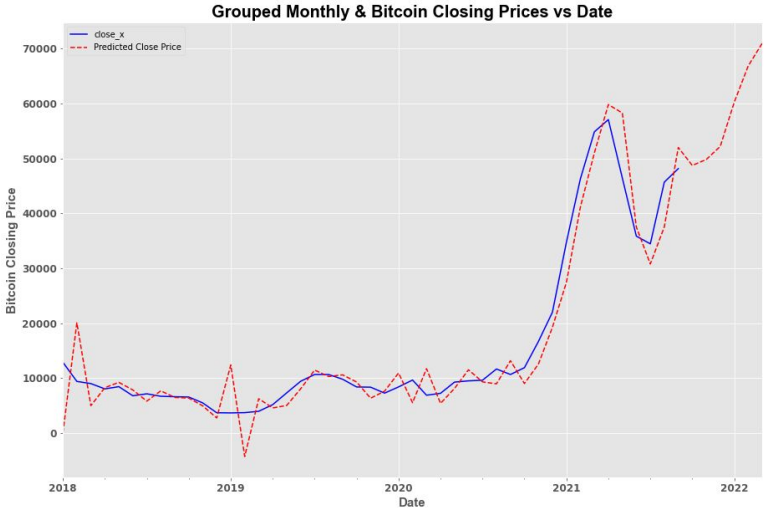
NUS National University of Singapore

Time Series Modelling

Plotting ACF and PACF plots



Prediction Result (SARIMA)



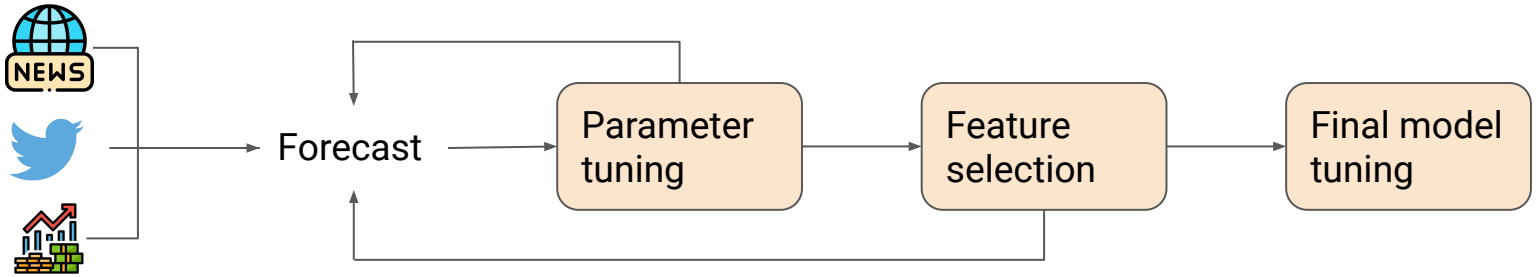
Mean Absolute Percentage Error (MAPE): **31%**
Root Mean Squared Error (RMSE): **4,478.52**

DBA5102 Group Project: Predicting Bitcoin Prices

NUS National University of Singapore

Long-Short Term Memory (LSTM) Model

Features Selection	Train RMSE	Val. RMSE	Val. MAPE
Sentiment Analysis (TextBlob library), Bitcoin Price Indicators	676	3,627	7.3%
Bitcoin Price Indicators	791	3,698	7.2%
Bitcoin Price Indicators, Industry Insights, Fiat-Based Asset Indicators BTC Trade, Sentiment Analysis	710	4,448	9%
...	...	...	...

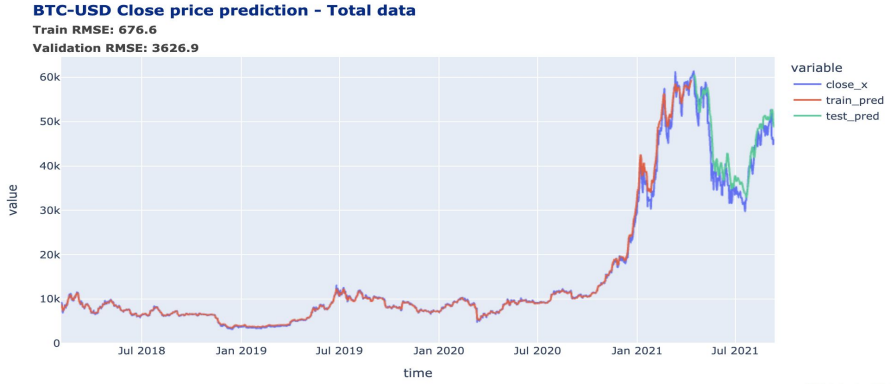


NUS National University of Singapore

Final Model

Long-Short Term Model (LSTM): 1-Day step forecasting trained on 6 features

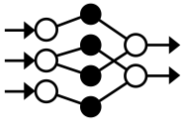
Feature Used	Time Frame	Hidden Layers	Dropout	Epochs	Data Used	Timestep	Validation Horizon
BTC Trade indicators, Sentiment Analysis	Daily	4	~20%	50	Past 4 years	3 days	5 months (Apr '21- Sep '21)



DBA5102 Group Project: Predicting Bitcoin Prices

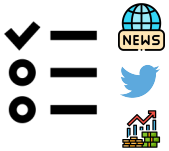
NUS National University of Singapore

Conclusion

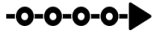


LSTM model:
- MAPE: 7%
- RMSE: 3627

Prediction frequency
- Daily step forecast



Feature selection
- 30 indicators
- Financial & NLP



Time horizon
- Feb 2018 - Sep 2021
- Including a boom/crash cycle

Potential improvements:

- Longer time horizon to include more BTC halving and cryptocurrency boom & crash cycles
- Reevaluate feature engineering and selection to get more value from existing data
- Higher tuning of chosen model
- Include S2F of other assets - such as Gold or Silver - as a comparison
- For sentiment analysis, assign weightage on selected influencers and number of Tweets (e.g. based on tweet likes/comments)

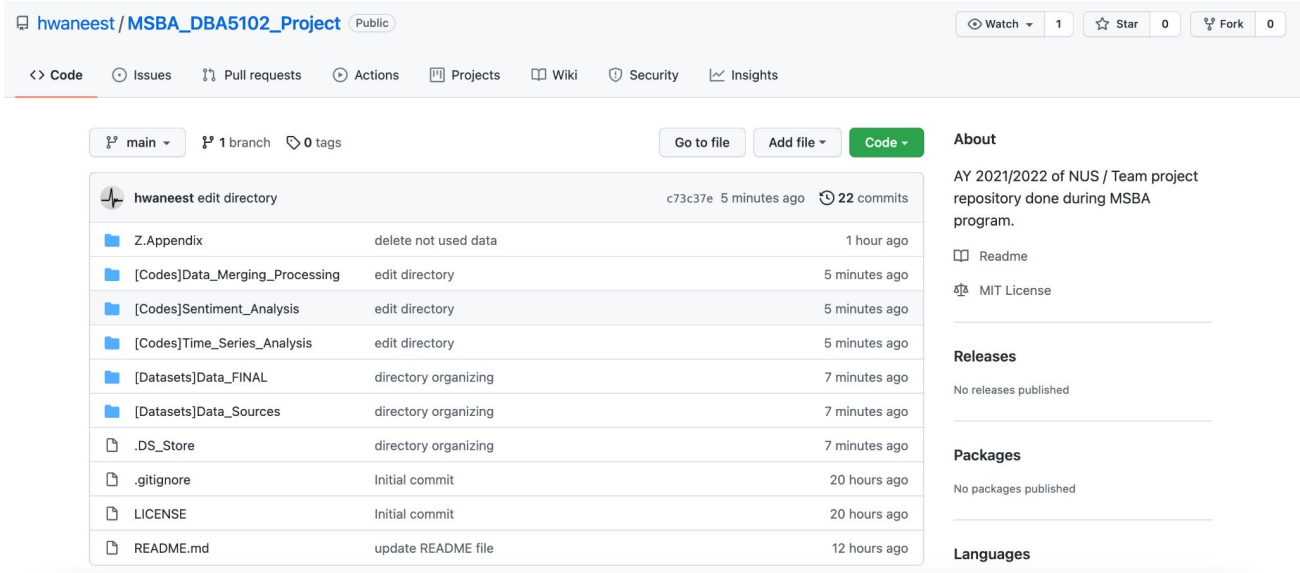
NUS National University of Singapore

Thank You

Appendix

Code & Data Repository

All datasets and codes can be found on our Github Repository:
https://github.com/hwaneest/MSBA_DBA5102_Project



NUS National University of Singapore

Data Sources Breakdown

Data Source	Details	Category/Used in
Coinmetrics.io	BTC Mining Reward (used to create S2F index)	Industry Insights
Blockchair	BTC Mining Reward (benchmark)	Industry Insights
Trading View	BTC and ETH Market Dominance	BTC Price Indicators
Coinmarketcap	Cryptocurrency Market Cap	BTC Price Indicators
Top 5 Cryptocurrency Exchanges (by Volume)	Scraped trading prices and indicators data from Coinbase, Binance, FTX, Huobi, KuCoin	BTC price indicators
NASDAQ	S&P 500 Data	Fiat-Based Assets Indicators
Currency.com	Gold Price Index, Silver Price Index	Fiat-Based Assets Indicators
TVC	VIC Index, Crude Oil Prices	Fiat-Based Assets Indicators

Data Sources Breakdown

Data Source	Details	Category/Used in
Twitter	**Bitcoin influencers selected:** Tone Vays, John Mcafee, Vitalik Buterin, Andreas M. Antonopoulos, Tim Draper, Roger Ver, Elon Musk, Cathie Wood, Jack Dorsey, Michael J. Saylor, CobraBitcoin, Erik Voorhees, Ben Armstrong, Vinny Lingham, Adam Back, Gavin Andresen, Nick Szabo **News station selected:** Bloomberg Technology, Bloomberg Crypto, Bloomberg, MarketWatch, The Wall Street Journal, The Economist, CNBC's Fast Money, CNBC, CNN Money, CNN business, CNN Breaking News, Financial Times, Forbes Crypto, Barron's, Yahoo Finance, Traders Magazine, MetaStock, Benzinga Crypto	Sentiment Analysis
Top 5 Cryptocurrency News Aggregator	CCN, NewsBTC, Cointelegraph, Coindesk, Forklog	Sentiment Analysis
Alternative.me	Greed & Fear Index	Sentiment Analysis
Kaggle	Raw data of BTC news aggregator	Sentiment Analysis

Brief Explanation on Selected Features

Indicators	Details
Relative Strength Index (RSI)	Momentum indicator to measure the magnitude of recent price changes, measuring whether an asset is oversold or overbought.
Moving Average Convergence & Divergence (MACD)	Indicator measuring the relationship between two moving averages of an asset's price, 26-period Moving Averages and 12-period Moving Averages.
Bollinger Bands	Statistical chart measuring possible upper and lower bound of asset prices.
Market Dominance (BTC/ETH)	Market cap dominance of a cryptocurrency.
On Balance Volume (OBV)	Momentum indicator that measures volume flow or changes to predict prices.
Cryptocurrency Greed & Fear Index	Market indicator that takes into account volatility, dominance, media sentiments, trends and volume.

Brief Explanation on Selected Features

Indicators	Details
Bitcoin S2F Index	Index that divides stock (the size of existing stockpiles) with flow (year production), used to measure Bitcoin scarcity.
Bitcoin S2F Index Predicted Price	Predicted price based on the Bitcoin S2F Index model (refer to “Asset & Study References” slide for more details).
Bitcoin Mining Reward	Bitcoin block reward against price at each period.
NLP Polarity	Classifies negative and positive sentiments. Values range between (-1,1)
NLP Compound	Classify the text if it is extremely negative or positive. Values range between (-1,1)
NLP Subjectivity	Distinguishes between facts and people opinions. Values range between (0,1)

Asset & Study References

[1] - **Bitcoin Stock-to-Flow Cross Asset Model by Plan B**
https://medium.com/@100trillionUSD/bitcoin-stock-to-flow-cross-asset-model-50d260feed12

[2] - **Stock Market Prediction Using LSTM Recurrent Neural Network by Adil Moghara & Mohamed Hamiche**
https://www.sciencedirect.com/science/article/pii/S1877050920304865

[3] - **MRC-LSTM: A Hybrid Approach of Multi-scale Residual CNN and LSTM to Predict Bitcoin Price by Qiutong Guo, Shun Lei, Qing Ye and Zhiyang Fang**
https://arxiv.org/pdf/2105.00707.pdf

[4] - **The Interval Slope Method for Long-Term Forecasting of Stock Price Trends by Chun-xue Nie & Xue-bo Jin**
https://www.hindawi.com/journals/amp/2016/8045656/

[5] - **Time Series Analysis in Python: An Introduction**
https://towardsdatascience.com/time-series-analysis-in-python-an-introduction-70d5a5b1d52a

Key Libraries Leveraged

Web crawling & data mining
Beautiful Soup
NLTK
Textblob
Alternative.me API
Blockchair API
Coinmetrics API
Twitter API

LSTM - Neural Network
Keras
Tensorflow

SARIMA - Time Series
Statsmodels

Visualisation
matplotlib
Plotly
Seaborn

Others
Multiprocessing
Numpy
Pandas
Pickle
Sklearn

Thank You