A Novel Crypto Trading Strategy with Sentimental Analysis

Rakeen Rouf, Tianji Rao, Mike Lund, Ching-Lung Hsu

Duke University

Math 585: Introduction to Algorithmic Trading

Prof. David Ye

April 29, 2024

Contents

1 Introduction 3

2 Literature Review 5

3 Data Descriptions 7

4 Sentiment Analysis 9

4.1 News Data 9

4.2 Twitter 11

4.3 Google Trend 12

5 Forecasting Model 13

5.1 Features and Target 13

5.2 Models: 16

5.3 Forecasting Results 19

6 Trading Strategy 22

6.1 Strategy Construction & Risk Management 22

6.2 Slippage Modeling 23

6.3 The Trading Algorithm 24

7 Backtest Results 25

8 Conclusion 31

9 Acknowledgements 32

10 Reference 32

Abstract

Cryptocurrencies are increasingly recognized as a compelling asset class, marked by swiftly growing market capitalization and low barriers to entry for trading. However, volatile and speculative cryptocurrency prices pose significant challenges for return prediction using traditional financial approaches. Our team undertook a novel investigation in this study to delineate the complex relationship between cryptocurrency returns and human sentiment. We developed a supervised machine learning framework that merges fundamental and technical indicators with cutting-edge sentiment analysis from various data streams. Our methodology incorporated a baseline linear regression model, further enhanced by a Random Forest model. These models underwent rigorous testing to confirm their robustness and reliability in making predictions. Our findings represent a substantial advancement in predictive accuracy and model interpretability, offering crucial insights for investment strategies within the cryptocurrency domain.

1. Introduction

Cryptocurrency is a digital asset that uses cryptography to safeguard the procedures involved in transactions and the creation of new units. Bitcoin, the world's inaugural decentralized cryptocurrency, was established in 2009. Fast forward fifteen years, and the landscape has expanded to include over 14,000 cryptocurrencies and tokens traded across more than 1,065 exchanges worldwide, amassing a total market capitalization of $2.48 trillion[1]. Entering the cryptocurrency market is relatively easy, akin to trading stocks, and doesn't require costly data setups for retail and institutional investors to engage actively. The varied trading environment presents an excellent opportunity to blend traditional financial portfolio strategies with innovative feature engineering, potentially discovering new alpha strategies.

One defining feature of cryptocurrencies is their inherent volatility. This instability is primarily due to the lack of fundamental backing for crypto assets as opposed to traditional financial assets like stocks or bonds. As a result, the market is often filled with speculators who are more interested in trading than investing for long-term value. The inherent volatility driven by people makes cryptocurrencies a high-risk investment (P Katsiampa (2019)). A naive buy-and-hold strategy potentially causes significant drawdown and losses. Therefore, a natural question is whether the returns of Bitcoin are predictable based on human sentiment. Accurate prediction allows investors to better capture timely speculative volatility in advance, effectively making a profit and mitigating potential loss.

[1] https://www.coingecko.com/en/global-charts

This report proposes a novel machine-learning model to predict cryptocurrency's future return better while uncovering the intricate interplay between cryptocurrency returns and human sentiment. Our work can be decomposed as follows: 1) Sentimental Analysis, 2) Predictive Modeling, and 3) Effective Trading Strategies. We first collect various news from Tingo News API and feed them as input to our sentiment model, Bidirectional Encoder Representations from Transformers (BERT), a powerful pre-trained large language model widely used in natural language processing (NLP) tasks. We integrated the output sentimental scores with other fundamental, technical indicators and Google Trend information as input predictors for our predictive models. We then implemented linear regression and random forest as the predictive model and chose the latter for its better predictive power. Our model refines prediction capabilities for the 15-day moving average daily return of Bitcoin, considering its inherent volatility and market complexity. We then design our trading strategy by considering the magnitude of sentimental scores and their transformation as a probability of success. Such transformation allows us to implement a modified Kelly criterion and increase profitability in the long run.

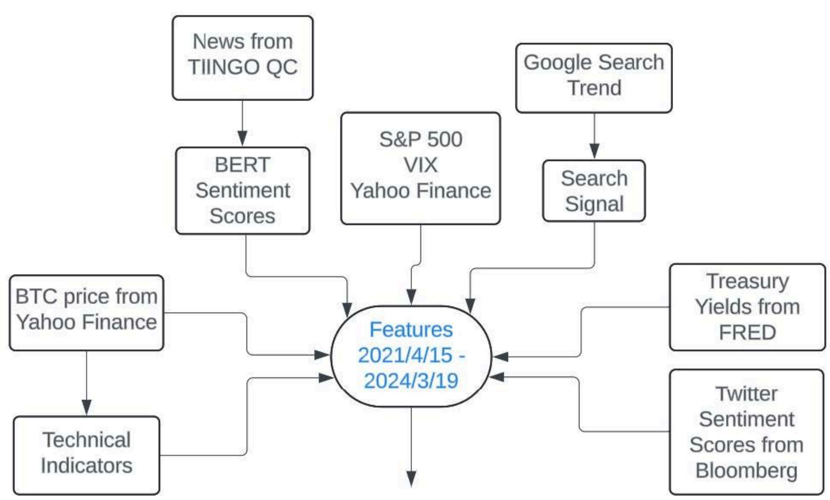


Figure 1: Features Construction

The remainder of this paper is organized as follows. Section 2 provides a literature review for models in sentimental analysis. We describe the exploratory data analysis in Section 3. In Section 4, we provide details about the implementation of BERT. We then present the predictive model and results in Section 5. In Section 6, we present the trading strategy and the risk management. In section 7, we show the backtest results, demonstrating the validity of our strategies. Finally, we conclude our work in Section 8, along with proposed future work.

2. Literature Review

Based on the development of NLP techniques, we can evaluate the sentiment from textual information and get a more accurate measure of investor sentiment. Bollen et al. (2011) demonstrated that Twitter's mood could predict the stock market, highlighting the potential of leveraging social media data for financial forecasts. According to Bellintani and Sophistor (2019), news contributes to market sentiment. The research aims to find the link between the price of a stock and the news. They have chosen a regression model with eleven independent variables and one dependent variable. Additionally, the study by Yadav et al. (2023) explored the unsupervised sentiment analysis of financial news. This approach mitigates some challenges associated with the labor-intensive process of labeling data in supervised learning models. For sentiment analysis, LLMs have been proven to take higher efficiency positions. Bozanta et al. (2021) employed textual data from another online platform - StockTwits, and further conducted sentiment analysis on posts for each company. Among BERT, DistillBERT, RoBERTa, and XLNet, RoBERTa outperformed traditional classifiers and deep learning algorithms regarding average F1 scores on classifying whether those tweets are “bearish” or “bullish.” In the Method section, we further introduced the usage and advantages of these models. However, among these researches, we seldom find some applications with LLM fine-tuned for finance. Lacking this step may cause bias when applying it to Financial text. In this research, FinBert, an LLM fine-tuned on financial texts, is a more field-specific method for accurate sentiment scoring on both Finance News and Twitter.

Recent studies have indicated the utility of machine learning approaches in predicting the return of Bitcoins. Ider and Lessmann (2022) implement Bidirectional Encoder Representations from Transformers (BERT) to predict cryptocurrency returns. Nasir et al. (2019) and Arratia et al. (2021) underscore the predictive power of search engine trends like Google Trends for Bitcoin. Similarly, Critien, Gatt, and Ellul (2022) and Peter Gabrovse et al. (2017) demonstrated the impact of Twitter sentiment on Bitcoin prices, affirming the relevance of social media sentiments in our analysis. Jing-Zhi Huang, William Huang, and Jun Ni (2019) advocate the inclusion of high-dimensional technical indicators in Bitcoin return predictions, enhancing our model's diversity. N. Y. Vo and G. Xu's（2017) research provides insights into Bitcoin's volatility and its relationship with the broader financial markets, guiding our integration of market indicators. Our model employs a Random Forest approach. It stands out for its greater transparency compared to those discussed above, making it preferable for investors seeking clarity and performance in their investment strategies.

3. Data Descriptions

We sourced Bitcoin price data from QuantConnect, a cloud-based algorithmic trading platform that enables users to design, backtest, and deploy trading strategies. QuantConnect collects Bitcoin data from various sources by integrating directly with multiple cryptocurrency exchanges, aggregating and normalizing this data for consistency. The calculation of returns utilizes the daily closing price, focusing on log returns to measure the rate of price change more accurately. Furthermore, according to the following figure, we can observe that bitcoin returns resemble a normal distribution with long tails, indicating some days of extreme returns.

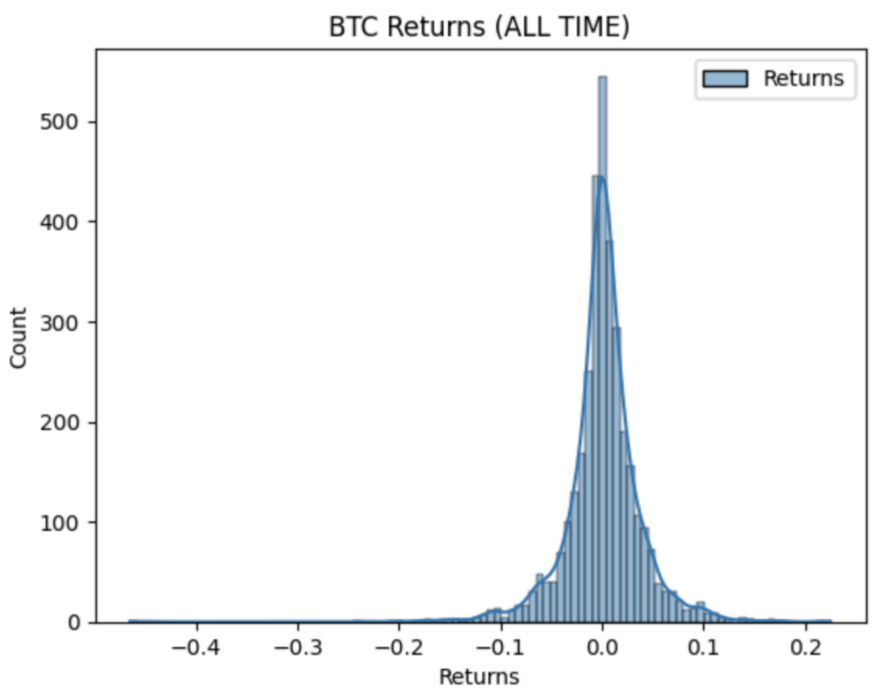


Figure 2: Distribution of all-time Bitcoin log returns

In our study, we've opted to forecast the 15-day moving average daily price return (15DAR) of Bitcoin and project its value to price movement direction and magnitude. This approach helps mitigate the volatility and noise in daily price movements, allowing us to capture more significant trends. It also enhances predictive power by smoothing out short-term fluctuations and revealing more meaningful patterns. Moreover, averaging returns over a longer period provides stability and reliability, which is crucial for risk management and decision-making. By focusing on the 15DAR, we improve the signal-to-noise ratio in our models, extracting more relevant information for investors with medium to long-term horizons. The figures below visualize the reduction in noise due to this modeling consideration.

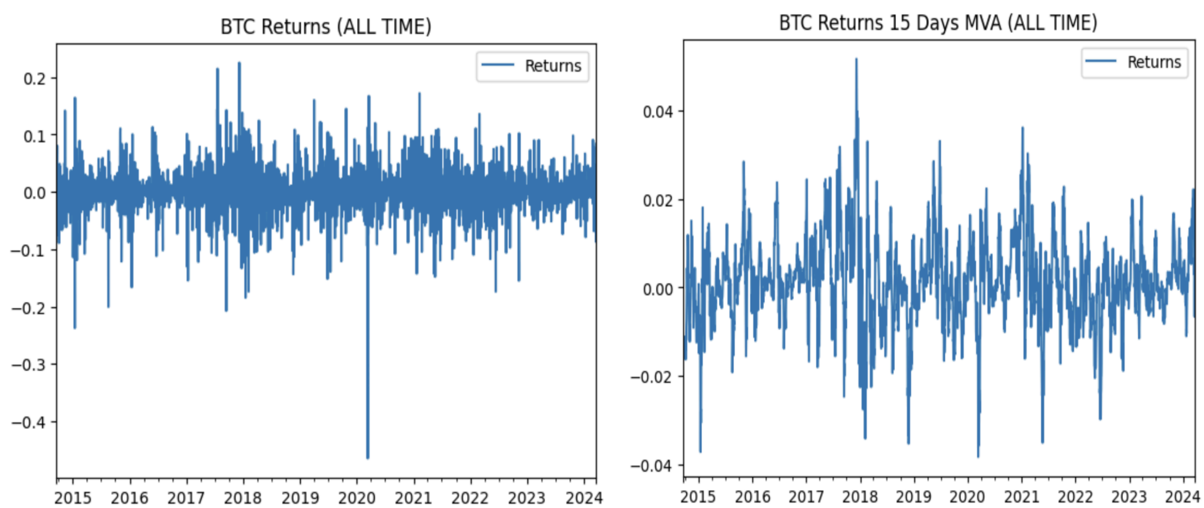


Figure 3: BTC Returns

The left panel is the daily Bitcoin returns (smaller signal-to-noise ratio), and the right panel represents 15 Day Moving Average (15DAR) Bitcoin returns (higher signal-to-noise ratio)

4. Sentimental Analysis

4.1 News Data

We sourced news data from TIINGO News API via the QuantConnect platform. Since TIINGO only stores news information regarding US Equities, Coinbase (the first publicly traded cryptocurrency exchange) was used to retrieve descriptions of all relevant news articles. Figure 4.1 shows the count of news articles retrieved that were relevant to Coinbase vs. time. From these articles, articles that contained any reference to Bitcoin/BTC (case-insensitive) were filtered to form the Bitcoin-exclusive news shown in Figure 4.2.

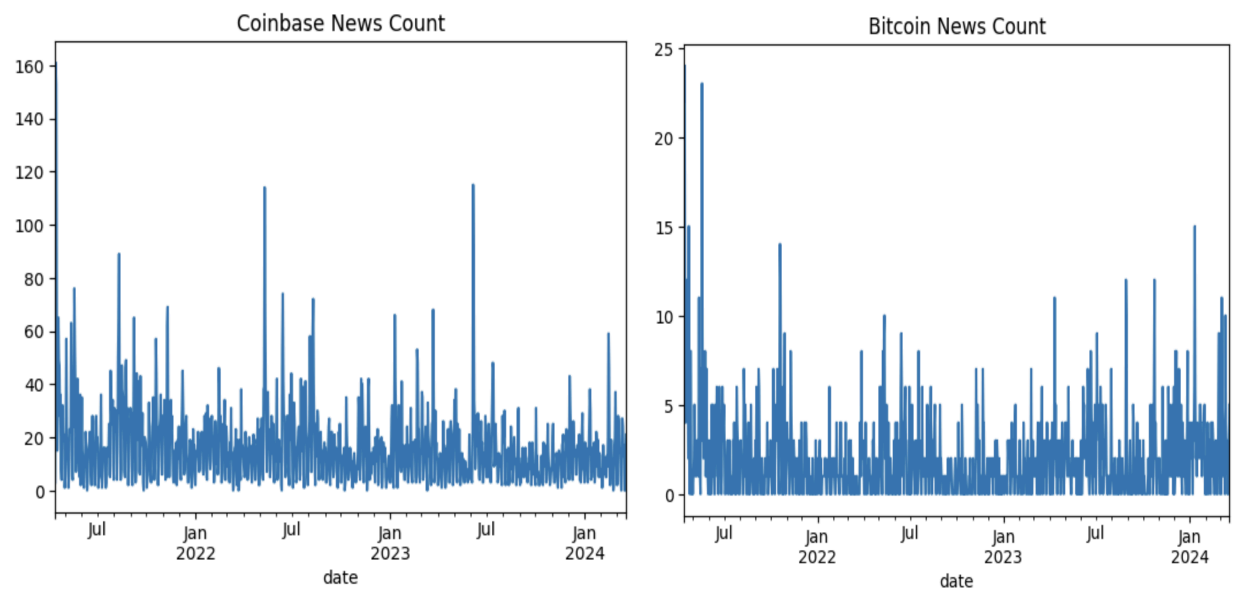


Figure 4: The left panel is the count of daily news containing “Coinbase,” and the right panel is the daily news containing “Bitcoin.”

Each news article (truncated to 512 tokens) was then passed through a fine-tuned version (on financial news data) of the DistilRoberta [11] to produce our news sentiment features. Figure 4 below shows the distribution of the resulting sentiment assignments.

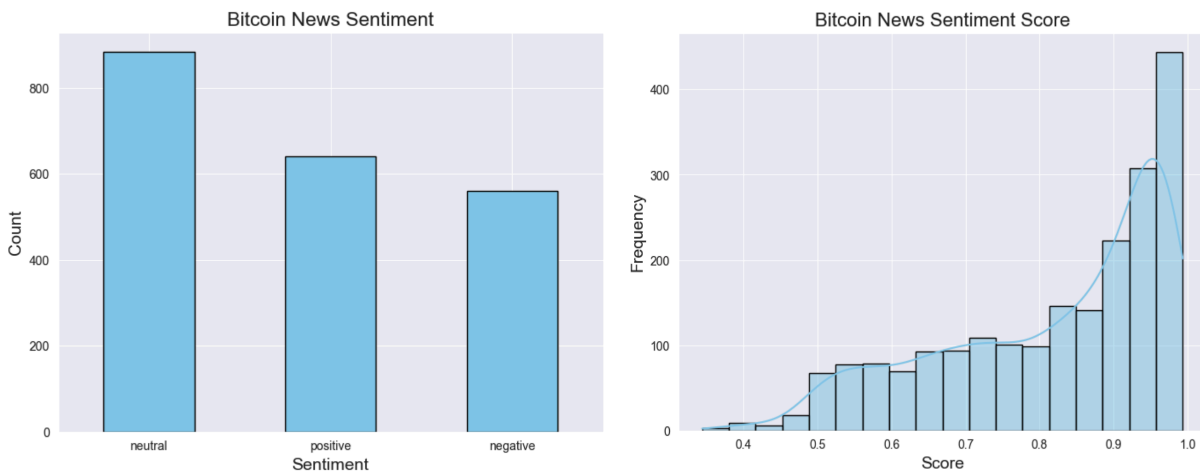


Figure 5: Distribution of generated sentiment labels and confidence scores using financial news fine-tuned version of the BERT Large Language Model

We then aggregated news sentiment data on a daily level. Some new features generated at this stage include "total_news_scores" and "signal (bitcoin_signal)." We calculated the Total News Score as a weighted sum of the average sentiment scores for positive and negative news. Total News Score measures the overall daily sentiment from the news. The Bitcoin signal is calculated similarly to the Google Trends signal below. Figure 6 below shows the distribution of the resulting daily features.

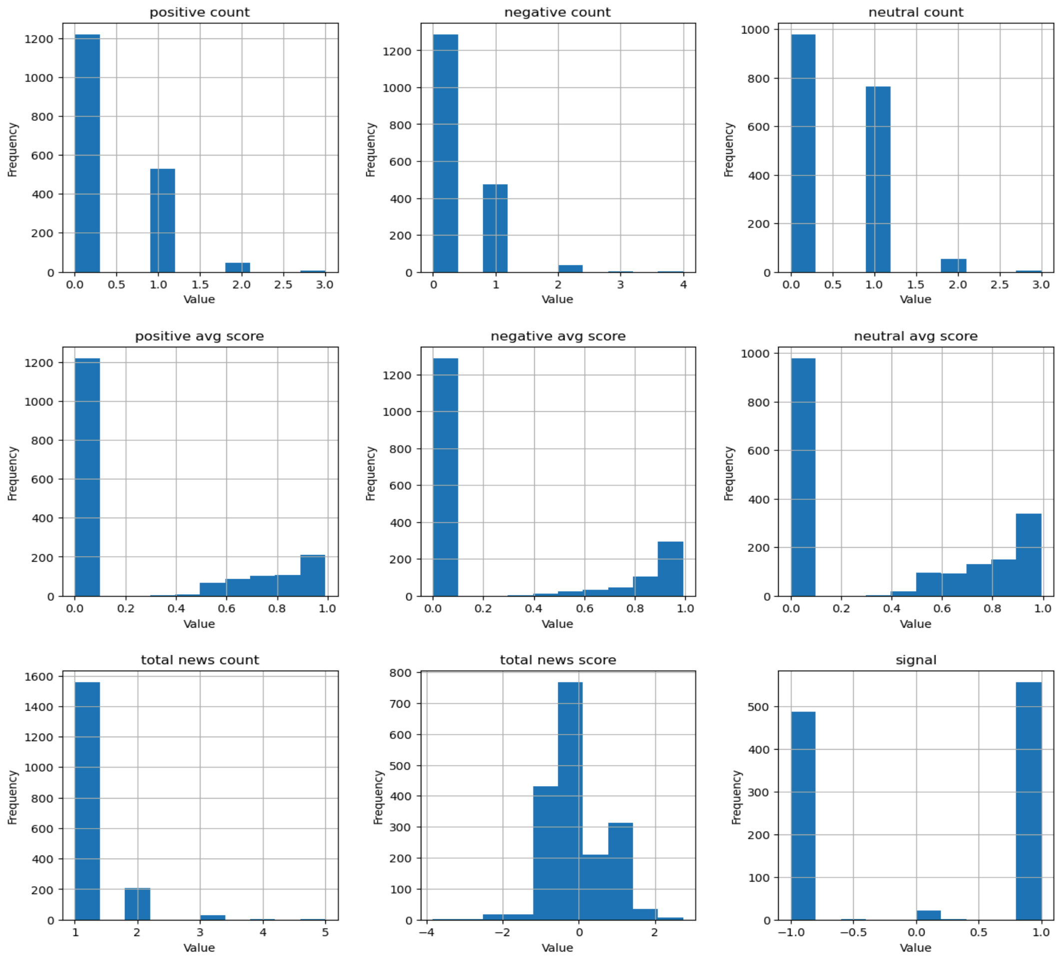


Figure 6: Distributions of all news sentiment features generated using a financial news fine-tuned version of the BERT Large Language Model [11].

4.2 Twitter

We obtained Twitter sentiment data from the Bloomberg terminal. Using a proprietary model, Bloomberg derived its own version of Total News Score, called News Sentiment Daily Average. Despite the differences in methodology, the distribution of the Bloomberg score closely resembles the distribution of the news sentiment score obtained from Tiingo. This similarity in distribution shapes serves as validation, indicating that proper data processing techniques have been employed across both methods. Figure 7 below shows the distributions of the Twitter sentiment features. The News Sentiment Daily Average, sourced from Bloomberg, and the Tiingo Total News Score, derived from BERT, exhibit similar distributions in their overall sentiment scores.

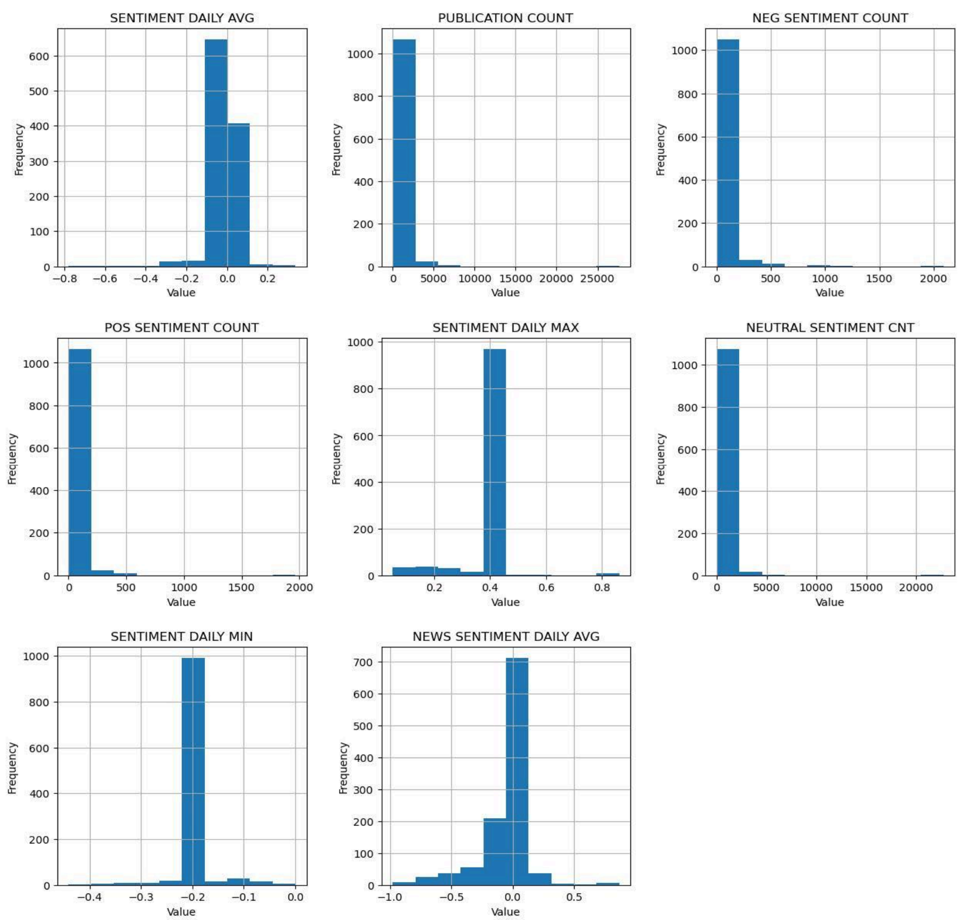


Figure 7: Distributions of all Twitter sentiment features extracted from the Bloomberg Terminal.

4.3 Google Trends

Google Trends is a website by Google that analyzes the popularity of top search queries in Google Search across various regions and languages. To build our Google Trends signal, we query the search volume for eight positive and eight negative search keywords listed in Table 1.

However, the raw keyword-level data spreadsheets from Google Trends need some preprocessing. We download the data at the weekly granularity and interpolate the data at a daily level.

Table 1: Positive and Negative Search Keywords

Positive	Negative
“bitcoin price”, “buy bitcoin”, “bitcoin”, “coin-base”, “bitcoin boom”, “make bitcoin”, “bitmex”, “long bitcoin”	“bitcoin crash”, “bitcoin loophole”, “sell bitcoin”, “bitcoin scam”, “short bitcoin”, “bitcoin bubble”, “bitcoin illegal”, “bitcoin bear”

Unlike stocks and other instruments in the financial market, crypto does not have fundamental variables like revenue, earnings, and so on, so investor sentiment may play a more important role in explaining the dynamic. Here, we highlighted how to develop a list of proxies for investor sentiment and use them as inputs for constructing forecasting models.

5. Forecasting Model

We further constructed forecasting models for the future Bitcoin movement using sentiment scores generated from the news. We transform the predicted 15DAR into daily price change direction and magnitude movement to compare the accuracy with existing research.

5.1 Features and Target

Our dataset was split into two parts: training data from April 15, 2021, to March 1, 2023, and testing data from March 1, 2023, to March 19, 2024. We meticulously ensured that no future information influenced the model by lagging all features by one day, ensuring predictions for any given day were based solely on data available up to the end of the previous day. Additionally, for rolling sums and averages, we dropped initial data points to ensure each calculation was based on a complete set of information, thereby preventing any potential data leakage and maintaining the integrity of our testing process.

We have developed four distinct feature sets for our experiments. The **first set** includes features commonly regarded as effective in market predictions alongside sentiment indicators. Examples are S&P 500 returns, counts and scores of positive and negative news, Google trends, and several technical indicators. This foundational set aligns with our study's objectives and helps prevent initial overfitting during model training.

The **second set** comprises all available features, including market prices, news scores, sentiment scores derived from the BERT model, technical indicators, and Google trend data. This comprehensive set allows us to assess model performance using all possible inputs and sets the stage for subsequent feature engineering to enhance model effectiveness.

The **third set** expands on the previous by incorporating memory features, which include rolling sums of all features over the past 10, 20, and 30 days. One of our key contributions is the development of memory features designed to address the limitations of lagged features used in literature, which only capture historical data as isolated snapshots. Our memory features accumulate data from the past days, effectively creating a continuous 'working memory' for our predictive model.

The **fourth set** consists of the top 30 significant features identified after applying a random forest model.

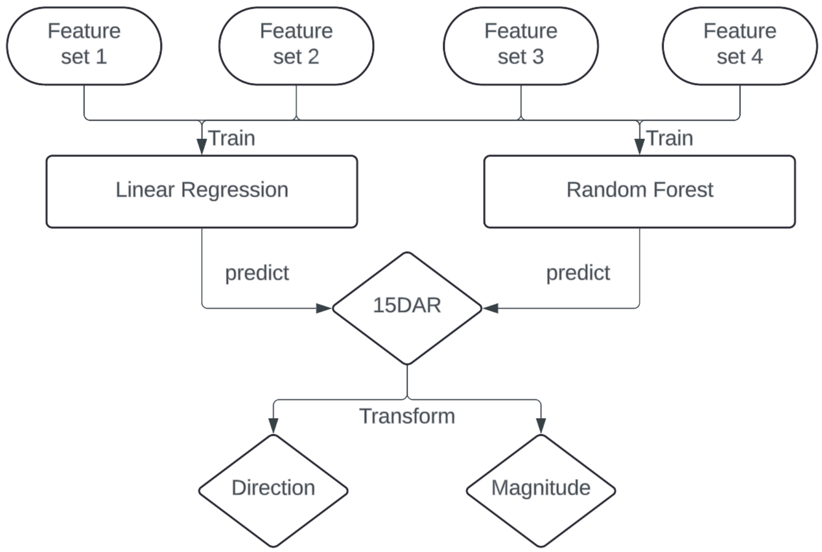


Figure 8: Experimentation Flow Chart

To compare the accuracy with existing research, we transform the predicted 15DAR into daily price change direction and magnitude movement. For direction, if the projected 15DAR is positive, we infer that the corresponding direction would also be positive (1) and conversely for negative predictions (0). For magnitude, we first multiply the predicted 15DAR by the previous day's price, resulting in the projected value of price movement, and then assign it to one of the classes below. Figures 8, 9, and Table 2 elaborate on the prediction workflow and Target construction.

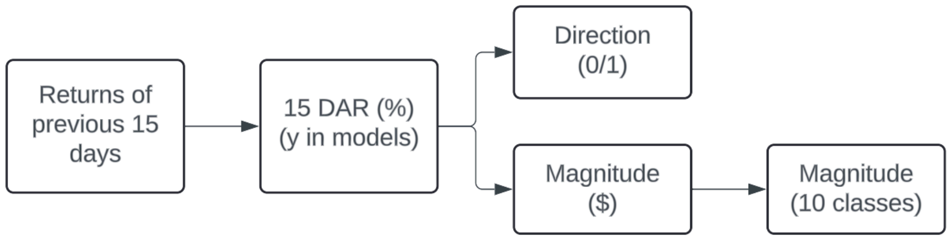


Figure 9: Target Variable Transformation

Table 2: Magnitude Bins for projected movement

Class	1	2	3	4	5	6	7	8	9	10
Range	Less than -$1320	-$1320 to -$990	-$990 to -$660	-$660 to -$330	-$330 to $0	$0 to $330	$330 to $660	$660 to $990	$990 to $1320	Greater than $1320

5.2 Models

We use Linear Regression as our initial model and the Random Forest Model as a more advanced option. Each model is fitted with a set of features. Linear regression is our first experimental model for the forecasting model due to its simplicity. Linear Regression takes less time to run and provides a baseline for understanding the influence of the independent variables on Bitcoin returns. Its straightforward interpretation and quick computation make it an ideal starting point for our experiments, allowing us to quickly establish a foundational understanding of the relationships within our data.

We applied the Random Forest model as an advanced machine-learning approach for Bitcoin return forecasting. This model, known for its robustness and ability to handle non-linear relationships, offers a more complex and nuanced analysis compared to Linear Regression. By aggregating decisions from many decision trees, Random Forests reduces the risk of overfitting and provides a more reliable prediction. This method is particularly effective due to its capability to handle large datasets with numerous features, making it highly suitable for the intricate dynamics of cryptocurrency markets.

To assess our model, we will evaluate its performance using several key metrics: R-squared, accuracy of direction prediction, and accuracy of magnitude prediction. R-squared will help us understand how much of the variance in our model's dependent variable can explain, providing a measure of how the model will likely predict well-unseen samples. The accuracy of direction prediction will assess whether our model can correctly predict the trend of the market movement—whether the price is going up or down. Finally, the accuracy of magnitude prediction will measure how well our model can estimate the size of the market changes. By analyzing these metrics, we can gain insights into our predictive approach's strengths and potential limitations, allowing for targeted improvements and refinement.

When configuring the Random Forest model, we select the top 30 most important features to form the fourth feature set. Feature Importance is shown in Figure 10. This approach is designed to retain the most predictive elements while excluding less relevant information, which helps prevent overfitting. This method ensures that our model remains robust and focused on the most influential variables affecting the outcome. Below is a graph illustrating the importance of features.

Figure 10 below provides valuable insight into a snippet of the decision-making process of our Random Forest model, particularly emphasizing the straightforward nature of its decision trees. Unlike the more common black-box models, such as Long Short-Term Memory (LSTM) found in much of the existing literature, our approach offers investors a clearer understanding of how decisions are made.

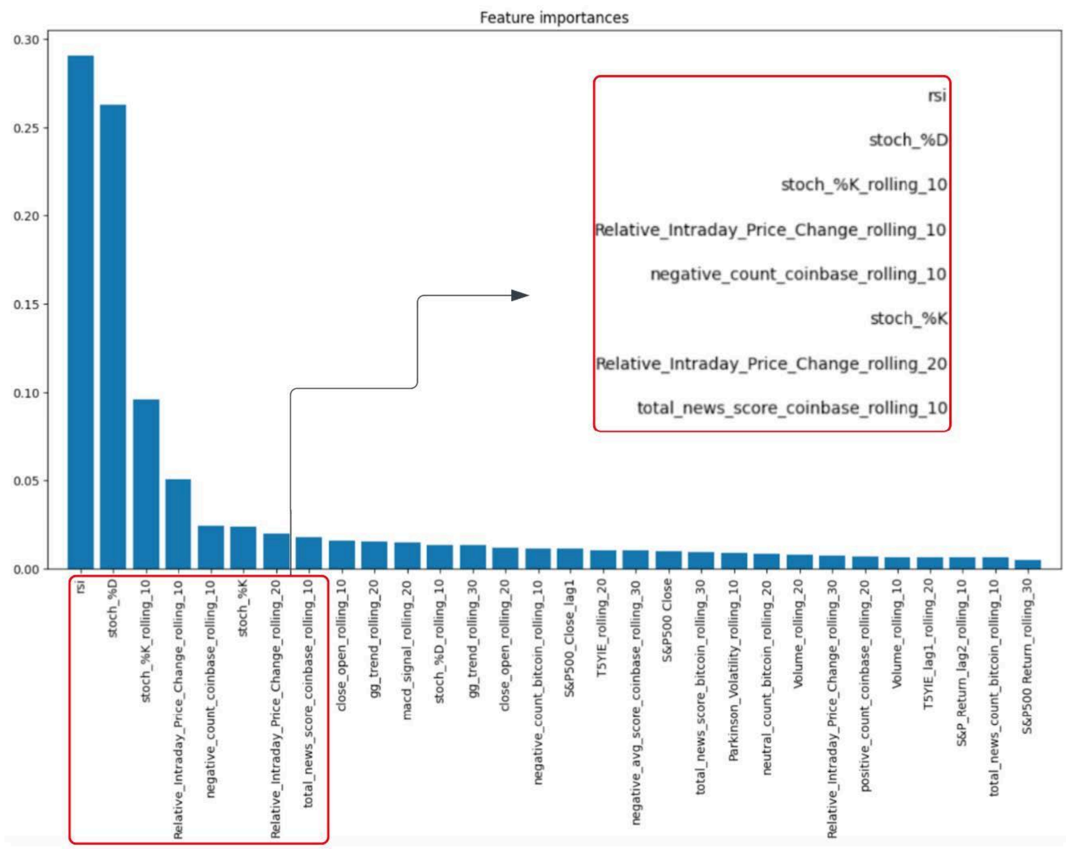


Figure 10: Feature importance of the final Refined Random Forest Model.

Despite Random Forests not typically being considered interpretable models, the split rules within the trees provide a window into the logic used, offering a tangible sense of how investment decisions are derived. This clarity is advantageous as it allows investors to see why their funds are allocated in certain ways, thus boosting their confidence and facilitating more informed decision-making. In our model, the primary splits focus on technical indicators, with sentiment features becoming more influential in subsequent layers of the decision process. This methodical unveiling of decision layers helps demystify the predictive process and sets our model apart by enhancing transparency and investor engagement.

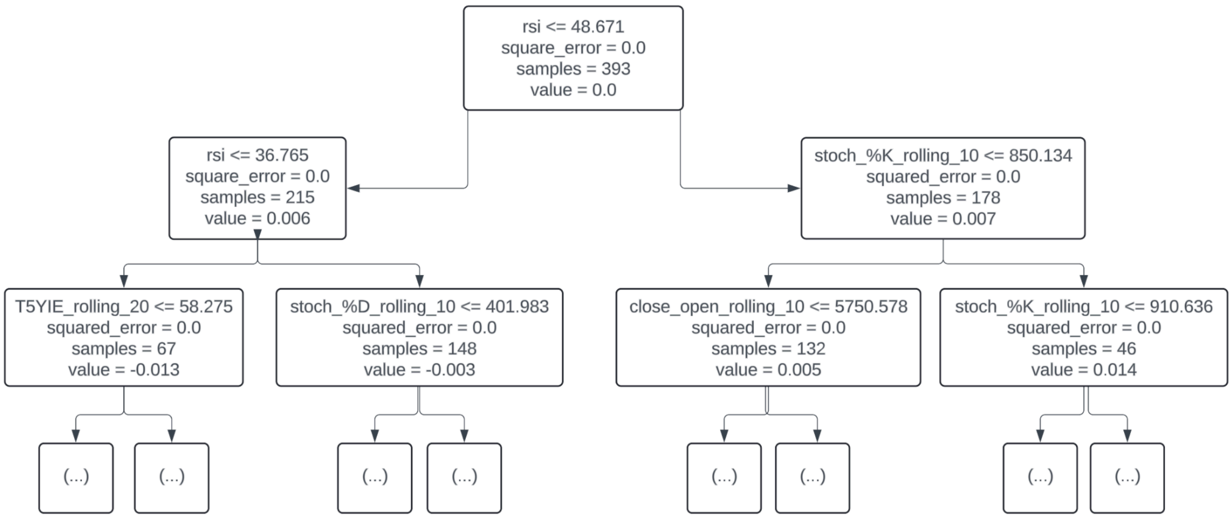


Figure 11: First two levels of a sample decision tree from Random Forest

5.3 Forecasting Results

To better compare with existing literature, we project the 15DAR to the direction and magnitude of price movement. Result tables for R-squared, direction accuracy, and magnitude accuracy are listed below.

Table 3: Forecasting Results for 15DAR R-square

15DAR R-square	Set 1	Set 2	Set 3	Set 4
Linear Regression	0.57	0.63	-41.82	0.59
Random Forest	0.59	0.67	0.73	0.73

Table 4: Forecasting Results for Direction Accuracy (random guess: 50%)

Acc Direction	Set 1	Set 2	Set 3	Set 4
Linear Regression	0.85	0.79	0.63	0.84
Random Forest	0.86	0.87	0.87	0.88

Table 5: Forecasting Results for Magnitude Accuracy (random guess: 10%)

Acc Magnitude	Set 1	Set 2	Set 3	Set 4
Linear Regression	0.21	0.22	0.13	0.22
Random Forest	0.22	0.22	0.21	0.22

As we can see from the result tables, with Feature Set 1, which includes basic market and sentiment features, both models show decent R-squared values, suggesting a reasonable fit to the data (actual 15DAR vs predicted 15DAR). However, Random Forest slightly edges out Linear Regression, and this trend persists across all feature sets, highlighting the robustness of the Random Forest model.

Feature Set 2, which uses all available features, shows improved R-squared values for both models, but notably, there's a dip in the direction prediction accuracy for Linear Regression. This may indicate that including more complex features without feature selection can introduce noise for simpler models. In contrast, the Random Forest model maintains high accuracy in direction prediction, demonstrating its capability to handle high-dimensional data without a significant loss of interpretability.

With Feature Set 3, where memory features are added, there is a significant anomaly in the R-squared for Linear Regression, suggesting that including these features could lead to model instability or inappropriate model fit in this instance. The Random Forest, however, shows a remarkable increase in R-squared, indicating its ability to utilize these temporal features effectively without overfitting.

Feature Set 4, which includes the top 30 features selected by the Random Forest, shows a convergence in performance for both models. This suggests that when the most predictive features are used, a simpler model like linear regression can achieve results comparable to those of a more complex model. This is particularly evident in the R-squared and magnitude prediction accuracy, although Random Forest maintains a slight advantage in direction prediction accuracy.

Table 6: Performance Comparison of Random Forest with Best from Literature Review

	Best from Literature Review	**Random Forest Model**
Daily Price Direction Forecast Accuracy	68.4%	88%
Magnitude Forecast F1 Score	14.21%	16.89%
Magnitude Forecast Accuracy	51.47%	22.25%

Table 6 compares the Random Forest performance with Features set 4 vs. the literature review. The Random Forest exhibits superior performance over the best model identified in the literature review in predicting the direction of daily price movements, achieving an accuracy of 88% compared to 68.4%. However, when forecasting the magnitude, the Random Forest displays a slightly lower accuracy rate of 22.25% versus 51.47% for the best model from the literature review. Despite this disparity, the RFM's higher F1 score suggests a better balance between precision and recall, indicating its potential to more effectively capture nuanced price movement patterns. The methods from the literature were not implemented on our dataset; however, the comparison remains meaningful as it provides a benchmark of model performance across different market conditions and asset classes. It's comparable because it allows us to evaluate the generalizability and robustness of our models against various approaches and data characteristics found in existing research. While the Random Forest may occasionally misclassify certain movements, it offers competitive performance and provides investors with valuable insights into Bitcoin price dynamics.

6. Trading Strategy

6.1 Strategy Construction and Risk Management

Utilizing our research results in the previous section, we tried to take advantage of the outstanding forecasting performance and further build a trading strategy. The strategy we used is threshold-based: Before the start of each day d, we will get the predicted returns of the 15-day moving average S_d in advance. We then decide whether to go for long, short, or nothing based on the value of S_d. If $S_d > 1\%$, we believe this prediction represents an upcoming Bullish market, while if $S_d < -0.5\%$, we interpret this as a signal of a Bearish market. Choosing a smaller value as a threshold for the negative predictive return allows us to be more careful about the potential downside. After deciding on the action (long, short, or do nothing), we convert S_d to $I_d = ABS(ImpFunc(S_d))$ such that I_d lies between 0 and 1. Our goal is to transform the predictive return S_d to the probability of the win rate for the action we decided on above. This transformation allows us to calibrate further how much we want to bet on the action with Kelly's criterion.

Kelly's criterion is a formula used to determine the optimal size of a series of bets to maximize the logarithm of wealth. Developed by John L. Kelly, Jr. in 1956, the criterion calculates the ideal fraction of a gambler's bankroll to wager, given the probability of winning p and the odds of the bet b. The mathematical formula for the Kelly criterion is: $kf_d = \frac{bp - (1-p)}{b}$, where kf_d represents the fraction of the bankroll to bet. This formula is widely used in gambling, investing, and risk management to balance the trade-off between maximizing returns and minimizing the risk of ruin. We set $p = max(I_d, 0.51)$ as the winning rate to calculate the best betting ratio kf_d with Kelly's criterion. Next, we execute the trading action with quantities proportional to kf_d for both Bitcoin and Ethereum. Lastly, we implement risk management to avoid potential loss. Either the portfolio value falls by 5% within a day or falls in nine consecutive days, we will liquidate our positions. One can refer to Algorithm 1 for more details.

6.2 Slippage Modeling

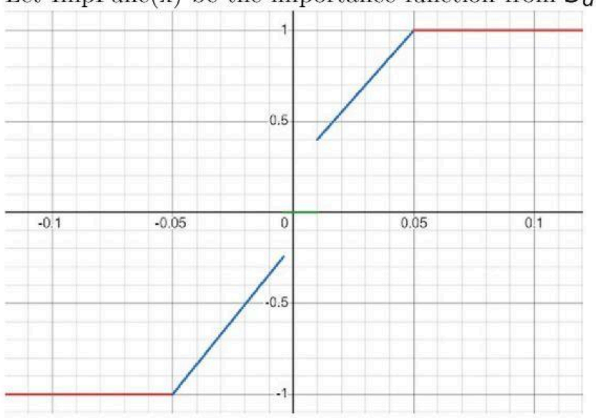
Slippage in QuantConnect simulates the difference between expected and actual execution prices of trades, providing more realistic backtesting and live trading results. More specifically, we consider Slippage = asset.price * 0.0001 * log(2 * order.absolute_quantity). This equation provides a reliable estimate for the potential slippage and makes the backtest more robust and accurate.

6.3 The Trading Algorithm

Algorithm 1 The Trading Strategy with Predicted Returns

Step 1: Making Trading Decisions
long = short = 0
for (each day $d > 0$) **do**
    Get tomorrow's predicted 15-day moving average S_d
    if $S_d > 1\%$ **then** ▷ Market is Bullish
        long = 1
        short = 0
    else if $S_d < -0.5\%$ **then** ▷ Market is Bearish
        long = 0
        short = 1
    else ▷ Market is Neutral
        long = 0
        short = 0

Step 2: Transform S_d to importance scores I_d
Let ImpFunc(x) be the importance function from S_d to $[-1, 1]$. Its graph is presented as follows:



Higher importance means more confidence.
I_d = ABS(ImpFunc(S_d)) ▷ $0 \leq I_d \leq 1$
if $I_d < 0.51$ **then**
    $I_d = 0.51$ # Kelly criterion doesn't make sense for smaller values

Step 3: Using Kelly Criterion to calculate the best Kelly fraction kf_d for optimal bet size.

Step 4: Trading Execution
Buy and Sell according to the signal and Kelly fraction kf_d
if long == 1 **then**
    Long Bitcoin $(0.8 * kf_d)$ of Portfolio Value
    Long Ethereum $(0.2 * kf_d)$ of Portfolio Value
else if short == 1 **then**
    Short Bitcoin $(0.5 * 0.8 * kf_d)$ of Portfolio Value
    Short Ethereum $(0.5 * 0.2 * kf_d)$ of Portfolio Value

Step 5: Risk Management
Liquidate portfolio if its value falls by more than 5% within a day.
Liquidate portfolio if its value falls for nine consecutive days.

7. Backtest Results

7.1 In-Sample (Training Period): 4/10/2022 — 3/31/2024

Return	34.53%
PSR	57.897%
Sharpe Ratio	1.15
Max Drawdown	13.8%
Compounding Annual Return	35.447%
Volume	$ 81,466,447
Net Profit	$ 3,453,094

Backtest link:

https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_7bcfb7777a07fd6aa450c2084f99c698.html

Remark: The alpha model used was trained on data from this period. Our good performance in the in-sample period serves as internal validity of our strategy and modeling.

7.2 Out-of-Sample A (OOS A): 4/10/2021 — 8/10/2021

Return	23.11%
PSR	78.624%
Sharpe Ratio	3.206
Max Drawdown	6.7%
Compounding Annual Return	132.776%
Volume	$84,828,989
Net Profit	$ 2,311,482

Backtest link:
https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_1e0d88c786d526e00e92fdfc1b924264.html

Remark: (Please look at remark of OOS A + B below)

7.3 Out-of-Sample B (OOS B): 9/10/2021 — 12/10/2021

Return	11.74%
PSR	58.926%
Sharpe Ratio	1.61
Max Drawdown	7.6%
Compounding Annual Return	54.606%
Volume	$37,993,254
Net Profit	$ 1,171,914

Backtest link:

https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_888b52eb63f9c41b6f4e39819de027fa.html

Remark: (Please look at remark of OOS A + B below)

7.4-1 Out-of-Sample A+B (Our Model): 4/10/2021 — 12/10/2021

Backtest link

7.4-2 Out-of-Sample A+B (Buy & Hold Benchmark): 4/10/2021 — 12/10/2021

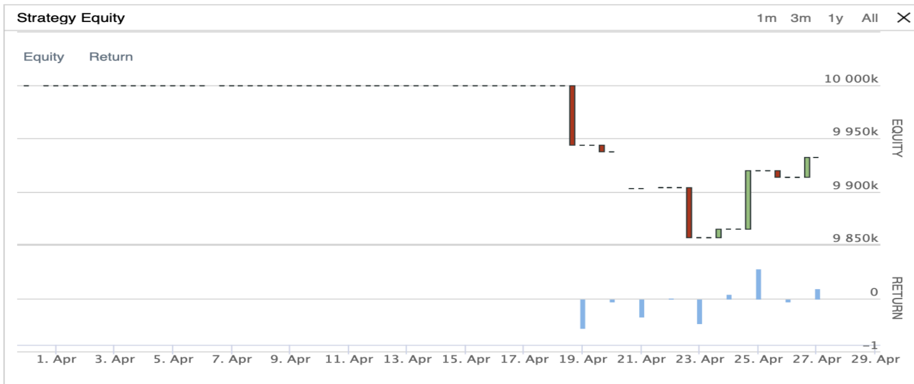
To establish a benchmark, we employ a buy-and-hold strategy that mirrors our portfolio composition throughout the entire trading period. This portfolio consists of 80% Bitcoin and 20% Ethereum, reflecting the cryptocurrency blend proportions utilized in our strategy.

[Backtest link]

Strategy	Return	PSR	Sharpe Ratio	Max Drawdown	Compounding Annual Return	Total Order
Buy & Hold	-5.24%	24.41%	0.256	43.4%	-8.85%	4
Our Model	**41.21%**	**72.1%**	**2.056**	**10.1%**	**81.23%**	**197**

Remark: Our model outperforms the buy-and-hold strategy in all the metrics, gaining more than 2 for the Sharpe ratio with a 41% Return and only a 10% potential drawdown. This shows a substantial gain over the passive benchmark strategy.

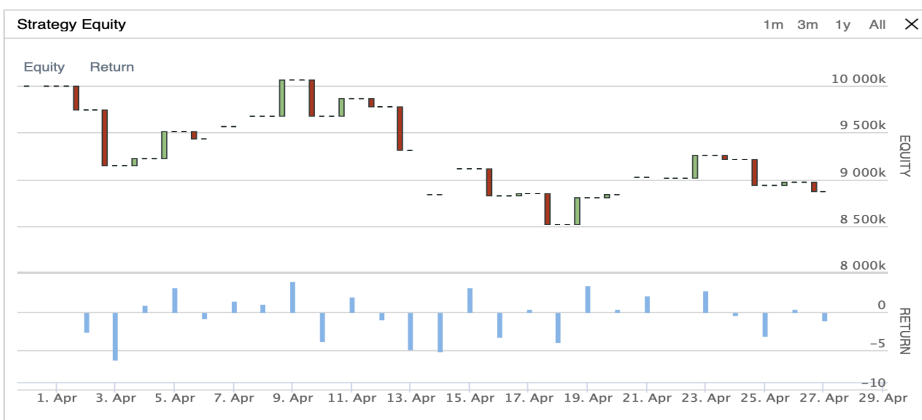
7.5-1 Out-of-Sample C, Blind (Our Model): 3/31/2021 — 4/27/2021



Backtest link

7.5-2 Out-of-Sample C (Buy & Hold Benchmark): 3/31/2021 — 4/27/2021

The same buy and hold strategy as above is followed here as well.



Backtest link

Strategy	Return	PSR	Sharpe Ratio	Max Drawdown	Compounding Annual Return	Total Order
Buy & Hold	-10.96%	14.36%	**-1.514**	15.4%	-76.99%	4
Our Model	**-0.68%**	**16.10%**	-3.714	**1.4%**	**-8.25%**	**6**

Remark: Our model outperforms the buy-and-hold strategy in all metrics except for the Sharpe Ratio due to the limited trading periods of OOS C. Our model did not get strong enough signals to trade in about two-thirds of the roughly month-long period. As a result, our portfolio was only active for approximately 10 days. Since the standard deviation of returns from 10 days is typically very high, this affects the Sharpe ratio. Therefore, for a medium to long term strategy like ours, we would need a portfolio that is active for much longer to get a useful sharpe ratio.

7.6-1 Stress Test (Our Model): 12/1/2021 — 7/1/2022

[Backtest link]

7.6-2 Stress Test (Buy & Hold Benchmark): 12/1/2021 — 7/1/2022

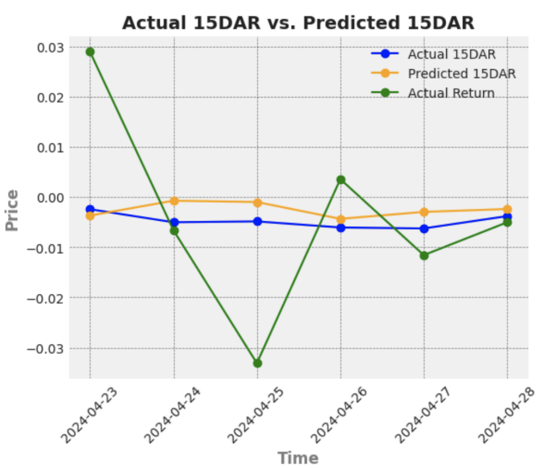
The same buy and hold strategy as above is followed here as well.

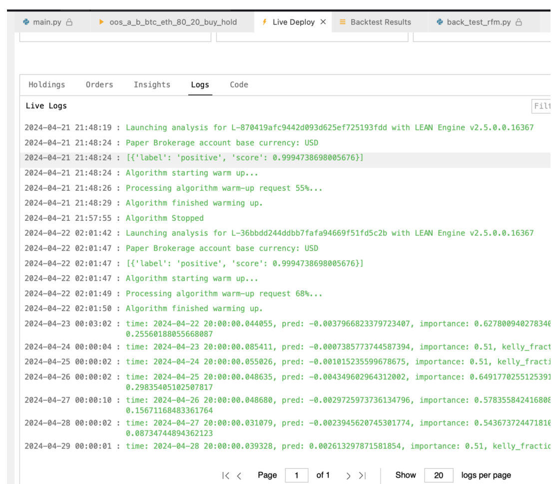
[Backtest link]

Strategy	Return	PSR	Sharpe Ratio	Max Drawdown	Compounding Annual Return	Total Order
Buy & Hold	-68.305%	0.261%	-1.194	69.0%	-85.938%	4
Our Model	**5.42%**	**30.65%**	**0.431**	**14.7%**	**9.41%**	**167**

Remark: Our model again outperforms the buy-and-hold strategy in all the metrics, gaining more than a positive Sharpe ratio (0.34) with a 5.42% Return and only a 14.7% potential drawdown. This result demonstrates our strategy could effectively avoid potential drawdown risk in the Bearish market conditions.

7.7 Live Trading: 4/23/2024 — 4/29/2024





Remark: The 15-day moving average return (15DAR) direction and value were predicted closely on all days during the live trading so far. Furthermore, we also predicted 4 out of 6 daily return directions correctly. Since the predicted 15DAR value is near zero, we have not executed any action yet, indicating a neutral crypto market (Remember, we only trade at predicted 15DAR < -05% & 15DAR > 1%).

8. Conclusion

This research introduces a pioneering approach to predicting cryptocurrency trends, emphasizing the intricate interplay between human sentiment and sophisticated machine-learning techniques. The combination of Bitcoin historical data, technical and fundamental indicators, and sentiment analysis, particularly from social media and financial news, has demonstrated a marked improvement in predicting crypto market movements.

We achieved a more nuanced understanding of market sentiments by incorporating advanced BERT classifiers for sentiment analysis over the traditional VADAR [10] model, a lexicon and rule-based sentiment analysis tool. This enhancement was crucial in improving the model's accuracy and interpretability.

The Random Forest model's superiority over other tested models was evident in its higher predictive accuracy and lower error rates. Employing this model, we developed a demonstrative trading strategy based on the model's signals, which successfully outperformed a simple buy-and-hold strategy (Section 7.5, OOS A+B). These results underscore the practical applicability of our model, showcasing its ability to offer clear, actionable insights to investors while balancing risk and return efficiently.

In conclusion, our study contributes a novel, robust model for Bitcoin return predictions, paving the way for future explorations in this intriguing and challenging arena of financial forecasting. By achieving a balance between advanced machine learning techniques and the critical need for investor-friendly model transparency and interpretability, we hope our work will inspire and inform subsequent efforts, advancing our collective understanding of the intersection between human sentiment and cryptocurrency market dynamics.

9. Acknowledgments

We are immensely grateful to Professor David Ye and TA Huanming Zhang for their invaluable assistance, continuous feedback, support, and the resources they provided throughout our project. Their guidance was instrumental in our success.

Additionally, we extend our thanks to all the guest speakers from diverse backgrounds who shared their extensive expertise in quantitative finance and technology. Their contributions enriched our learning experience significantly. We would also like to express our gratitude to the QuantConnect Team for their timely response and for helping.

Lastly, we are profoundly thankful for an extraordinary and insightful class. The stimulating and inspiring atmosphere fostered by all teams allowed everyone to excel and produce remarkable work.

10. Reference

1. Arratia, A., López-Barrantes, A.X.(2021). Do Google Trends forecast bitcoins? Stylized facts and statistical evidence. J BANK FINANC TECHNOL 5, 45–57 (2021). https://doi.org/10.1007/s42786-021-00027-4
2. Bellintani, L., & Sophistor, J. (2019). The Effect of News on Intra-Day Stock Prices & their Volatility.
3. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.
4. Couzin-Frankel, J. (2019). Questions churn about vaping's long-term risks. Science, 366(6469), 1059–1060. https://doi.org/10.1126/science.366.6469.1059
5. Critien, J.V., Gatt, A. & Ellul, J.(2022). Bitcoin price change and trend prediction through twitter sentiment and data volume. Financ Innov 8, 45 (2022). https://doi.org/10.1186/s40854-022-00352-7
6. Denning, P., & Little, J. (2017). Over the influence: The harm reduction guide to controlling your drug and alcohol use (2nd ed.). Guilford Press.
7. Devlin, Chang, Lee, & Toutanova (2018), Bidirectional Encoder Representations from Transformers. https://arxiv.org/abs/1810.04805
8. Duygu Ider, Stefan Lessmann (2022). Forecasting Cryptocurrency Returns from Sentiment Signals: An Analysis of BERT Classifiers and Weak Supervision. https://arxiv.org/abs/2204.05781
9. Earp, J. (Director) & Katz, J. (Presenter) (2013). Tough guise 2: Violence, manhood & American culture [Film; educational DVD]. Media Education Foundation.

10. Gabrovsek, Peter; Aleksovski, Darko; Mozetic, Igor; Grcar, Miha (2017). Twitter Sentiment around the Earnings Announcement Events. https://arxiv.org/abs/1611.02090
11. Howard, J. (2019, December 16). Vaping linked with long-term risk of respiratory disease in new study. CNN. https://www.cnn.com/2019/12/16/health/vaping-respiratory-disease-study/index.html
12. Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).
13. Katsiampa, P. (2019), An empirical investigation of volatility dynamics in the cryptocurrency market. https://www.sciencedirect.com/science/article/abs/pii/S0275531919300637
14. Karthikeyan, D., Sivamani, B. A., Tummala, P. K., & Arumugam, C. (2021). Time Series for Forecasting Stock Market Prices Based on Sentiment Analysis of Social Media. International Conference on Computational Science and Its Applications (pp. 353-367). Cham: Springer International Publishing.
15. Lim, F., Paguirigan, M., & Cernivani, D. (2018). Delivering LGBTQ-sensitive care. Nursing Critical Care, 13(4), 14–19. https://doi.org/10.1097/01.CCN.0000534918.70677.9c
16. McKee, A. (2015, May 15). Introduction to criminal justice, section 1.2 [Video]. YouTube. https://youtu.be/ZXnp8OQwgH8
17. Merriam-Webster. (n.d.). Insanity. In Merriam-Webster.com dictionary. Retrieved January 9, 2020, from https://www.merriam-webster.com/dictionary/insanity

18. Nasir, M.A., Huynh, T.L.D., Nguyen, S.P. et al.(2019) Forecasting cryptocurrency returns and volume using search engines. Financ Innov 5, 2 (2019). https://doi.org/10.1186/s40854-018-0119-8
19. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System; 2008. https://bitcoin.org/bitcoin.pdf.
20. Weissinger, S. E., & Mack, D. (Eds.). (2018). Law enforcement in the age of Black Lives Matter: Policing black and brown bodies. Lexington Books.
21. World Health Organization. (2018, December). Gaming disorder. https://www.who.int/news-room/q-a-detail/gaming-disorder
22. Yadav, A., & Kumar, S. (2023). Unsupervised sentiment analysis of financial news. Journal of Artificial Intelligence Research, 71, 45-59.