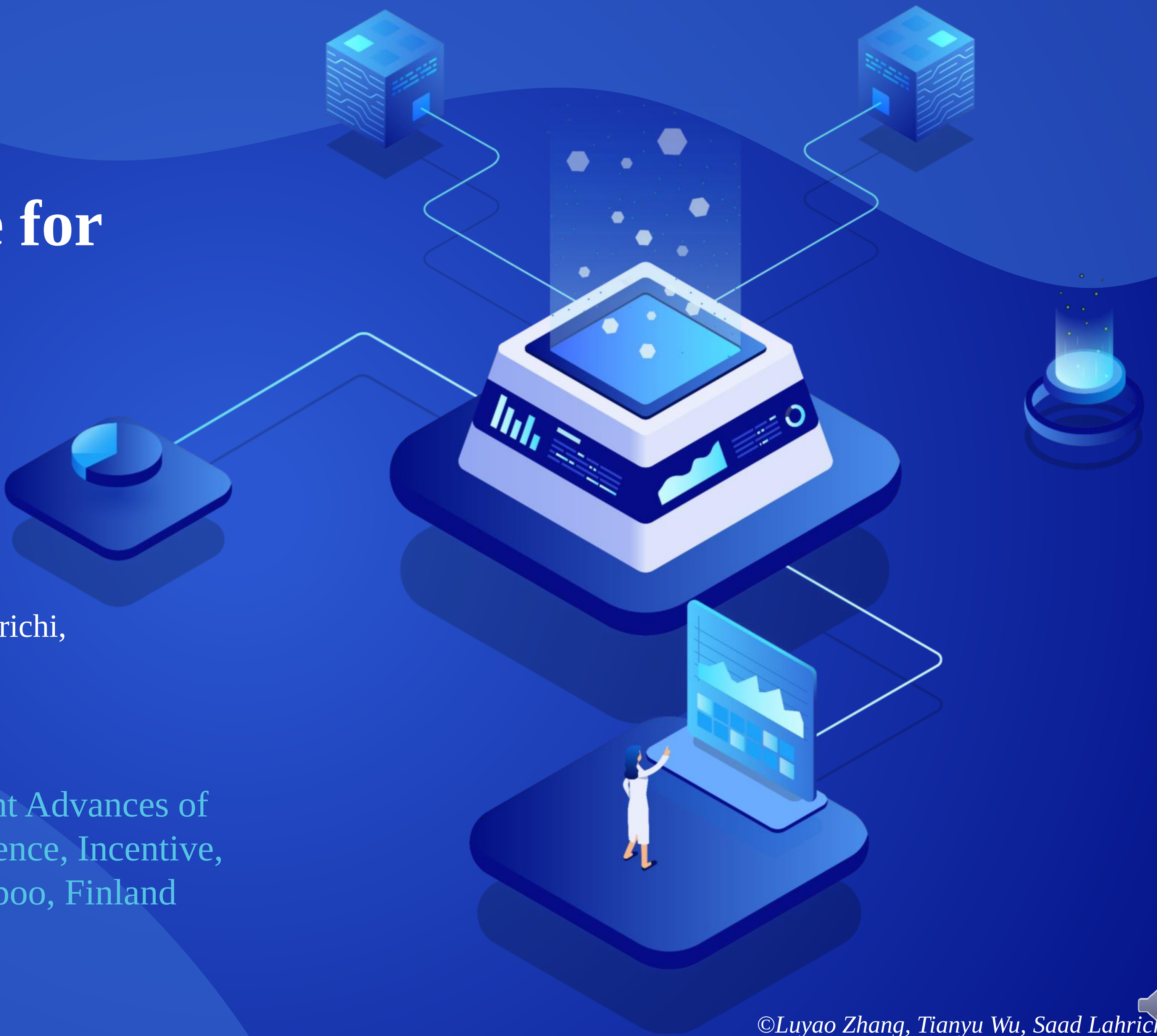
A Data Science Pipeline for Algorithmic Trading:

A Comparative Study in Applications to Finance and Cryptoeconomics

Co-authors: Luyao Zhang*, Tianyu Wu, Saad Lahrichi,
Carlos-Gustavo Salas-Flores, and Jiayi Li



The First International Symposium on Recent Advances of Blockchain Evolution: Architecture, Intelligence, Incentive, and Applications (BlockchainEvo 2022), Espoo, Finland
Date: August 22-25, 2022

CONTENTS

0

At a Glance

The Team

Luyao Zhang*
Data Science Research Center,
Social Science Division,
Duke Kunshan University
China

SciEcon CIC
United Kingdom

Tianyu Wu
Duke Kunshan University
China

SciEcon CIC
United Kingdom

Saad Lahrichi
Duke Kunshan University
China

Jiayi Li
Duke Kunshan University
China

Carlos-Gustavo Salas-Flores
Duke Kunshan University
China

✓ Data Science, Finance, Cryptocurrencies
✓ Cultivate undergraduate research
✓ Across different time zones

** Corresponding Author*
Luyao, Tianyu, and Saad are joint 1st authors, and Jiayi and Carlos-Gustavo are 2nd authors.

Background and Research Questions

The Big Picture

1. Artificial Intelligence – **an important technique** of algorithmic trading in finance and cryptoeconomics.

2. Lack of **established pipeline** leads to process inconsistency and makes ceteris-paribus comparison difficult.

3. No **open-source coding algorithms** to evaluate and compare different trading strategies.

Our Questions

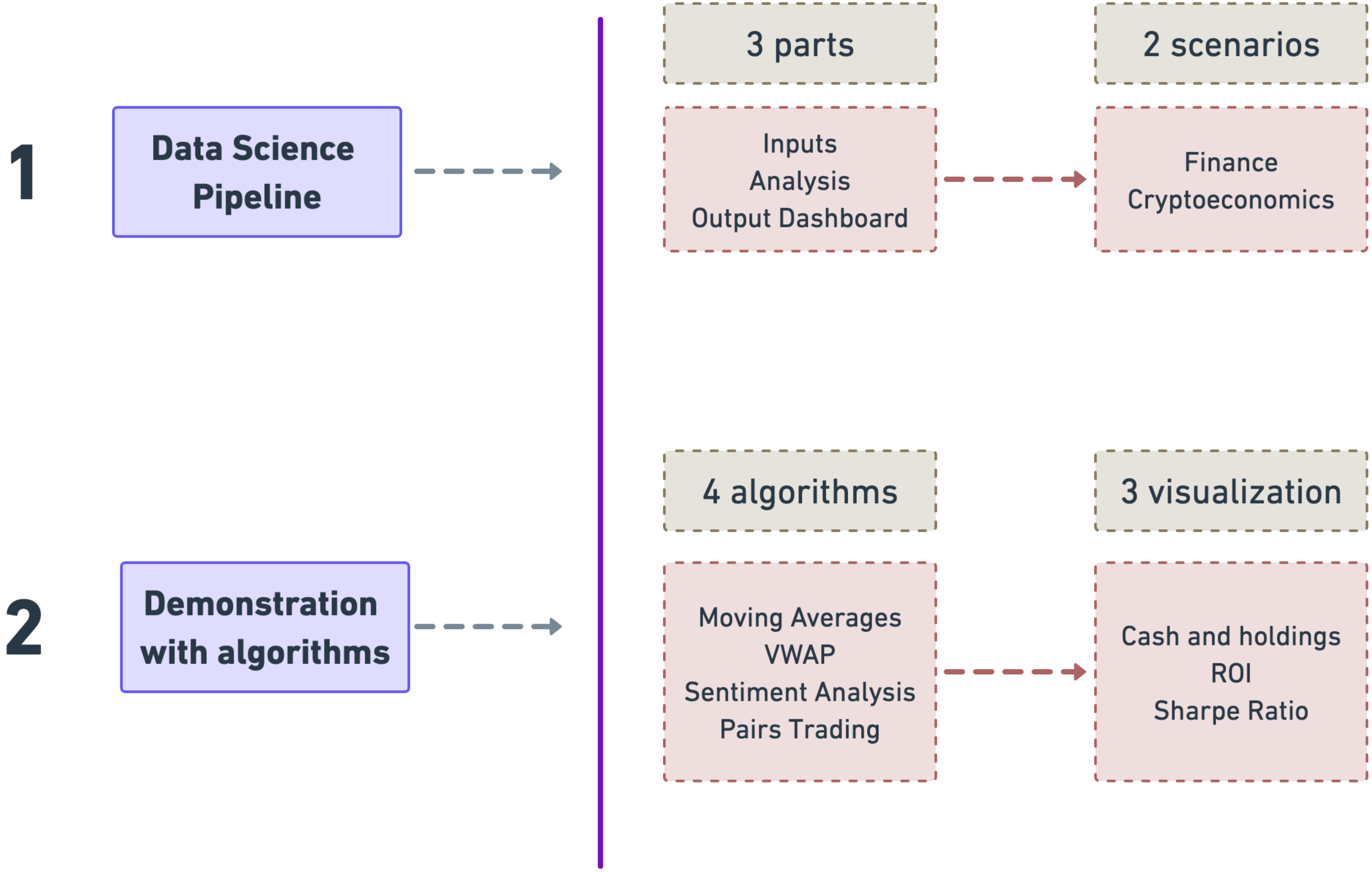
1. **A generally applicable data science pipeline:**

   What are the inputs, analysis, and output dashboard in this workflow?

2. **A comparative study in Applications to Finance and Cryptoeconomics:**

   How can this data science pipeline be generally applicable to design, program, and evaluate algorithmic trading of stock and crypto assets with conventional algorithms?

Main Results



https://whimsical.com/fig1-GkT91dQfNvYLfY1pTJ5T6g

Data Source and Methodology

Moving Averages (MA) Crossover

Data Sources

API Documentation | Alpha Vantage

Assets

Tesla (TSLA)
Apple (AAPL)
Bitcoin (BTC)
Ether (ETH)

Calculate Index

Short-window Moving Average (SWMA)
Commonly 50-Days

Long-window Moving Average (LWMA)
Commonly 200-Days

Buy Signal

SWMA crosses above LWMA

Input Variable

Historical Close Price
Frequency: Daily
Start date: 2020/01/01
End date: 2022/02/04

Trading Condition

Initial capital: 10,000

max_buy: 10,000,000
max_sell: 10,000,000
gas fee: 0.1%

Sell Signal

SWMA crosses below LWMA

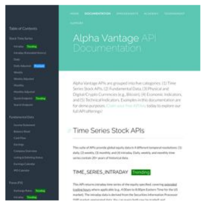
Evaluations

ROI
Sharpe Ratio

$$\mathrm{SMA}_t^n = \frac{1}{n} \sum_{i=t-n+1}^{t} p_i$$

Volume-Weighted Average Price (VWAP)

Data Sources



API Documentation | Alpha Vantage

Input Variables

Historical Price
Frequency: intraday
Start date: 2022-02-08
End date: 2022-02-11

Assets

Ethereum (ETH)
Bitcoin (BTC)

Calculate index

VWAP

Trading Condition

Initial Capital: 100,000 USD
Max_buy: 1,000,000
Max_sell: 1,000,000
Gas fee: 0.1%

Buy Signals

Price cross above VWAP

Sell Signals

Price cross below VWAP

Evaluations

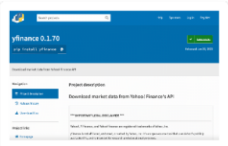
ROI
Sharpe Ratio

$$VWAP = \frac{\sum P_t \cdot Q_t}{\sum Q_t}$$

Data Source and Methodology

Sentiment Analysis

Data Sources 1



yfinance PyPI

Input Variable 1

Historical Prices

Frequency: Daily
Start date: 2021-01-01
End date: 2021-06-01

Calculate Index

EMA

Buy Signal

Price-EMA< -std(Price-EMA) &
Positive Sentiment

Data Sources 2

Github: snscrape: A social networking service scraper in Python

Input Variable 2

Tweets

Frequency: 1 tweet
Start date: 2021-01-01
End date: 2021-06-01

NLP

VADER

Sentiment Category

(Positive, Negative)

Sell Signal

Price-EMA > std(Price-EMA) &
Negative Sentiment

Evaluation

ROI
Sharpe Ratio

Assets

Apple (AAPL)
Bitcoin (BTC)
Ether (ETH)

Trading Condition

Initial capital: 100000
max_buy: 10000000
max_sell: 10000000
gas fee: 0.1%

Pairs Trading

Data Sources

API Documentation | Alpha Vantage

Input Variables

Historical Close Prices
Frequency Daily
Start date: 4/22/2021
End date: 10/22/2021

Trading Condition

initial_capital: 10,000 USD
max_number_of_pairs: 20
transaction_fee: 0.01
gas_fee: 0.001
borrow_fee: 0.01
max_buy: 1000000
max_sell: 1000000
max_short: 1000000

Assets

Dataset 1: S&P500
Dataset 2: Digital Currencies

Pairs Selection using Machine Learning

Compute ROI → Dimensionality Reduction (PCA) → Unsupervised Machine Learning (OPTICS) → Mean Reversion & Cointegration Tests

Pairs Selection using Brute Force

Sum of Standard Deviations (SDD) → Select top k pairs with the smallest SDD

Calculate Simple Spread

Spread_simple = PriceA - PriceB

Calculate Normalized Spread

Spread_normal = PriceA/PriceA_0 - PriceB/PriceB_0

Define Thresholds

Define:
mu = mean(Spread)
sigma = standard deviation (Spread)

buy threshold = mu - 2 sigma
short threshold = mu + 2 sigma
exit = mu

Define:
sigma = standard deviation (Spread)

buy threshold = - 2 sigma
short threshold = 2 sigma
exit = 0

Buy A and Short B Signal

Spread < buy threshold

Short A and Buy B Signal

Spread > short threshold

Sell and Payback Signals

Spread = exit

Evaluations

ROI
Sharpe Ratio

1

Introduction to Data Science Pipeline

Introduction to Data Science Pipeline

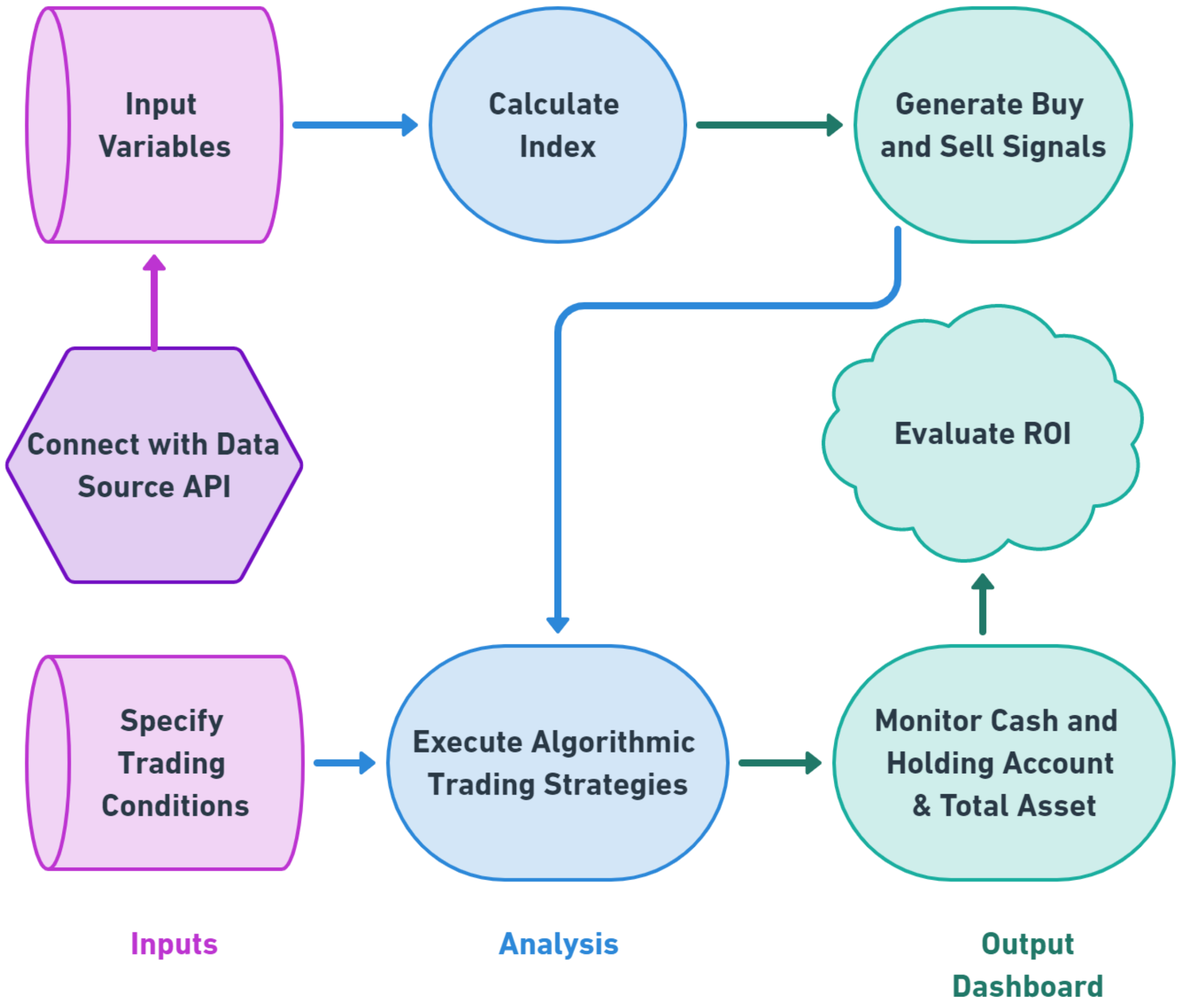


Figure 1: Our proposed data science pipeline for algorithmic trading

2

Demonstration with Four Conventional Algorithms

Moving Averages Crossover
Volume-Weighted Average Price
Sentiment Analysis
Pairs Trading

Moving Averages Crossover

Variable	Frequency	Unit	Description
Date	daily	YYYY-MM-DD	Date and time for which the data were recorded
Close	daily	USD	Price at which the stock ended trading in a given time period
Short MA	daily	USD	Average price of a security within a certain period, typically 50 days
Long MA	daily	USD	Average price of a security within a certain period, typically 200 days
Signal	-	-	Buy-and-sell signal (e.g., TSLA, AAPL for stock, BTC, ETH for crypto)

Table 1: Moving Average Crossover: Data
Data Source: Alpha Vantage API

Moving Averages Crossover

Take Ether (ETH) for example:

Figure 2: Buy-and-Sell Signal: ETH moving averages crossover

Figure 3: Portfolio time series: ETH moving averages crossover

Moving Averages Crossover

Take Ether (ETH) for example:

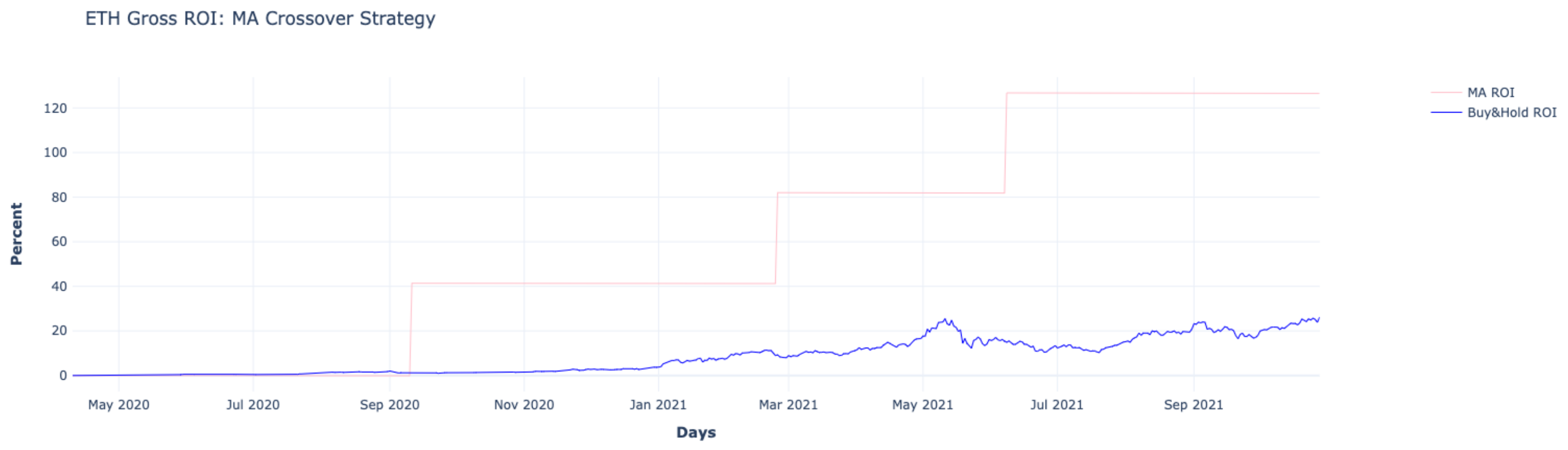


Figure 4: Gross ROI: ETH moving average crossover vs. buy-and-hold

Moving average crossover strategy ROI: 849.84%
Buy & hold strategy ROI: 11.97%

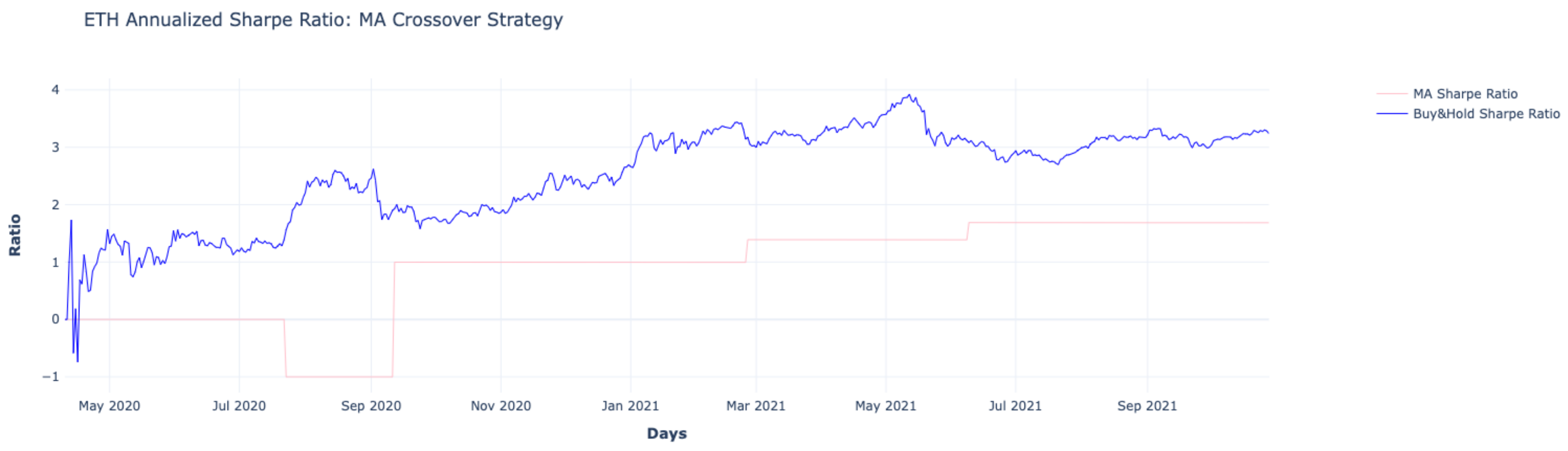


Figure 5: Sharpe Ratio: ETH moving average crossover vs. buy-and-hold

Moving average crossover strategy Sharpe ratio: 0.98
Buy-and-hold strategy Sharpe ratio: 2.60

Volume-Weighted Moving Average

Variable	Frequency	Unit	Description
Date	5 min	YYYY-MM-DD HH:MM:SS	Date and time for which the data were recorded
Close	5 min	USD	Price at which the stock ended trading in a given time period
VWAP	5 min	USD	Average price of a security within a day, adjusted for its volume. Available for an API call only for traditional stocks; manually calculated using the formula for crypto
Ticker	-	-	Stock symbol (e.g., TSLA, AAPL for traditional, BTC, ETH for crypto)
Interval	5 min	min/hr/day	Time difference between two data points

Table 2: Volume-Weighted Moving Average: Data
Data Source: Alpha Vantage API

Volume-Weighted Moving Average

Take Ether (ETH) for example:

Figure 6: Buy-and-Sell Signal:
ETH volume-weighted moving average

Figure 7: Portfolio time series:
ETH volume-weighted moving average

Volume-Weighted Moving Average

Take Ether (ETH) for example:

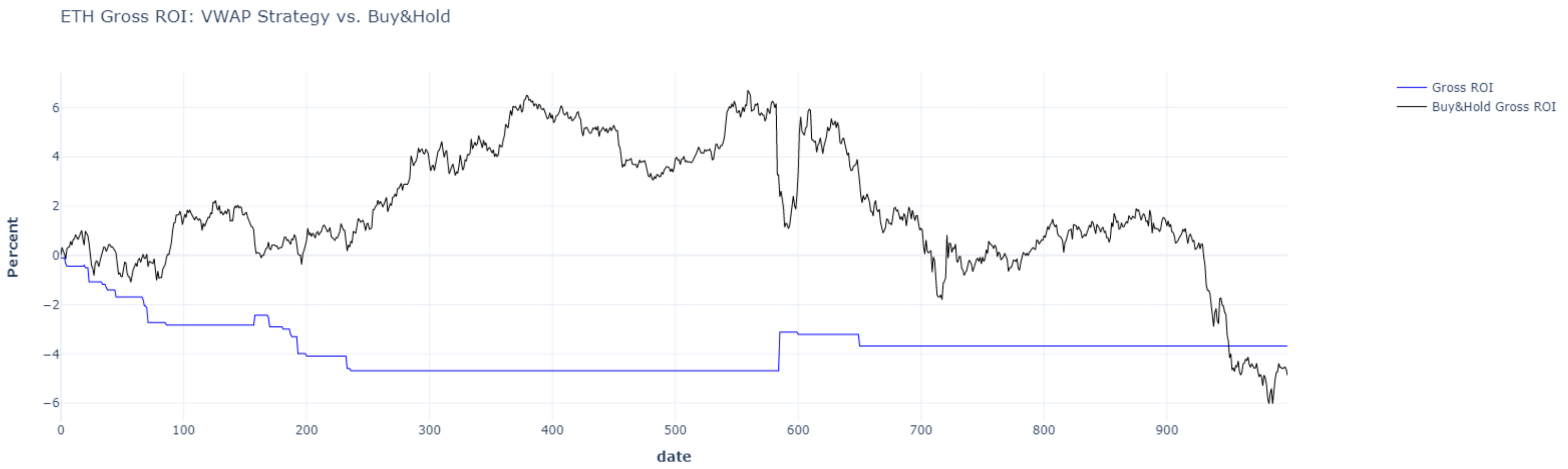


Figure 8: Gross ROI: ETH volume-weighted moving average vs. buy-and-hold

Moving average crossover strategy ROI: -3.93%
Buy & hold strategy ROI: -4.26%

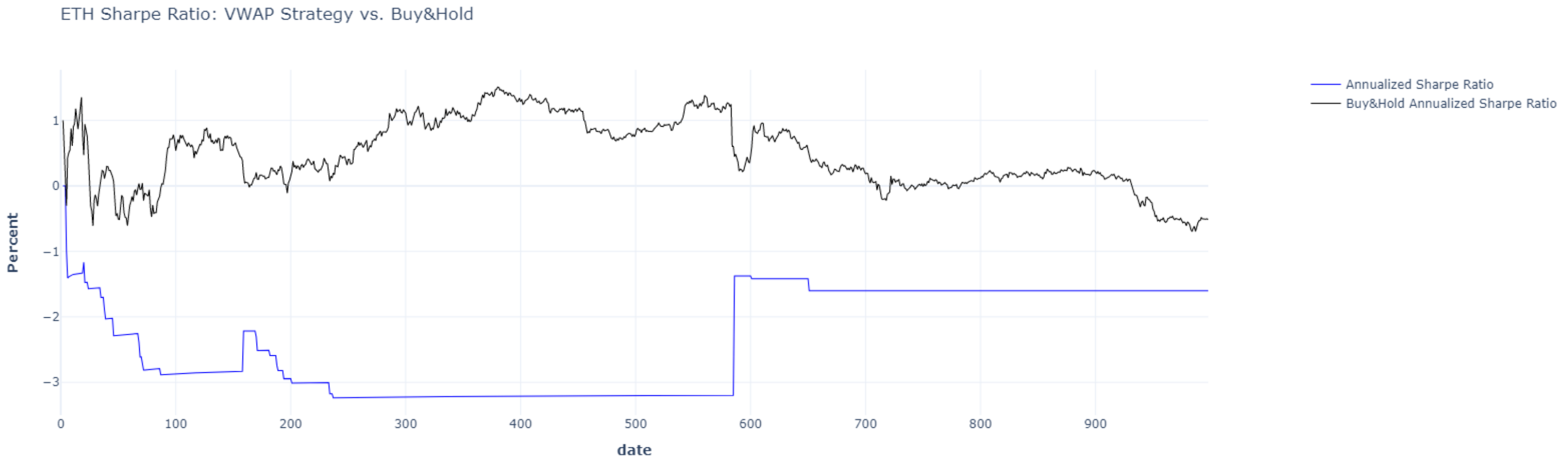


Figure 9: Sharpe Ratio: ETH volume-weighted moving average vs. buy-and-hold

Moving average crossover strategy Sharpe ratio: -1.62
Buy-and-hold strategy Sharpe ratio: -0.52

Sentiment Analysis

Variable Name	Frequency	Unit
Date	daily	YYYY-MM-DD
Tweets	-	-
Stock Price	daily	USD
Crypto Price	daily	USD

Table 3: Sentiment Analysis: Raw Variables
Data Sources: Snscrape API, Yahoo Finance API, Alpha Vantage API

ID	Frequency	Unit	Description
P_{comp}	per tweet	-	A normalized compound score that sums every lexicon rating and takes values from -1 to 1
EMA_t (10-day)	daily	USD	Exponential moving average of adjusted closing price, where $EMA_0 = P_0$ and $EMA_t = (1-\alpha)EMA_{t-1} + P_t, \alpha = 2/(s+1)$. For span $s \geq 1, s$: decay in terms of span P_t : day t closing price
Raw Trading Position	daily	USD	$P_t - EMA_t$
Sentiment Category	daily	Negative, Positive, Neutral	Negative: $P_{comp} < \bar{P}_{comp} - 0.2\sigma_{P_{comp}}$ Positive: $P_{comp} > \bar{P}_{comp} + 0.2\sigma_{P_{comp}}$ Neutral: $P_{comp} \geq \bar{P}_{comp} - 0.2\sigma_{P_{comp}}$ or $P_{comp} \leq \bar{P}_{comp} + 0.2\sigma_{P_{comp}}$
Trading Positions	daily	-	1 represents buy, and -1 represents sell

Table 4: Sentiment Analysis: Calculated Variables

Sentiment Analysis

Take Bitcoin (BTC) for example:

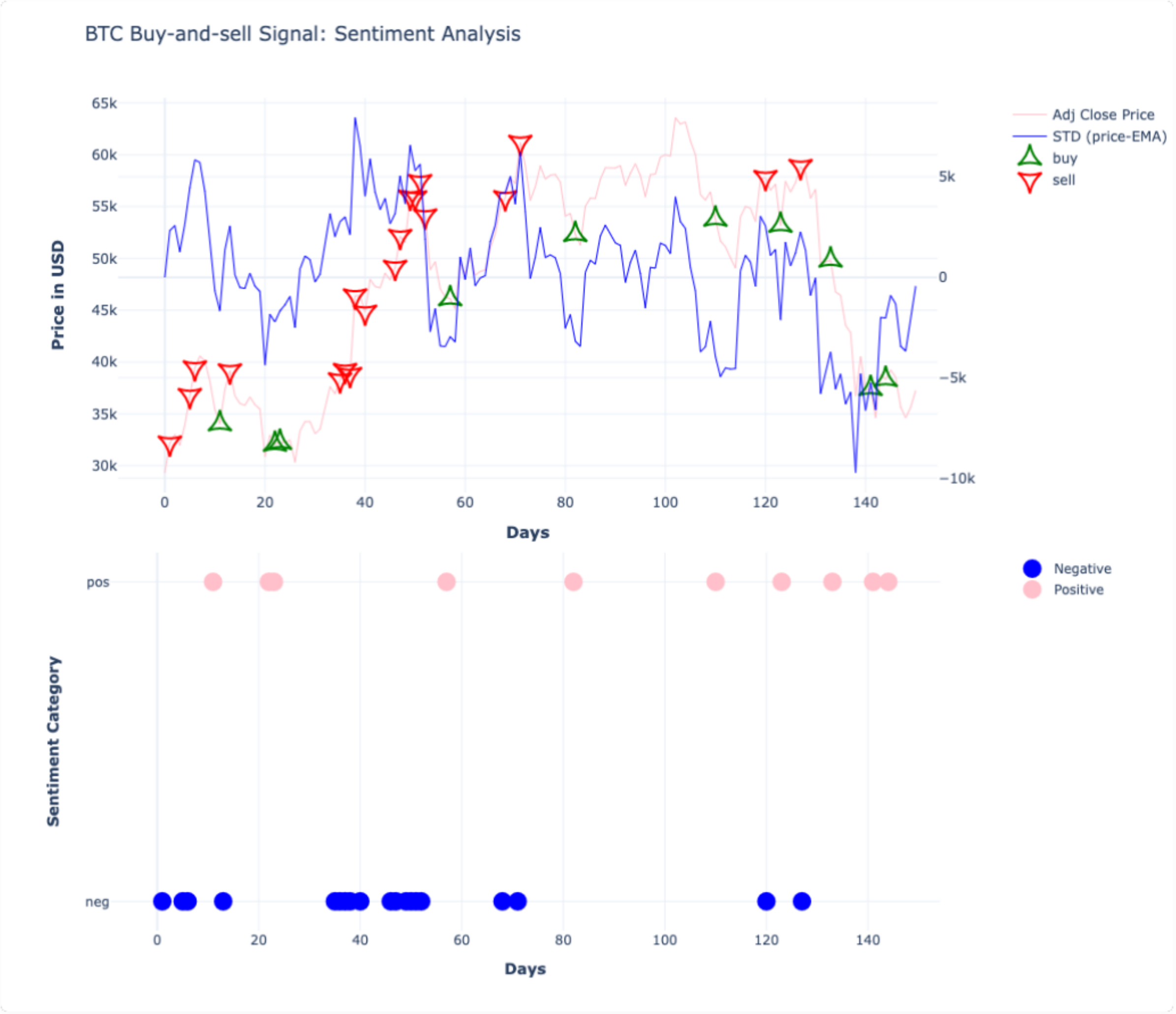


Figure 10: Buy-and-Sell Signal: BTC Sentiment Analysis

Sentiment Analysis

Take Bitcoin (BTC) for example:

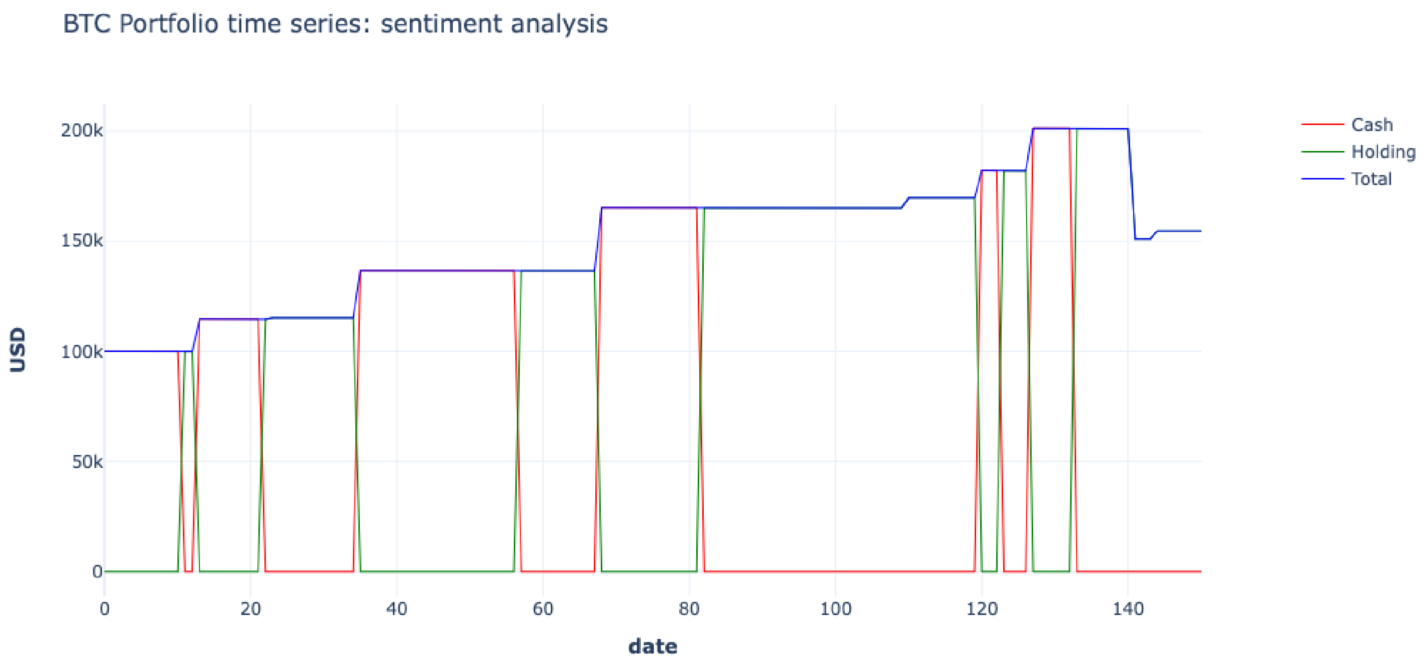


Figure 11: Portfolio time series: BTC Sentiment Analysis

Sentiment Analysis

Take Bitcoin (BTC) for example:

Figure 12: Gross ROI: BTC Sentiment Analysis vs Buy-And-Hold

Sentiment strategy ROI: 54.57%
Buy & hold strategy ROI: 27.00%

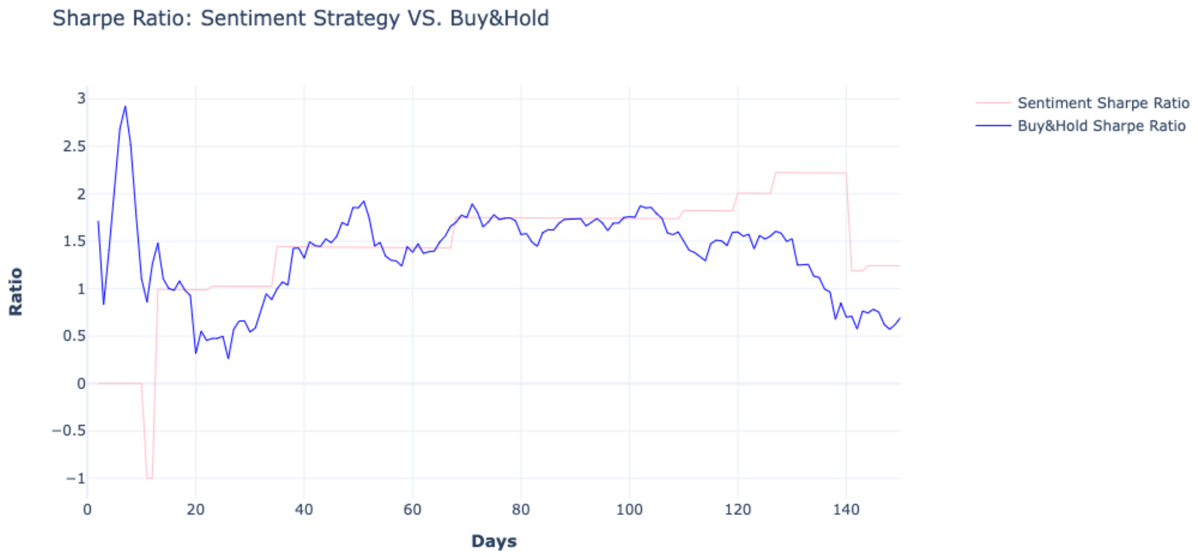


Figure 13: Sharpe Ratio: BTC Sentiment Analysis vs Buy-And-Hold

Sentiment strategy Sharpe Ratio: 1.24
Buy & hold strategy Sharpe Ratio: 0.69

Pairs Trading

Variable name	Frequency	Unit
Date	daily	YYYY-MM-DD
Close	daily	USD
Asset Name	-	-

Table 5: Pairs Trading: Data
Data Source: Alpha Vantage API

Variable name	Description
P-value	P-value from Engle and Granger's cointegration test
Hurst Exponent	The Hurst exponent is a value between 0 and 1 that serves to evaluate the mean-reversion property of a time-series
Half-life	A value λ that tells how long it takes for a time series to mean-revert
Times Crossing the Mean	Times the spread goes across the mean during the formation period

Table 6: Pairs Trading: Calculated Variables

Threshold	Value Using $\mathrm{S}_{\mathrm{simple}}$	Value Using $\mathrm{S}_{\mathrm{normal}}$
Short	2σ	$\mu + 2\sigma$
Buy	-2σ	$\mu - 2\sigma$
Exit	0	μ

Table 7: Pairs Trading: Trading Signals

Pairs Trading

Take Binance USD Token (BUSD) & EOS.IO Token (EOS) pair for example:

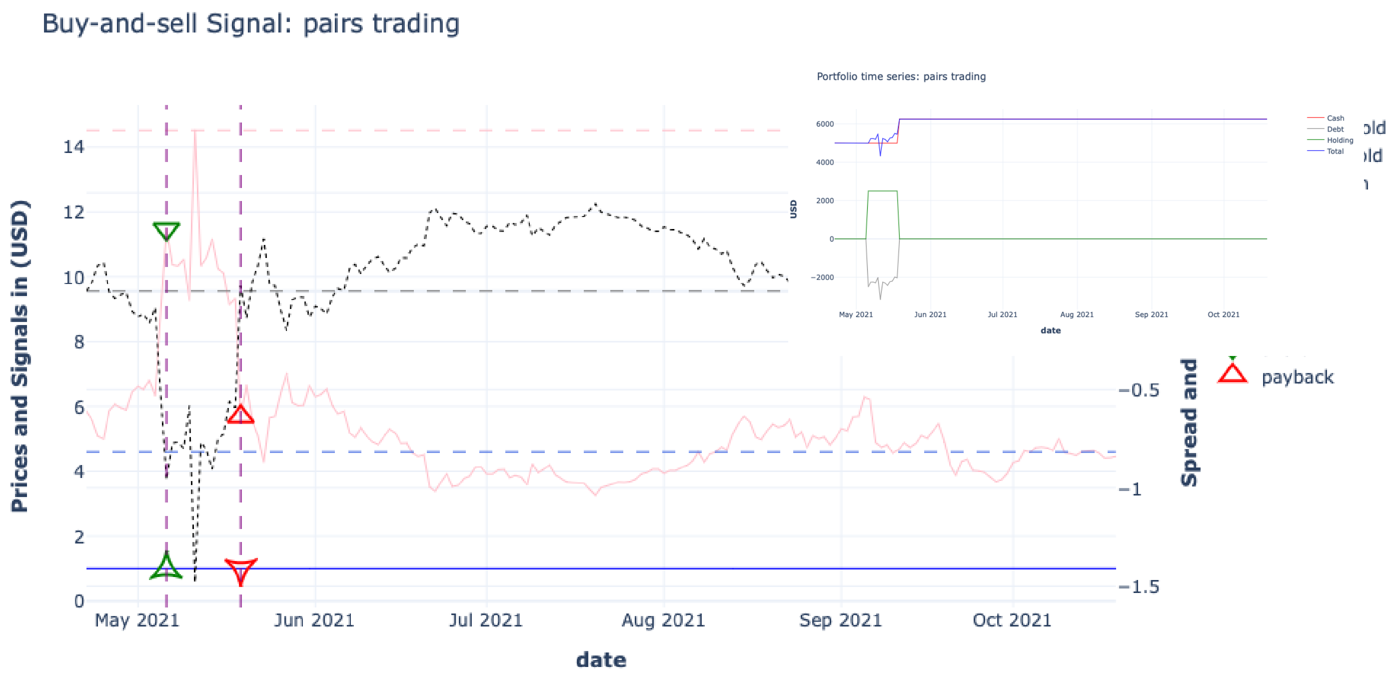


Figure 14: Buy-and-Sell Signal: BUSD & EOS Pair Trading

Pairs Trading

Take Binance USD Token (BUSD) & EOS.IO Token (EOS) pair for example:

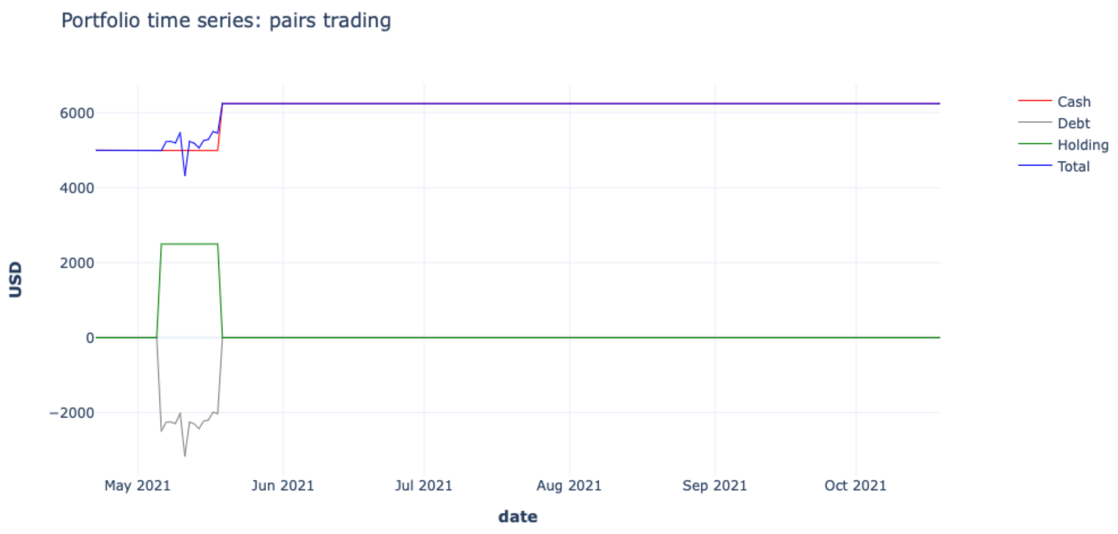


Figure 15: BUSD & EOS Portfolio time series: Pairs Trading

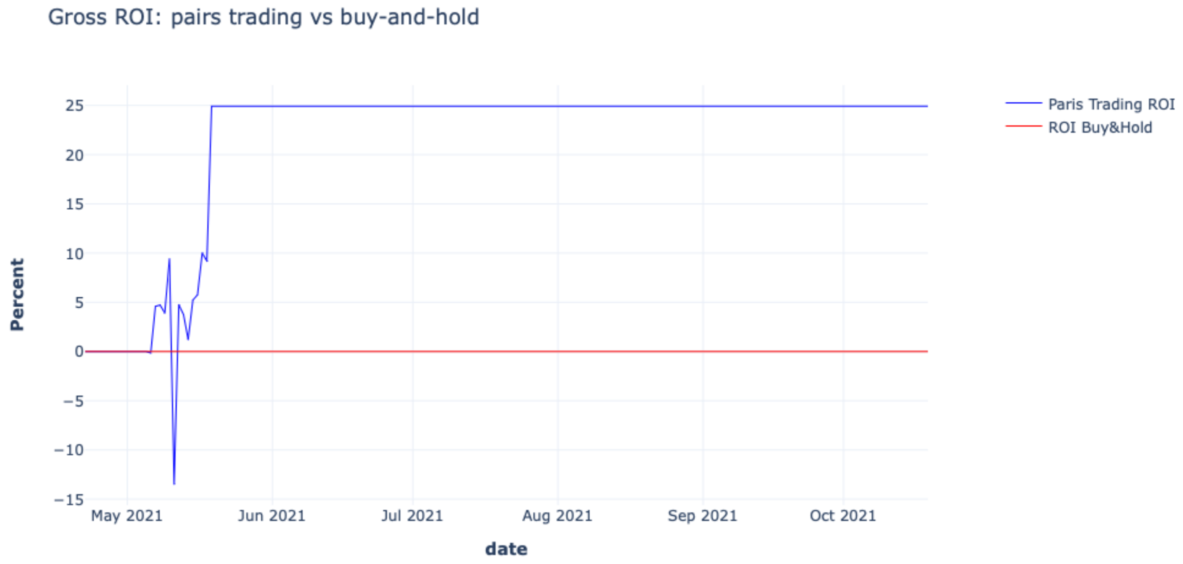


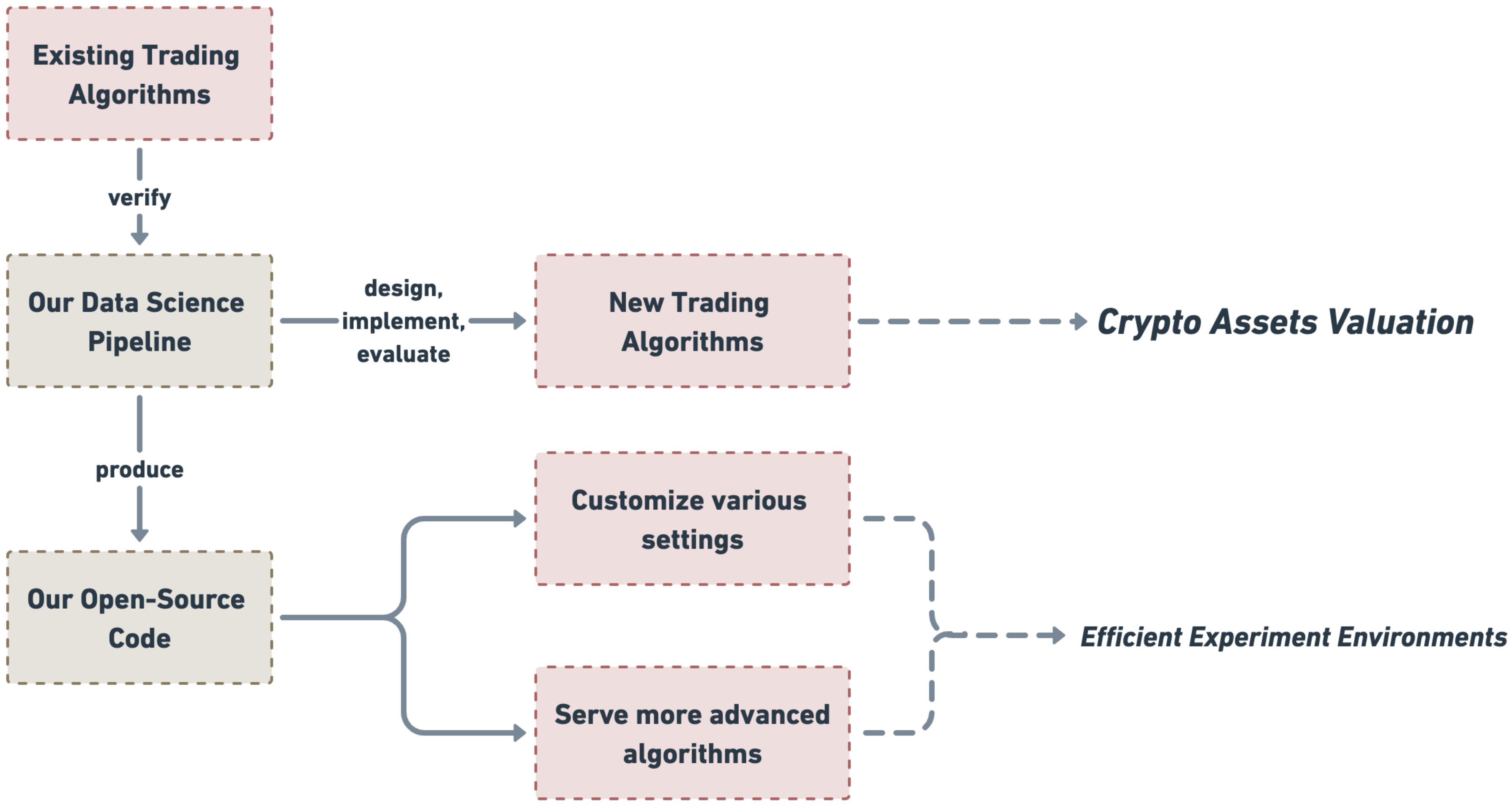
Figure 16: BUSD & EOS Gross ROI: Pairs Trading vs buy-and-hold

Pairs Trading strategy ROI: 24.92%
Buy & hold strategy ROI: 0.00%

3

Future Research and Practices

Future Research and Practices



https://whimsical.com/fig2-N8PAxb9LxMDSoeW4yyzrdF

Thank you

Working Paper: https://arxiv.org/abs/2206.14932

Data and Code: https://github.com/sciecon/srs2021

Corresponding to Luyao (Sunshine) Zhang:
lz183@duke.edu
yulinzurich@gmail.com

Duke Scholar Page for Sunshine:
http://scholars.duke.edu/person/luyao.zhang

Industry 4.0 Open Educational Resource Initiatives:
https://ie.pubpub.org/
https://ce.pubpub.org/
https://ie.pubpub.org/