Twitter Sentiment Analysis & Bitcoin Price Prediction

Oleg Ananyev, Briana Hart, Neil Prabhu

# *Introduction*

## Problem Statement

Bitcoin is a very liquid financial asset that is traded through centralized exchanges. As with any financial asset, there is opportunity to earn a return through the buying and selling of the asset or its derivatives. Financial institutions of all different types and scales have complex financial models that attempt to predict future prices of financial assets. These models can be based on fundamental valuation or technical indicators, or a mix of both. Because Bitcoin doesn't have any fundamental intrinsic value, we would like to determine if we can build a model to determine Twitter sentiment on Bitcoin. This would further allow us to determine if individuals are “bullish” (i.e., in favor) of Bitcoin or “bearish” (i.e., against) and determine if this impacts the price.

In addition, economic indicators (e.g., GDP, money supply, VIX, etc.) and technical indicators (e.g., Moving Averages, MACD, RSI, etc.) can impact financial assets and markets, very significantly in certain circumstances. We will incorporate some of these indicators in our models to determine if they have an impact on Bitcoin prices.

## Objective

The objective of this study is to predict Bitcoin prices through the use of Twitter sentiment and economic and technical indicators.

## Approach Methodology

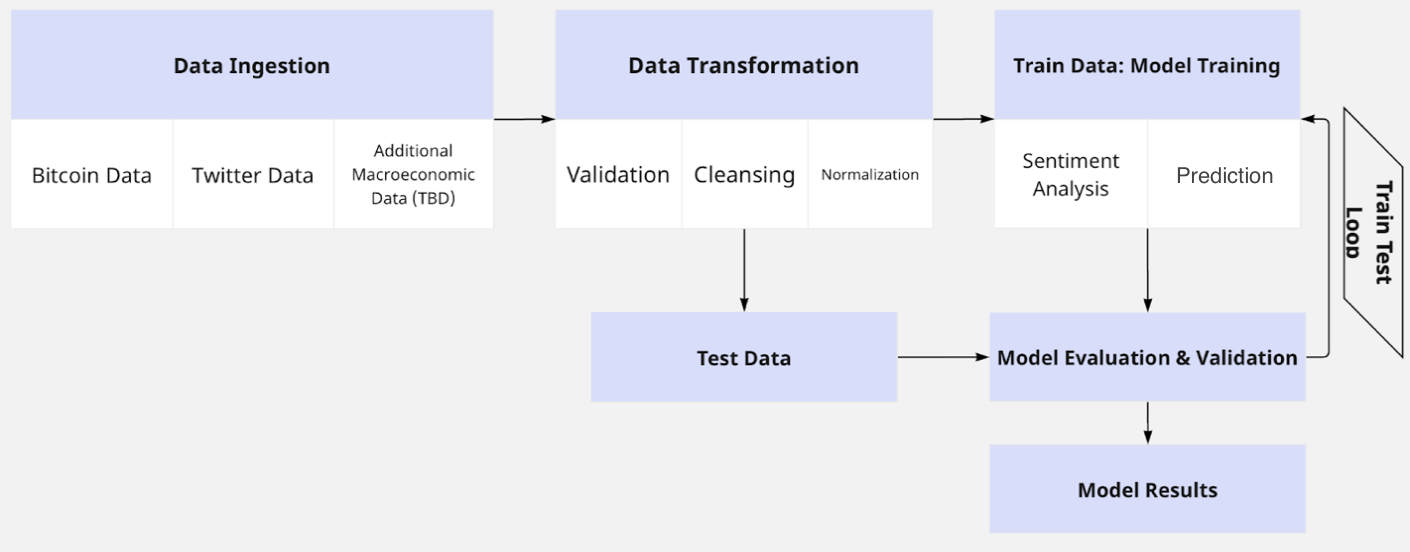
We will model Twitter sentiment using tweets related to Bitcoin. We will also include macroeconomic indicators and technical indicators to determine their impact on predicting Bitcoin prices. Examples of these indicators include money supply, GDP growth rates, consumer confidence, federal bond issuance, S&P 500, and many others.

For Sentiment modeling, we will leverage a K-means clustering algorithm to group Tweets with similar meanings into a cluster. We will create a two and three cluster model to split the tweets into positive/negative and positive/negative/neutral clusters, respectively.

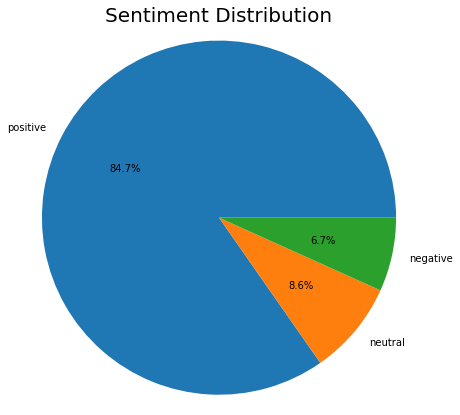
For Bitcoin price prediction, we will leverage the following algorithms and compare their effectiveness:

* Logistic regression
* Bagging
* Gradient boosting
* Random Forests
* Voting - Top 3 performing algorithms

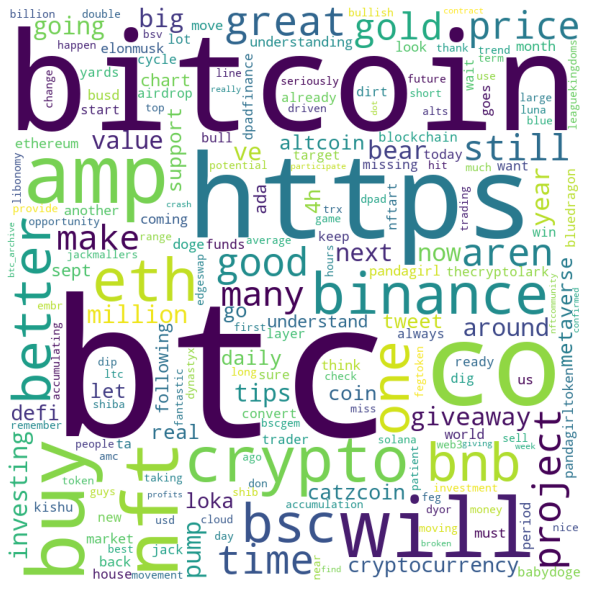
## Block Diagram

****

## Images



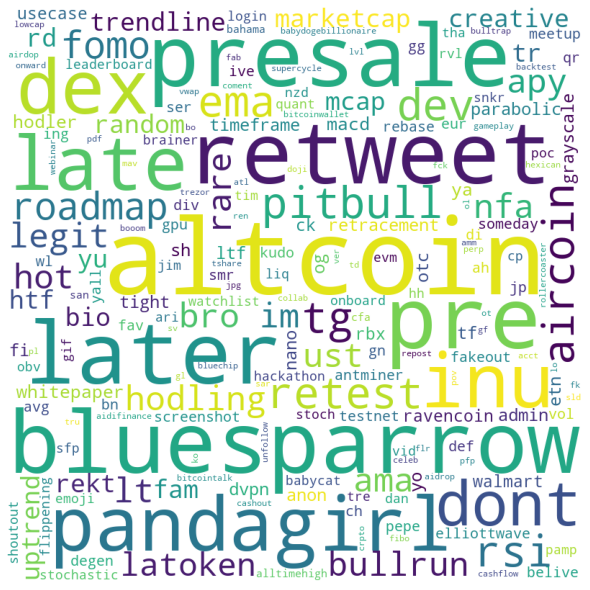
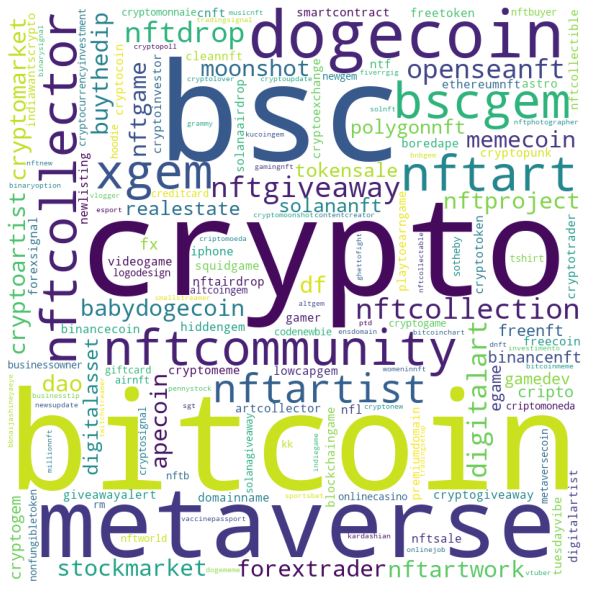
***Chart 1:*** *Sentiment Distribution of All Tweets*



***Images 1 & 2:***

*Left: Wordcloud of Words from Positive Tweets*

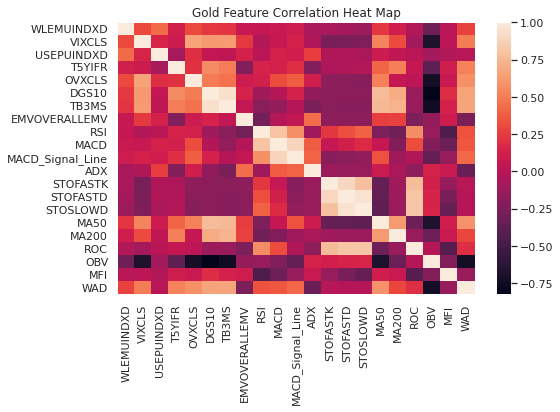
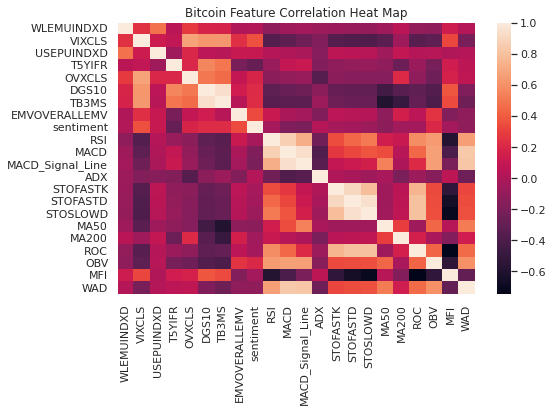
*Right: Wordcloud of Words from Negative Tweets*



***Images 3 & 4:***

*Left: Sentiment Analysis Positive Words*

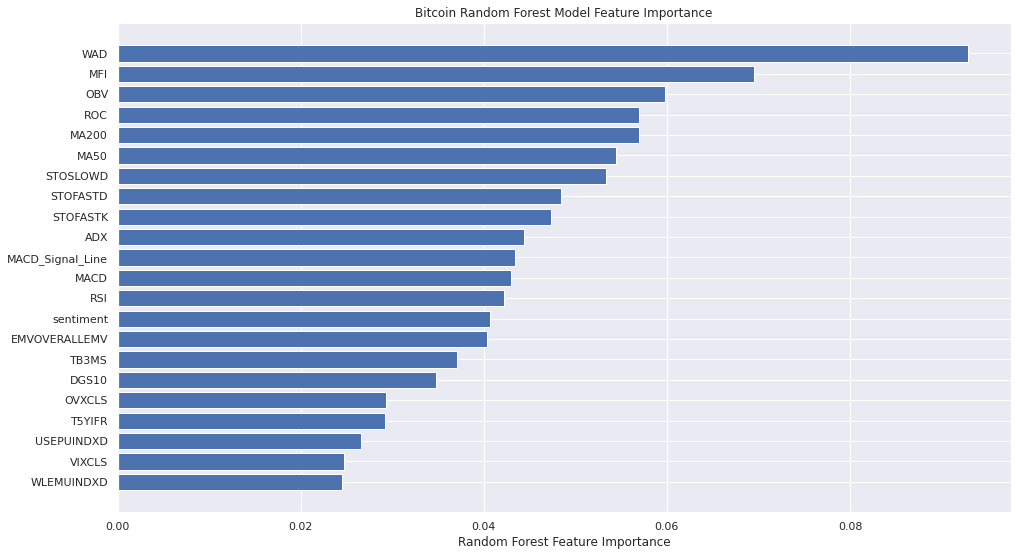
*Right: Sentiment Analysis Negative Words*



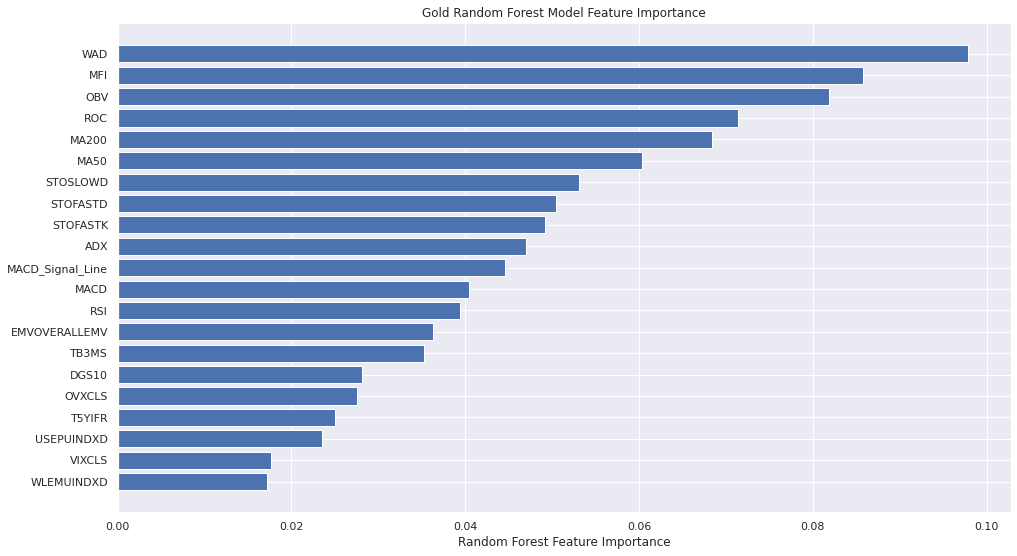
***Charts 2 & 3:***

*Left: Bitcoin Feature Correlation Heat Map*

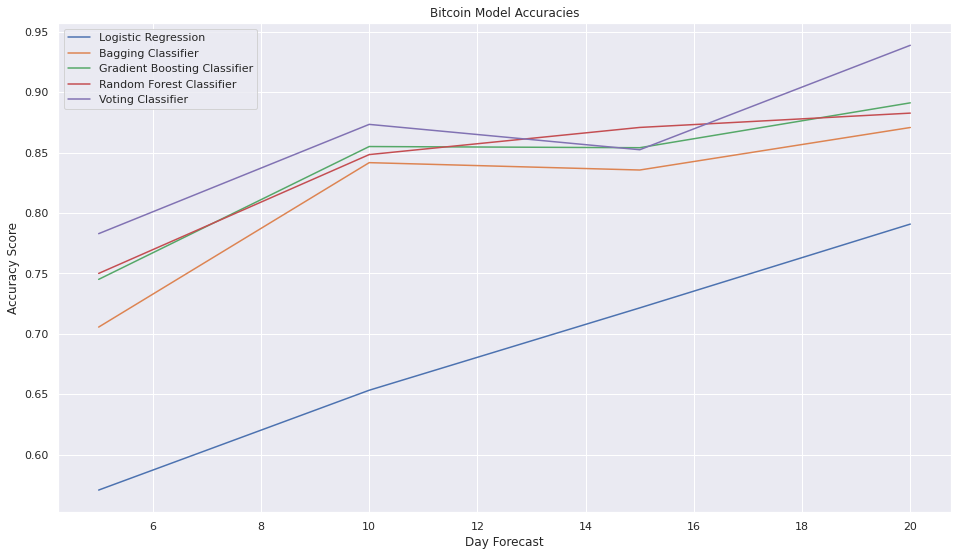
*Right: Gold Feature Correlation Heat Map*



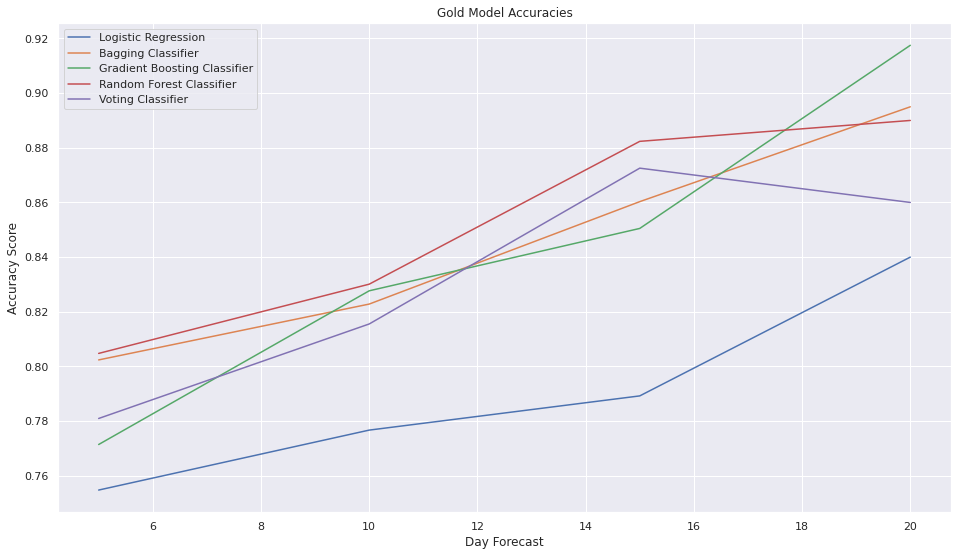
***Chart 4:*** *Bitcoin Random Forest Feature Importance*



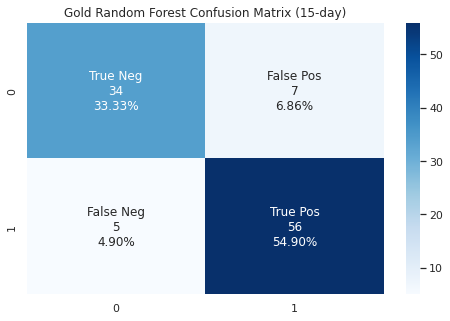
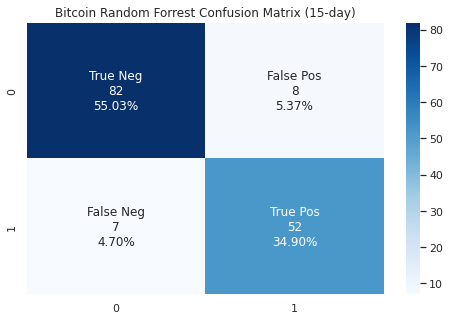
***Chart 5:*** *Gold Random Forest Feature Importance*



***Chart 6:*** *Bitcoin Model Accuracies*



***Chart 7:*** *Gold Model Accuracies*



***Chart 8:*** *Random Forest Confusion Matrices for Bitcoin and Gold (15-day forecast)*

|  | Bitcoin | Gold |
| --- | --- | --- |
| Accuracy | 0.90 | 0.88 |
| Mis Classification | 0.1 | 0.12 |
| Sensitivity | 0.88 | 0.92 |
| Precision | 0.91 | 0.83 |
| F1 | 0.90 | 0.87 |

***Chart 9:*** *Random Forest Model Metrics for Bitcoin and Gold (15-day forecast)*

# Data & Evaluation

## Datasets

### Twitter Data

For Twitter Bitcoin Sentiment Analysis, we utilized a Kaggle database with Bitcoin related Twitter data. The data ranges from February 2021 to July 2022.

* Twitter data: <https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets>

### Financial Data

For macroeconomic indicators, we utilized data from the Economic Research Federal Reserve Bank of St. Louis. For our technical indicators, we used calculated metrics commonly used on stocks in the stock market. For our main two data sources we had Bitcoin and gold price datasets from Yahoo! Finance.

Macroeconomic Indicators

* Inflation, Consumer Prices: <https://fred.stlouisfed.org/series/FPCPITOTLZGUSA>
* Equity Market Volatility Tracker: <https://fred.stlouisfed.org/series/EMVOVERALLEMV>
* 3-Month Treasury Bill: <https://fred.stlouisfed.org/series/TB3MS>
* 5-Year Forward Inflation: <https://fred.stlouisfed.org/series/T5YIFR>
* Mrkt Yield on U.S. Treasury Securities (10-yr): <https://fred.stlouisfed.org/series/DGS10>
* Equity Market Econ. Uncertainty: <https://fred.stlouisfed.org/series/WLEMUINDXD>
* Econ Policy Uncertainty: <https://fred.stlouisfed.org/series/USEPUINDXD>
* CBOE Volatility Index: <https://fred.stlouisfed.org/series/VIXCLS>
* Crude Oil ETF Index: <https://fred.stlouisfed.org/series/OVXCLS>

Bitcoin and Gold Price Data Sets

* Bitcoin Yahoo! Finance: <https://finance.yahoo.com/quote/BTC-USD/>
* Gold Yahoo! Finance: <https://finance.yahoo.com/quote/BTC-USD/>

Technical Indicators

* RSI - Relative Strength Index
* MACD - Moving Average Convergence Divergence
* MACD Signal - Moving Average Convergence Divergence Signal Line
* ADX - Average Directional Moving Index
* STOFASTK - Stochastic Fast K value
* STOFASTD - Stochastic Fast D value
* STOSLOWK - Stochastic Slow K value
* STOSLOWD - Stochastic Slow D value
* MA50 - Moving Average for 50 days
* MA200 - Moving Average for 200 days
* ROC - Price Rate of Change
* OBV - On Balance Volume
* MFI - Money Flow Index
* WAD - Williams Accumulation Distribution

## Success/Failure Criteria

We will judge our success or failure on our ability to accurately predict a price movement. We will consider a successful model to have at least 50% accuracy, with a goal accuracy in mind of at least 75%. If the model is not able to accurately predict at least 50% of the time, this would be a failed model and would not be useful to speculate an asset’s value or earn a return.

## Evaluation Parameters

In order to evaluate which model performs best for classification, we chose the confusion matrix which in turn gave us Accuracy, Precision, and F1 score. These were the primary metrics used to determine if the model was performing well. Day forecasting was another parameter we had used to determine how our models perform over time. This parameter will be discussed in more detail in the results section.

# Model & Results

## Modeling

As the goal of the project was to examine Bitcoin sentiment on Twitter and use it to predict Bitcoin price classification, we started with a sentiment analysis on the Twitter data. We first cleaned the Twitter data for Sentiment modeling. This included removing NA values, which was a very small amount of data, and making all text lowercase to make sure the sentiment models don't differentiate between uppercase and lowercase letters. We then filtered for Tweets where the official source was “Twitter for iPad'', “Twitter for iPhone”, “Twitter for Android”, or “Twitter for Web App” to increase the chance of analyzing Tweets by legitimate users. We then reviewed the minimum length of tweets and it appeared as though Twitter has minimum limits in place (~8 characters), so we decided not to alter this. Additionally, we replaced all URLs, ticker symbols, and user names with a space. We removed single letter words and replaced words with root words (e.g., buying vs buy). Lastly, we removed stop words and filtered for non-alphanumeric words. This data cleansing allowed us to have a dataset ready to model Bitcoin sentiment. This sentiment was used as an input into the Bitcoin price prediction models.

We then built a Word2Vec model of the tweets. We used the Word2Vec model to build a K-means model of the individual words in the tweets with a two-cluster and a three-cluster model. We tested various words that are known to be positive or negative (positive: rocket, bullish, successful, winning, etc.; negative: garbage, bearish, losing, sink, etc.), and confirmed that the clusters accurately classified the words from tweets. The results section will provide more information on how this furthered our initial hypothesis.

When deciding on which ML classification models to use for the Bitcoin prediction model, we found Logistic Regression was a good base model and Bagging, Gradient Boosting, and Random Forest Classification were more effective for predicting success or failure.

When fitting the Logistic Regression model, we had to normalize input variables for models to compile. Bagging, Gradient Boosting, and Random Forest models worked well without any tuning. For Gradient Boosting we found the default learning rate of 0.1 to be best and for Random Forests the default number of estimators of 100 to be ideal while raising the number of trees did not have a much lower error rate. For Bagging, we did not tune any of the parameters.

In addition to the four models, we created another model, the Voting Classifier. This model would take in as input different estimators (models) and choose the best performing model to output the best prediction. We will touch more upon how this model performed in the results sections.

To supplement our Bitcoin Models we decided to also utilize gold price data to create similar models on gold to compare both and interpret the results. The gold models and Bitcoin models would be identical in features except for one feature, sentiment, which was calculated within our Twitter sentiment model.

## Experiments

For Bitcoin price prediction, in order to get to our final product we experimented with choosing different date ranges, filling NA values (e.g. data sets with varying rows like monthly vs. daily data), and changing the number of epochs.

When deciding on a date range, we initially had done time series analysis on our data sets to choose the most suitable date range. However, we found that a few of our macroeconomic indicators had end dates much sooner than our Bitcoin, gold, and Twitter data. To resolve the problem, we removed the indicator with the issue. This ended up not affecting our models' evaluation parameters. Additionally, the Twitter bitcoin sentiment dataset only spanned from February 2021 to July 2022, whereas, the Bitcoin and gold datasets pulled from Yahoo! Finance was able to go as far back as 2019. The Twitter sentiment dataset was only a single indicator amongst twenty-one other indicators, so we tried moving our beginning date further back than February 2021 and removed the Twitter sentiment dataset to see if this would improve or deteriorate our model’s accuracy. Our finding was that the model did slightly improve in accuracy and consistency especially when noting the average accuracy over the training epochs.

A secondary experiment was deciding upon whether or not we wanted to remove rows with NAs or fill them with the average. This testing consisted of two parts: filling NAs in the macroeconomic datasets and filling NAs after calculating the technical indicators of the given Yahoo! Finance data (i.e. Open, Close, Volume High, Low etc.). For the former, we found that all NAs were rows that had data for the next row and so we decided to fill the NA with the next row’s data. For the latter, we found that filling NAs using the average of the technical indicators generally provided a better result.

## Results

Using Voting, Random Forests, Gradient Boosting, Bagging, and Logistic Regression gave us a broad spectrum of results for both gold and Bitcoin. Before getting into results, it is important to note day forecasting was done in increments of 5, starting from 5 and going up to 20 days. N forecasted days were calculated as the labels and the N number of days subtracted from the respective dataset to account for the number of days forecasted.

From Charts 6 and 7, Logistic Regression gave the lowest accuracy score regardless of the day forecast. This ranged from ~ 55% at 5 days to 78% at 20 days forecasting.

For our other models (Voting, Random Forests, Gradient Boosting and Bagging), we saw competition. At 20-day forecast, the Voting Classifier was usually ranked at the top in accuracy along with Random Forests and Gradient Boosting Classifiers tying for 2nd and 3rd and lastly the Bagging Classifier in 4th. As seen in both charts the accuracy of the models trend upwards as the N number of days forecasted goes up as well. This means that the farther ahead the forecast, the more accurate the prediction. The sweet spot is around the 15 day mark where we see Random Forests usually coming out 1st in accuracy.

An interesting note here is that the Voting Classifier for both gold and Bitcoin models included only Random Forests, Gradient Boosting and Bagging Classifiers as Logistic Regression was not worth including because it acted as a baseline. Both graphs have similar patterns where Bagging, Gradient Boosting and Random Forests perform within a few percent accuracy of each other.

When discussing feature importance, we will observe charts 4 and 5. When determining feature importance, we used the model with the average best performance, Random Forest. For Bitcoin, the technical indicators were the most important, sentiment (two-cluster weights) indicator was in the middle, and macroeconomic indicators were the least important, filling up the lower half. For gold, there are relatively similar results to Bitcoin, except the sentiment indicator. The technical indicators are still ranked among the top and macroeconomic indicators toward the bottom.

In general, we found both gold and Bitcoin models produce similar results and that over the 5 to 20 day forecasting periods there was an upward trend in accuracy. The feature importances were nearly the same in indicator types and ranking, with the addition of sentiment for Bitcoin. As we continue we will discuss the value of including gold as a comparison.

## Constraints

The primary constraint in this project is that we were creating ML models using a very elementary understanding of them and the coding logic to create them. Time was an added constraint as we did not get a lot of time to play around with the model to further tune and improve it. The primary constraint related to time was that we did not have the time to process and clean Tweets from the API, as the process would have taken months, so we had to use already cleaned Tweets from Kaggle.

## Standards

Seaborn: sns - Version 0.11.2

TensorFlow: tf - Version 2.9.1

Sklearn: sklearn - Version 1.1.1

Statsmodel: statsmodels.api - Version 0.13.0

Matplotlib: Pyplot - Version 3.4

Nltk: WordNetLemmatiser - Version 3.7

Nltk: Stopwords - Version 3.7

Mlxtend: Scatterplotmatrix

Gensim: Word2Vec - Version 0.11.1

Pandas: pandas\_ta - Version 0.3.14b0

Plotly: plotly - Version 5.5.0

Yahoo Finance: yfinance - Version 0.1.74

## Comparison

As we achieved the goal of applying Twitter Sentiment data on Bitcoin, we wanted to compare our results and models from Bitcoin to another sought after commodity, gold. Gold price data was easily accessible and we applied many of the same indicators we had done on Bitcoin. After applying models for both Bitcoin and gold we could see relatively similar results in model performance and data. The only difference which could have explained the difference in accuracy would be the addition of sentiment from the Twitter Sentiment Analysis to our Bitcoin models.

Below are the charts and metrics for Voting Classification on Bitcoin and gold (Note: Voting Classification combines our top three models in performance: Bagging, Gradient Boosting, and Random Forest Classifiers. Voting Classifier chooses the best model for the given prediction).

# Conclusion

With Bitcoin drawing public attention, the idea of forecasting its price direction seemed to be important. Our goal was to predict Bitcoin price direction with macroeconomic and technical indicators and Twitter Sentiment Analysis data. In doing so, we were able to conclude which Machine Learning algorithms work best. Random Forests, on a consistent basis, predicted at the highest accuracy, with Gradient Boosting and Bagging as runner ups. Technical indicators such as WAD, MFI, OBV, MA200, and MA50 were useful when making predictions for Bitcoin and when comparing our Bitcoin models to gold models. We found the models had similar results, as shown in Charts 6 and 7. To conclude, we found value in using data from Twitter, technical indicators, and macroeconomic indicators to predict price for Gold and Bitcoin. This highlighted how both sources of currency can be treated similarly when doing prediction, and how both can be used as hedges during inflation.

## **Limitations of the Study**

The primary limitation of the study is a bias in the Twitter Sentiment modeling. Because Twitter is used by a younger crowd and messages can often be cryptic or misinterpreted by a machine, it can be difficult to capture the real meaning of a Tweet. In addition, the ML models on predicting Bitcoin prices take into account only technical and economic indicators, they do not take into account the increasing adoption of Bitcoin within the broader economy and whether this has an impact on its price.

## Future Work

Given additional time, we would first try to reduce the number of features to simplify the model. Secondly, we would like to hypertune the parameters to see if any adjustments affect model performance, such as adding additional time data, trying out more powerful models such as recurrent neural networks, and adding a validation set to improve accuracy.

## Scholarly Articles

Basher, Syed Abul and Sadorsky, Perry, Forecasting Bitcoin Price Direction With Random Forests: How Important Are Interest Rates, Inflation, and Market Volatility? (June 6, 2022). Available at SSRN: <https://ssrn.com/abstract=4128509> or <http://dx.doi.org/10.2139/ssrn.4128509>