Trading Crypto with FinBERT & Random Forest

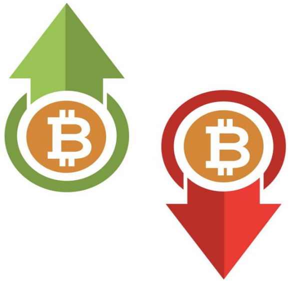
Group 4

Rakeen Rouf, Tianji Rao,

Ching-Lung Hsu, Mike Lund

Objective

1. Predict daily Bitcoin returns with a special focus on Human Sentiment.
2. Beat best performance levels identified in existing literature.
   -
   -



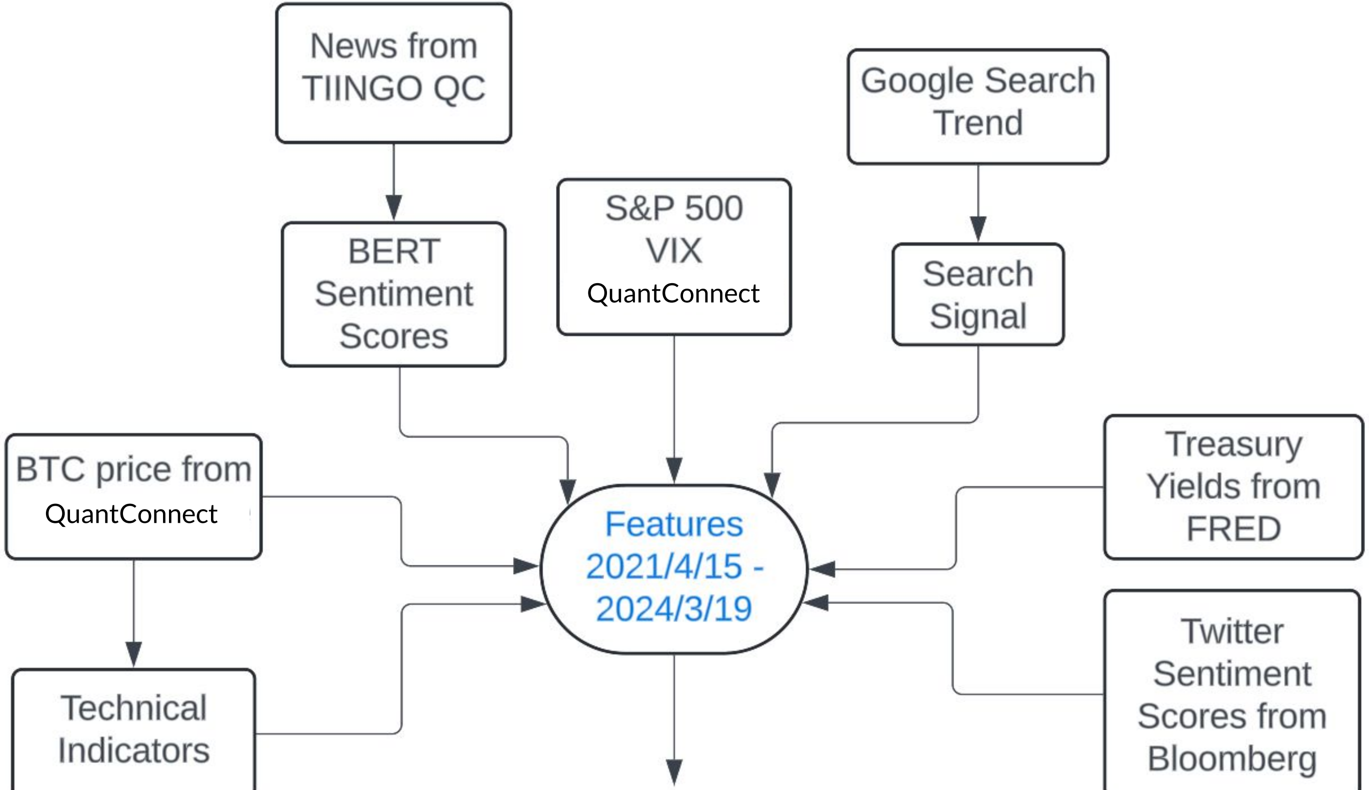
Class	1	2	3	4	5	6	7	8	9	10
Range	Less than -$1320	-$1320 to -$990	-$990 to -$660	-$660 to -$330	-$330 to $0	$0 to $330	$330 to $660	$660 to $990	$990 to $1320	Greater than $1320

3. Use prediction as Alpha signal for a cryptocurrency trading strategy

Agenda

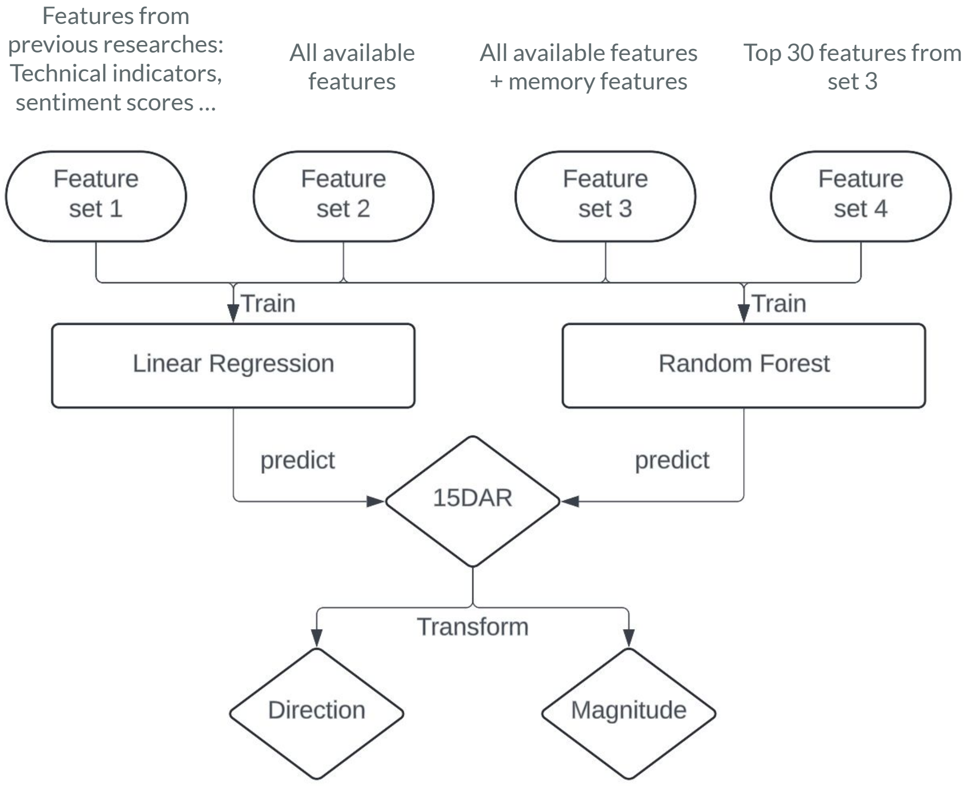
1. ~~Objective~~
2. Data
3. Signal Generation Experimnetation
4. Model Implementation
5. Strategy & Backtest Results
6. Key Contributions
7. Future Work

Data



News from TIINGO QC
BERT Sentiment Scores
S&P 500 VIX QuantConnect
Google Search Trend
Search Signal
BTC price from QuantConnect
Technical Indicators
Features 2021/4/15 - 2024/3/19
Treasury Yields from FRED
Twitter Sentiment Scores from Bloomberg

Signal Generation Experimentation

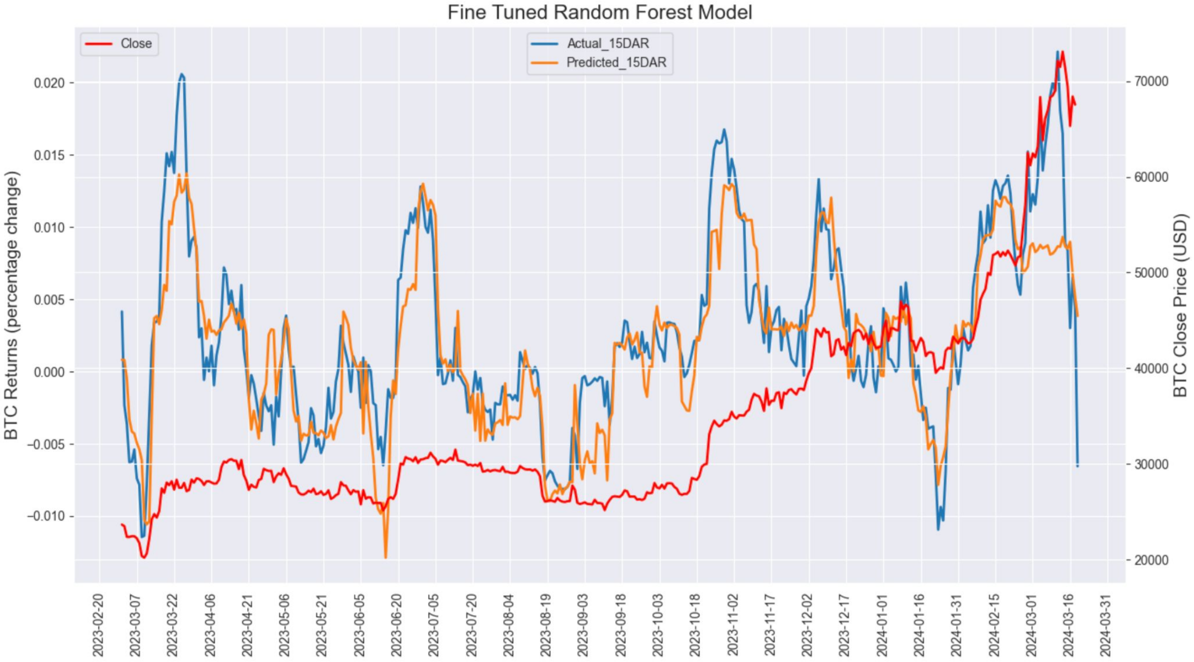


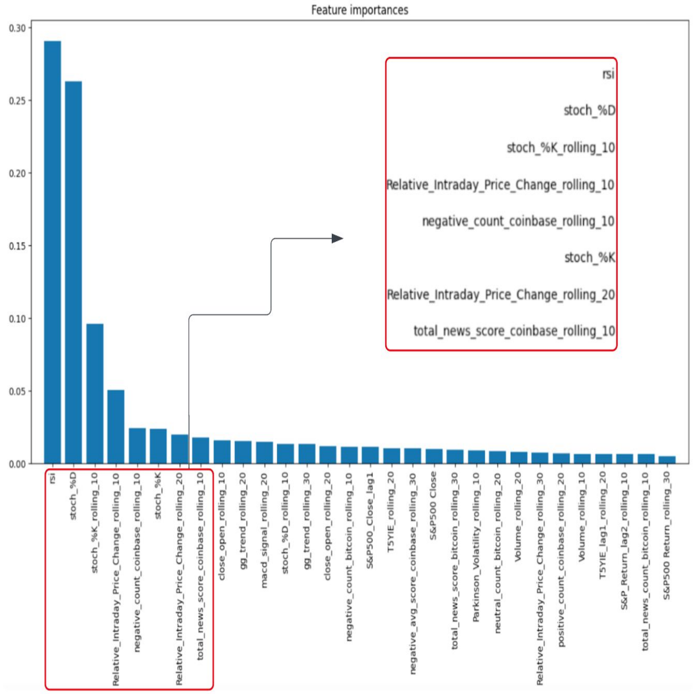
15DAR R-squared	Feature set 1	Feature set 2	Feature set 3	Feature set 4
Linear Regression	0.57	0.63	-41.82	0.59
Random Forest	0.59	0.67	0.73	0.73

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

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

Results Vs Literature

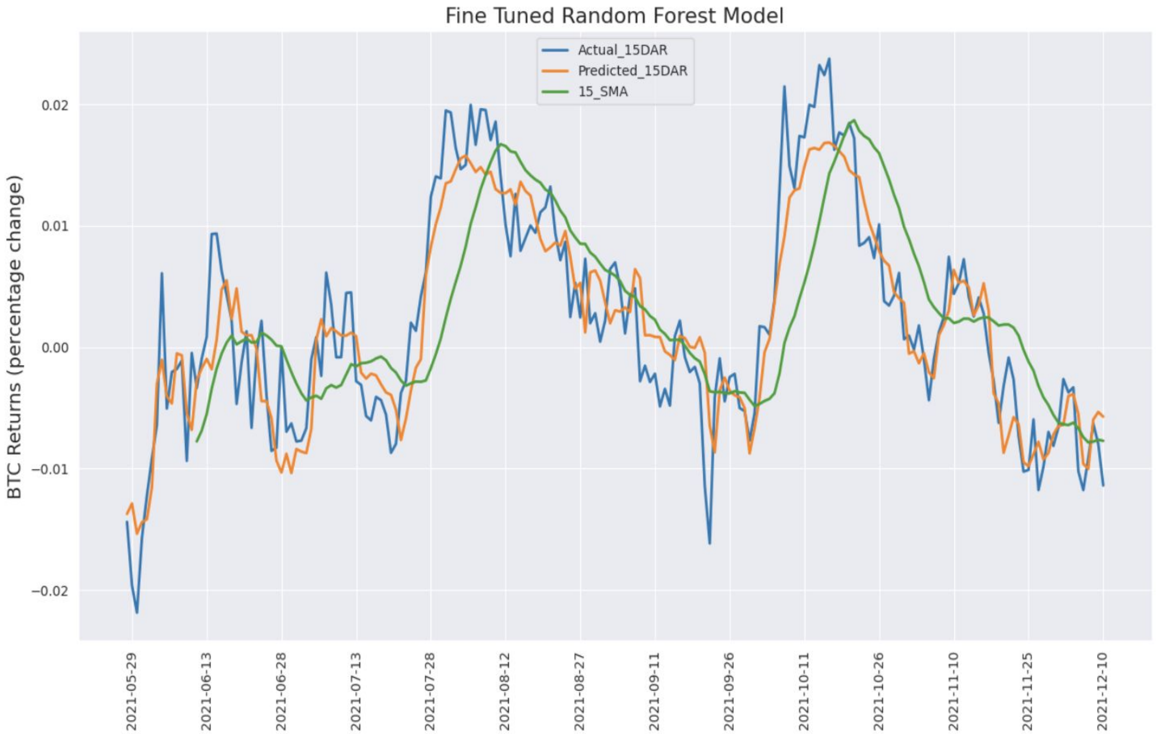


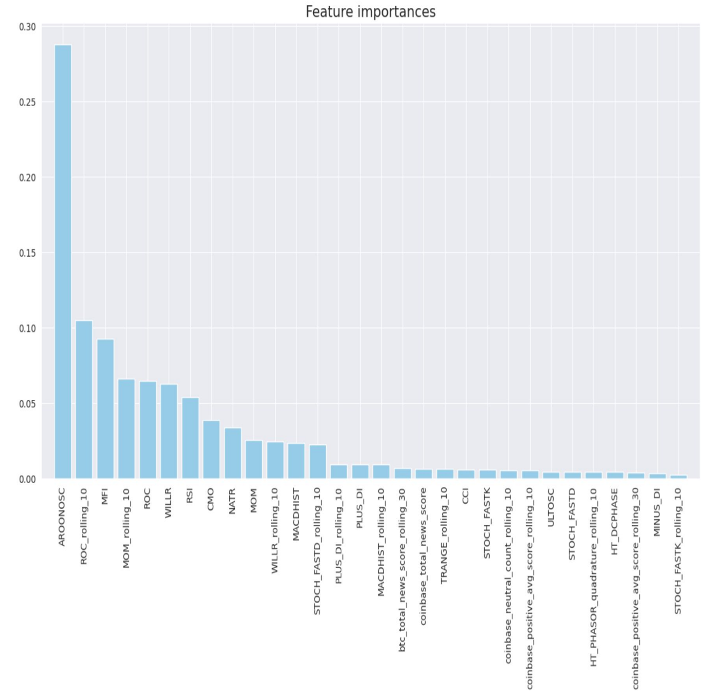


	Best from Literature Review	Refined Random Forest Model
Daily Price Direction Forecast Accuracy	68.4%	87%
Daily Price Movement Magnitude Forecast F1 Score	14.21%	16.89%
Daily Price Movement Magnitude Forecast Accuracy	51.47%	22.25%

QC Model Implementation (OOS A & B)

```
Mean Squared Error: 1.4387767253005934e-05
R Squared: 0.8253809634139997
```





Our Strategy

Step 1: Make the Trading Decision

long = short = 0 # (indicators)
For each day d:

1. Get tomorrow's predicted 15 day moving average (**S_d**)
2. if **S_d** > 1 %:

        # Market is bullish
        long = 1
        short = 0

    elif **S_d** < -0.5%:

        # Market is bearish
        long = 0
        short = 1

    else:

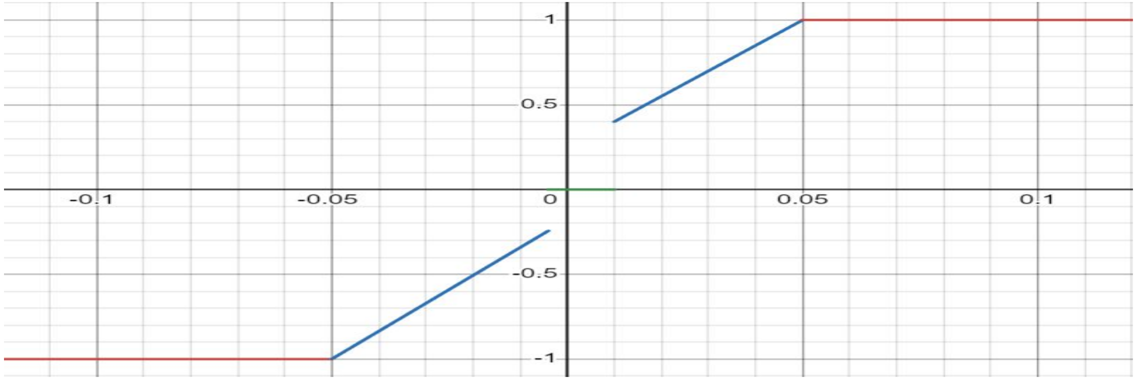
        long = 0
        short = 0

Step 2: Transform **S_d** to importance scores I_d

I_d = ABS(ImportanceFunc(S_d)) # (0 to 1)
higher importance means more confidence.

if **I_d** < 0.51:

    # Kelly_c doesn't make sense for smaller values
    I_d = 0.51



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

Our Strategy (Continued)

Step 4: Trading Execution

Buy and Sell according to the signal and kelly fraction **kf_d**

if long == 1:

- Long Bitcoin (0.8 * **kf_d**) of Portfolio Value
- Long Ethereum (0.2 * **kf_d**) of Portfolio Value

elif short == 1:

- 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 value falls by more than 5% in a day.
- Liquidate if portfolio value falls for nine consecutive days

Slippage Model

1. [Integration with Trading Strategy]
    - Estimate the potential slippage for each order

2. [Model]

Slippage = asset.price * 0.0001 * log(2 * order.absolute_quantity)

- Adjustment on order size

BTC (.8) ETH (.2) Buy & Hold (OOS A & B)

Metric	Value
Sharpe Ratio	0.256
Compounding Annual Return	-8.852%
Max Drawdown	43.4%

OOS A - April 10, 2021 to August 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

OOS B - September 10, 2021 to December 10, 2021

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

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

OOS A & B - April 10, 2021 to December 10, 2021

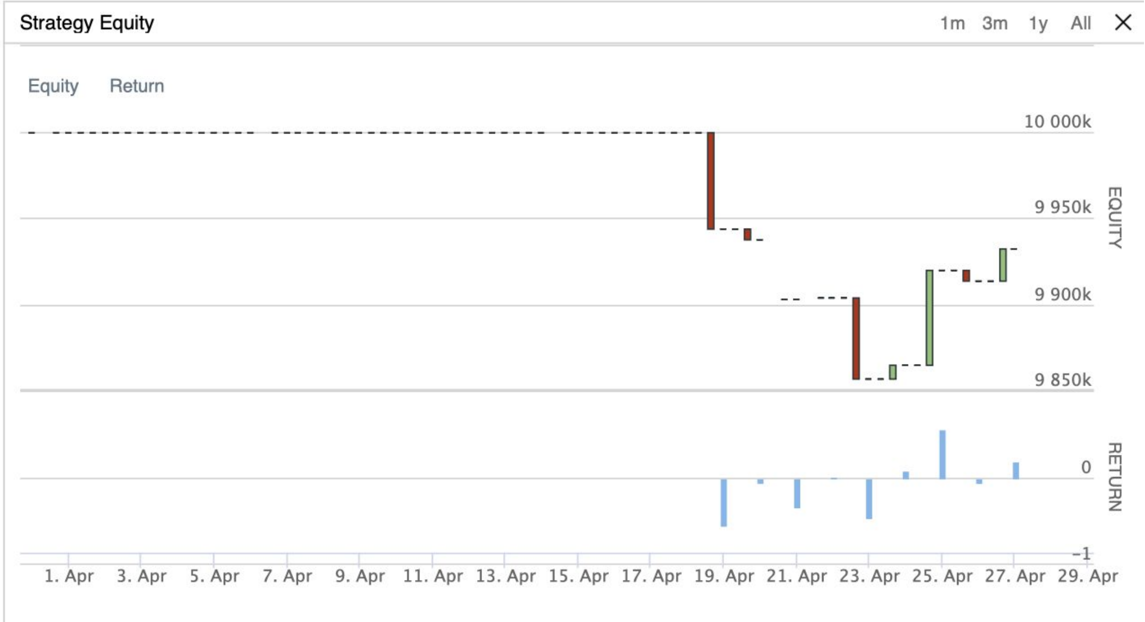
Return	41.21%
PSR	72.093%
Sharpe Ratio	2.056
Max Drawdown	10.1%
Compounding Annual Return	81.225%
Volume	$ 96,843,370
Net Profit	$ 4,120,863

Backtest link:
https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_78aa6cba58fcfd7865f5bd0c64928241.html

BTC (.8) ETH (.2) Buy & Hold (OOS C)

Metric	Value
Sharpe Ratio	-1.514
Compounding Annual Return	-76.99%
Max Drawdown	15.4%

OOS C (Blind) - March 31, 2021 to April 27, 2021



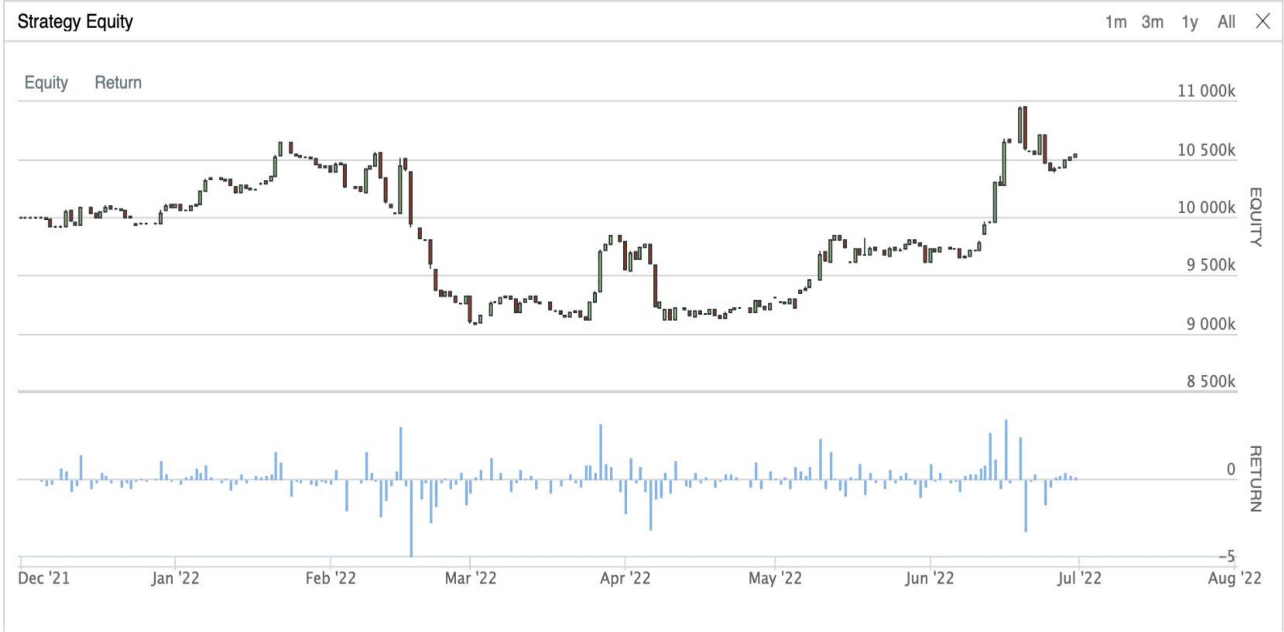
Return	-0.68%
PSR	16.10%
Sharpe Ratio	-3.714
Max Drawdown	1.4%
Compounding Annual Return	-8.25%
Volume	$ 3,369,402
Net Profit	$ -67,843

Backtest link:
https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_17158a451e216352043a4311447e4086.html

BTC (.8) ETH (.2) Buy & Hold (Stress Test Period)

Metric	Value
Sharpe Ratio	-1.194
Compounding Annual Return	-85.938%
Max Drawdown	69%

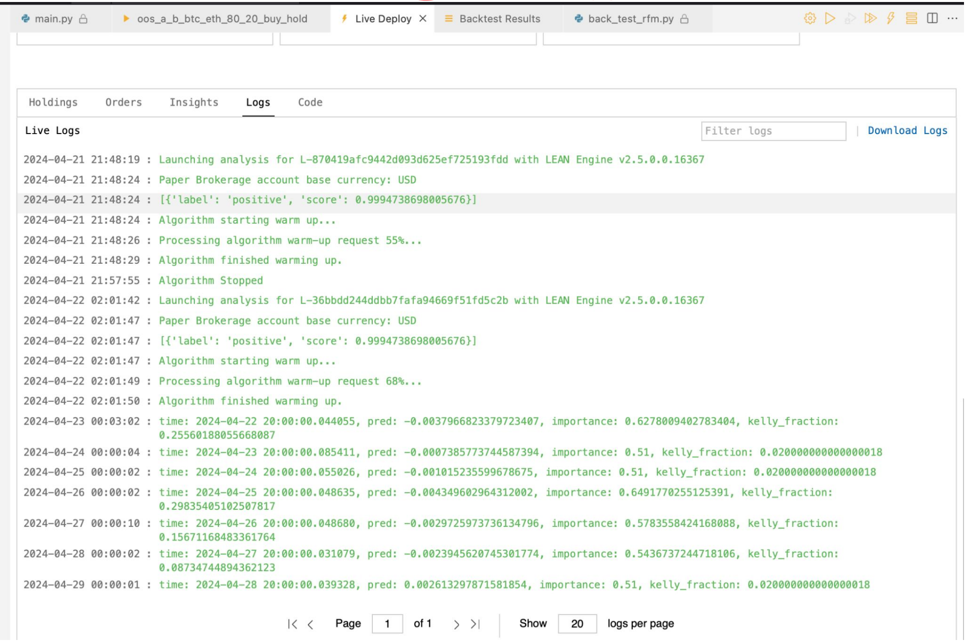
Stress Test (Blind) - Dec 1, 2021 to Jul 1, 2022

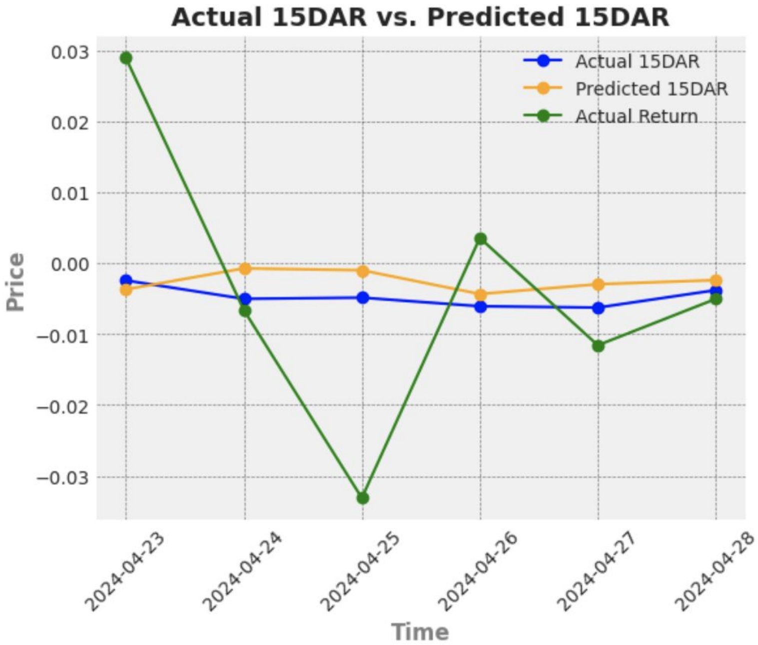


Return	5.42%
PSR	30.65%
Sharpe Ratio	0.431
Max Drawdown	14.7%
Compounding Annual Return	9.41%
Volume	$ 61,927,757
Net Profit	$ 540,866

Backtest link:
https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_95e01ed2bf0b70dbeeec18fc4b0db5d1.html

Live Trading - April 23, 2024 to April 28, 2024





- Return direction was predicted accurately on 4 out of 6 days
- Not enough days to account for variance
- Since predictions were all neutral, no trades were made

Key Contributions

- Enhancing Signal to Noise Quality in Returns Prediction
- Advanced Sentiment Analysis using Local LLMs
- Incorporating Working Memory Features
- Original Trading Strategy Based on Modified Kelly Criterion & Signal Importance

Future Work

- Try Alternative Trading Frequencies
- Making additional models to specialize in other cryptocurrencies
- Fine tune BERT LLM on only cryptocurrency data
- Incorporate more social sentiment data sources (reddit, facebook, etc.)

Thank you!

References

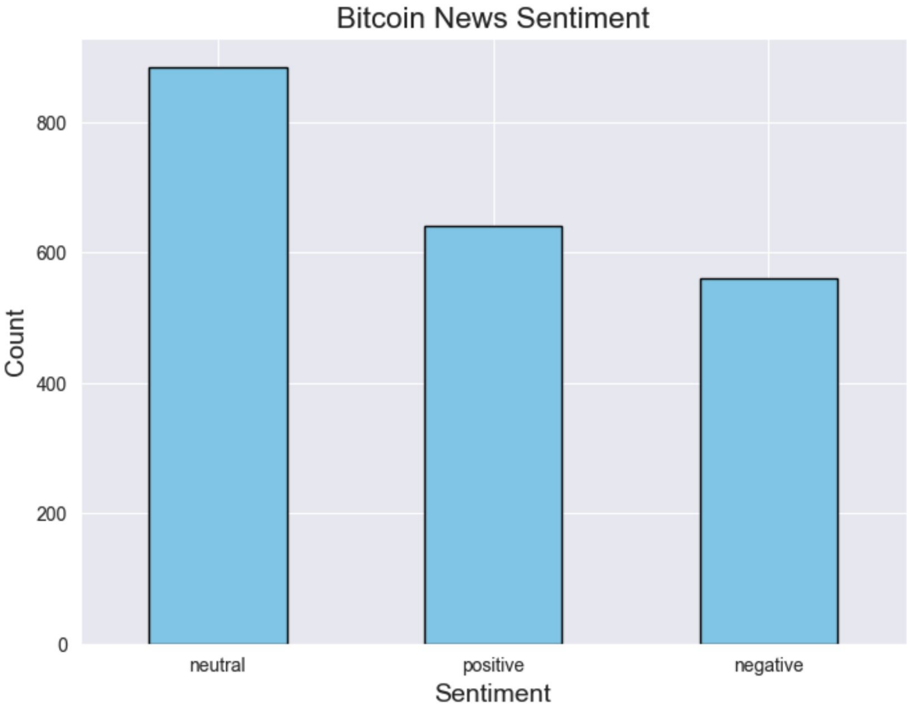
1. Arratia, A., López-Barrantes, A.X.(2021). Do Google Trends forecast bitcoins? Stylized facts and statistical evidence. J BANK FINANC TECHNOL 5, 45–57 (2021). https://doi.org/10.1007/s42786-021-00027-4
2. Couzin-Frankel, J. (2019). Questions churn about vaping's long-term risks. Science, 366(6469), 1059–1060. https://doi.org/10.1126/science.366.6469.1059
3. 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
4. Denning, P., & Little, J. (2017). Over the influence: The harm reduction guide to controlling your drug and alcohol use (2nd ed.). Guilford Press.
5. Devlin, Chang, Lee, & Toutanova (2018), Bidirectional Encoder Representations from Transformers. https://arxiv.org/abs/1810.04805
6. 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
7. Earp, J. (Director) & Katz, J. (Presenter) (2013). Tough guise 2: Violence, manhood & American culture [Film; educational DVD]. Media Education Foundation.
8. Gabrovsek, Peter; Aleksovski, Darko; Mozetic, Igor; Grcar, Miha (2017). Twitter Sentiment around the Earnings Announcement Events. https://arxiv.org/abs/1611.02090
9. 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
10. Katsiampa, P. (2019), An empirical investigation of volatility dynamics in the cryptocurrency market. https://www.sciencedirect.com/science/article/abs/pii/S0275531919300637
11. 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
12. McKee, A. (2015, May 15). Introduction to criminal justice, section 1.2 [Video]. YouTube. https://youtu.be/ZXnp8OQwgH8
13. Merriam-Webster. (n.d.). Insanity. In Merriam-Webster.com dictionary. Retrieved January 9, 2020, from https://www.merriam-webster.com/dictionary/insanity
14. 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
15. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System; 2008. https://bitcoin.org/bitcoin.pdf.
16. Weissinger, S. E., & Mack, D. (Eds.). (2018). Law enforcement in the age of Black Lives Matter: Policing black and brown bodies. Lexington Books.
17. World Health Organization. (2018, December). Gaming disorder. https://www.who.int/news-room/q-a-detail/gaming-disorder

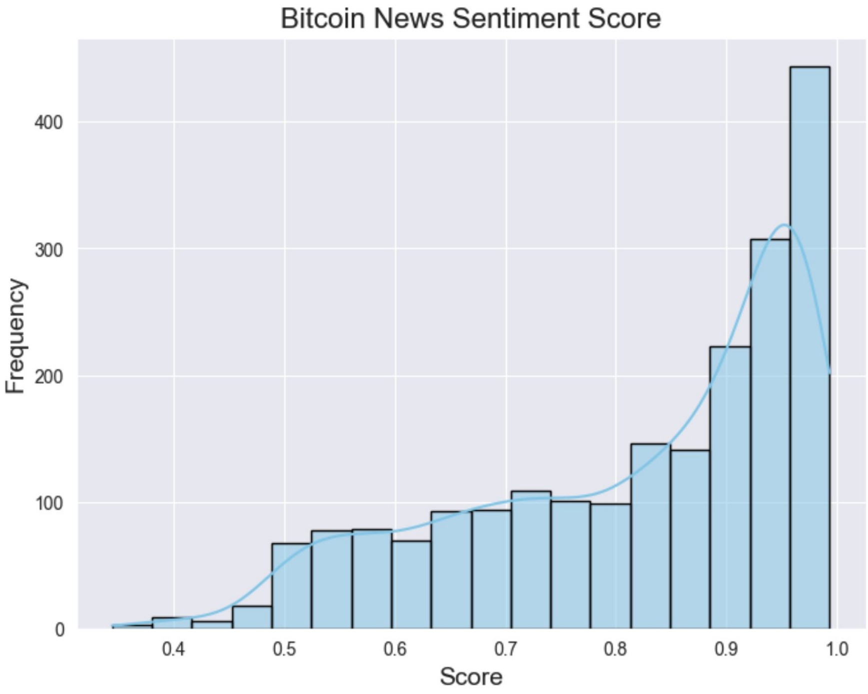
INS - April 10, 2022 to March 31, 2024

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

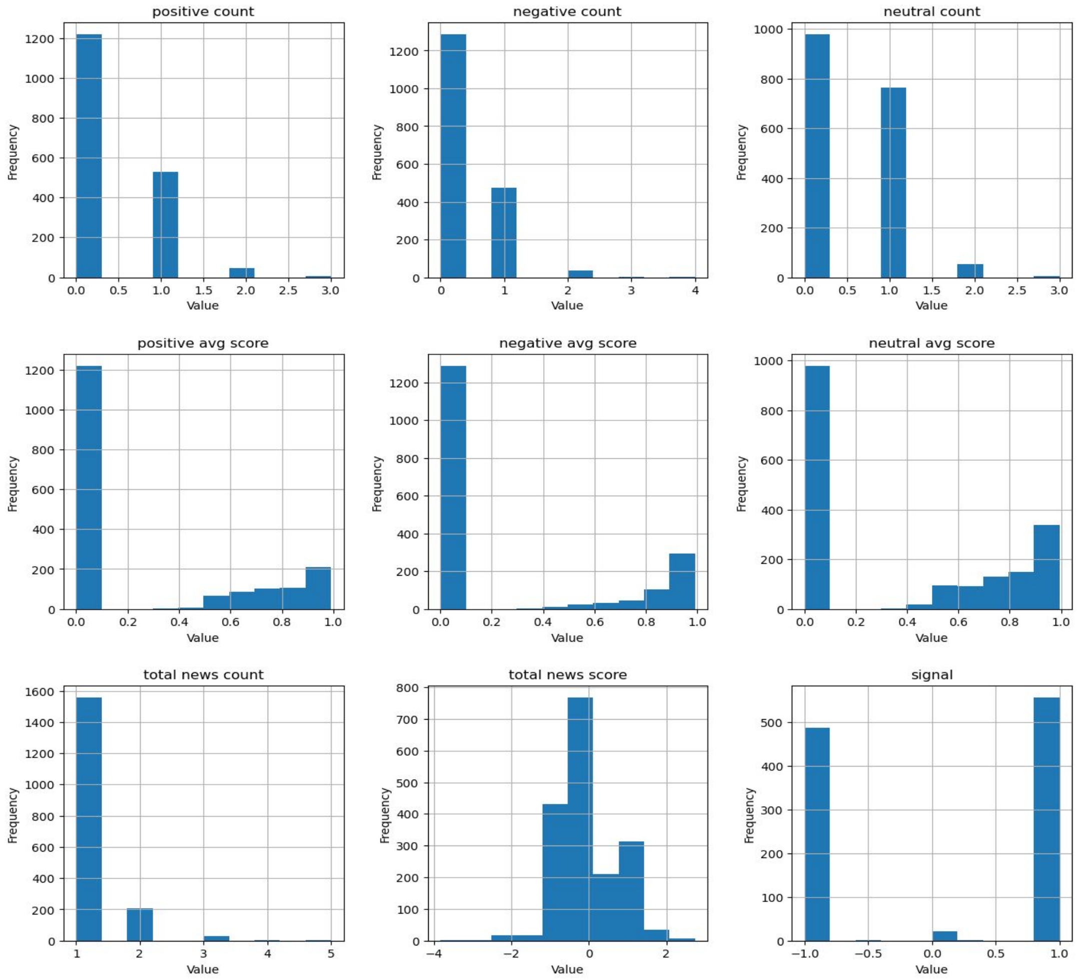
Backtest link:
https://www.quantconnect.com/terminal/processCache?request=embedded_backtest_7bcfb7777a07fd6aa450c2084f99c698.html

Distribution of news article Sentiments & Scores assigned by BERT





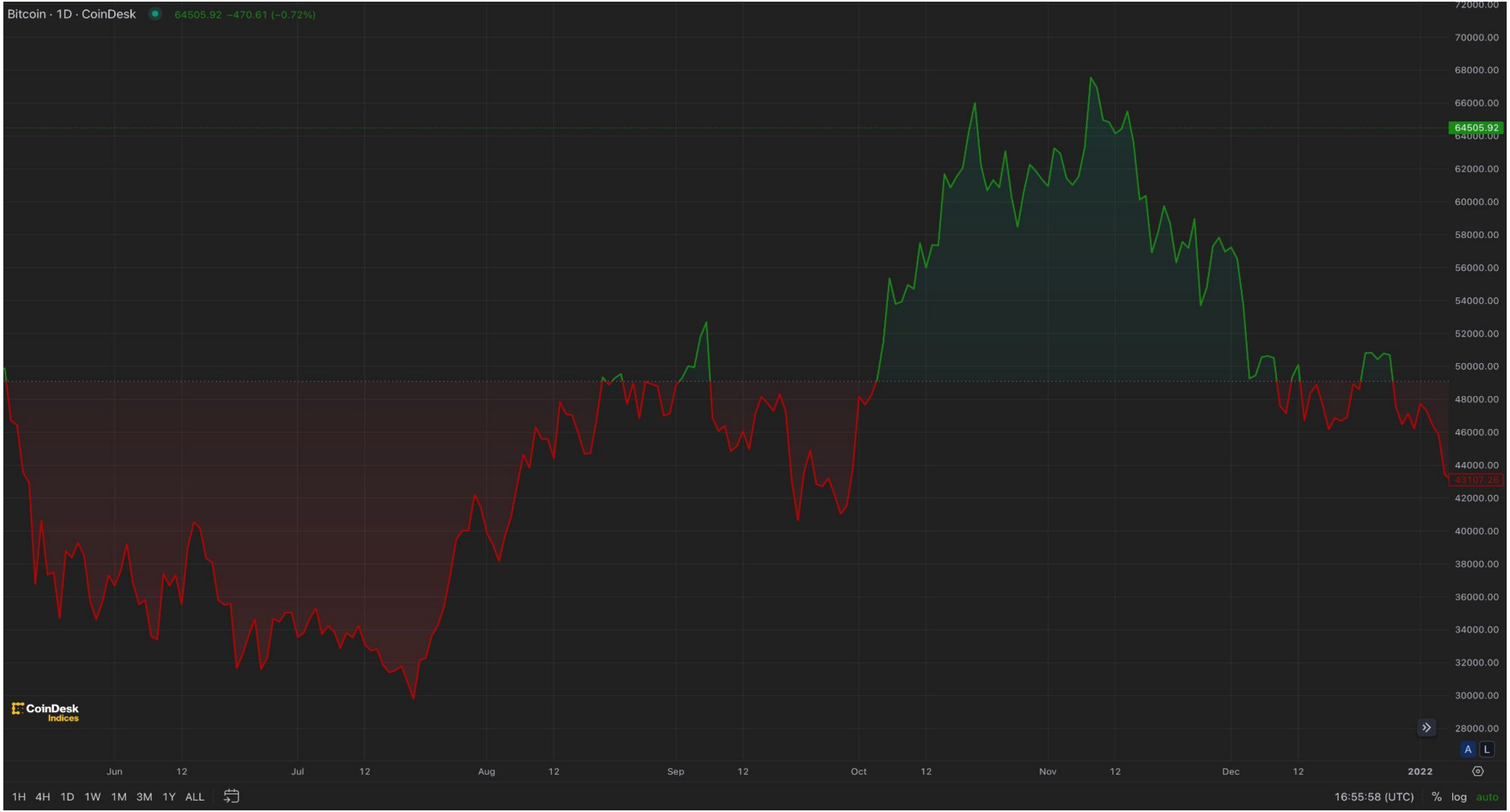
Distribution of daily aggregated news Sentiments Scores



Team Structure

- **Project Manager (Rakeen):** Oversees project timelines, sets milestones, and ensures team cohesion. Responsible for integrating the work of all team members and maintaining communication channels among the team and with external stakeholders if necessary.
- **Research Analyst (Tianji):** Focus on gathering, analyzing, and synthesizing relevant research, including academic papers and market analyses on sentiment analysis and its impact on crypto prices. This role requires a deep understanding of the cryptocurrency market and the ability to extract actionable insights from complex datasets.
- **Data Scientist (Tianji):** Specialize in data modeling, including the creation and optimization of algorithms for sentiment analysis and price prediction. These team members should be proficient in Python and familiar with libraries and packages relevant to data analysis and machine learning.
- **Trading Strategist (Rakeen):** Develop and refine trading strategies based on the predictive models provided by data scientists. This involves backtesting strategies against historical data, optimizing for risk management, and suggesting adjustments based on performance.
- **Risk Management Specialist (Ching-Lung):** Focus on identifying, assessing, and mitigating financial risks associated with the trading strategies. They will use various risk measurement tools, like Value-at-Risk (VaR) and stress testing, to ensure the sustainability of the trading model.
- **Portfolio Manager (Ching-Lung):** In charge of capital allocation among various strategies and managing the overall portfolio to achieve a balanced risk-reward ratio. This role requires a deep understanding of portfolio theory and the ability to make decisions based on the team's risk appetite and investment goals.
- **Software Engineer (Mike):** Responsible for coding the trading algorithms, setting up the trading infrastructure, and ensuring the seamless execution of trades. This includes integrating the trading models into a trading platform like QuantConnect and maintaining the codebase.
- **Communications Officer (Mike):** Handles documentation, presentation preparation, and internal reports. This role ensures that complex technical concepts are translated into understandable language for presentations and reports.

CoinDesk BTC Price (OOS A & B)



Literature Review

Type	Previous Paper
Data and Inputs	**Technical Indicators** (Huang et al., 2018) Huang et al., 2018 used 124 technical indicators, including momentum, volatility, cycle, and other indicators, to predict the movement of bitcoin price. **Sentiment Analysis** (Ider & Lessamnn, 2022, Da et al, 2011) Sentiment scores from Twitter, News, Google Trends, and other social media can boost the forecasting of the bitcoin price. **Other Financial Markets** (Vo & Xu, 2017)
Models	**Sentiment Analysis:** Lexicon-based approach/ Machine Learning approach: (Renault, 2022)/ Deep Learning approach: BERT (Devlin, 2018; Hu, et al., 2019; Sousa et al., 2019) **Predictive Model:** Time series model: ARIMA/ Machine Learning model: Random Forest, XGBoost/ Deep Learning model: LSTM, CNN, Transformer

Type	Previous Student Projects
Crypto	1. Cryptocurrencies Trading Strategy based on Sentiment Analysis (2019) 2. Advanced Volatility Arbitrage (2019)
Sentimental Analysis	1. Deep Learning for Structural Break Prediction in Equities Pairs Trading (2022) 2. ETF Trading based on factor investment (2023) 3. Extreme Gradient Profit 4. Trading, Dispersion, & Correlation (2023) 5. Sentiment-Boosted Pairs Trading (2023)