



**Trinity College Dublin**

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

# Sentiment analysis with Large Language Models for predicting trends in Bitcoin

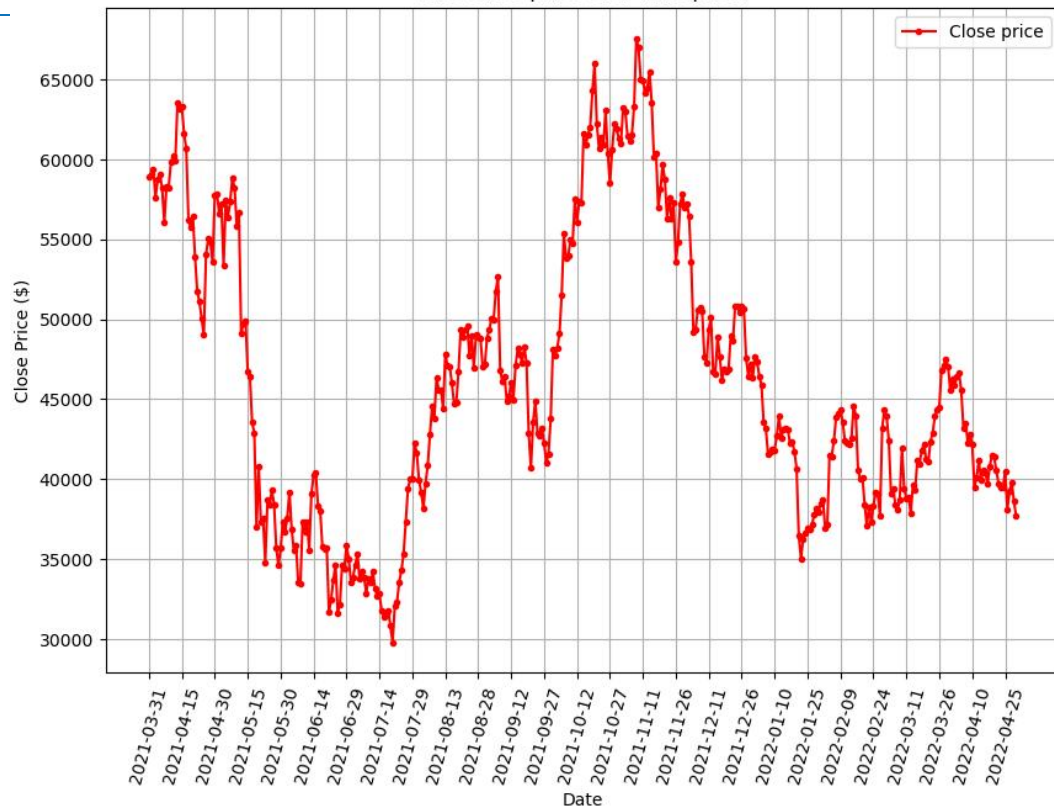
Ziang Liu 24333476

Supervisor: Prof. Anil Kokaram, Dr. Claire Gallagher

May.21 2025

# Introduction and Aim

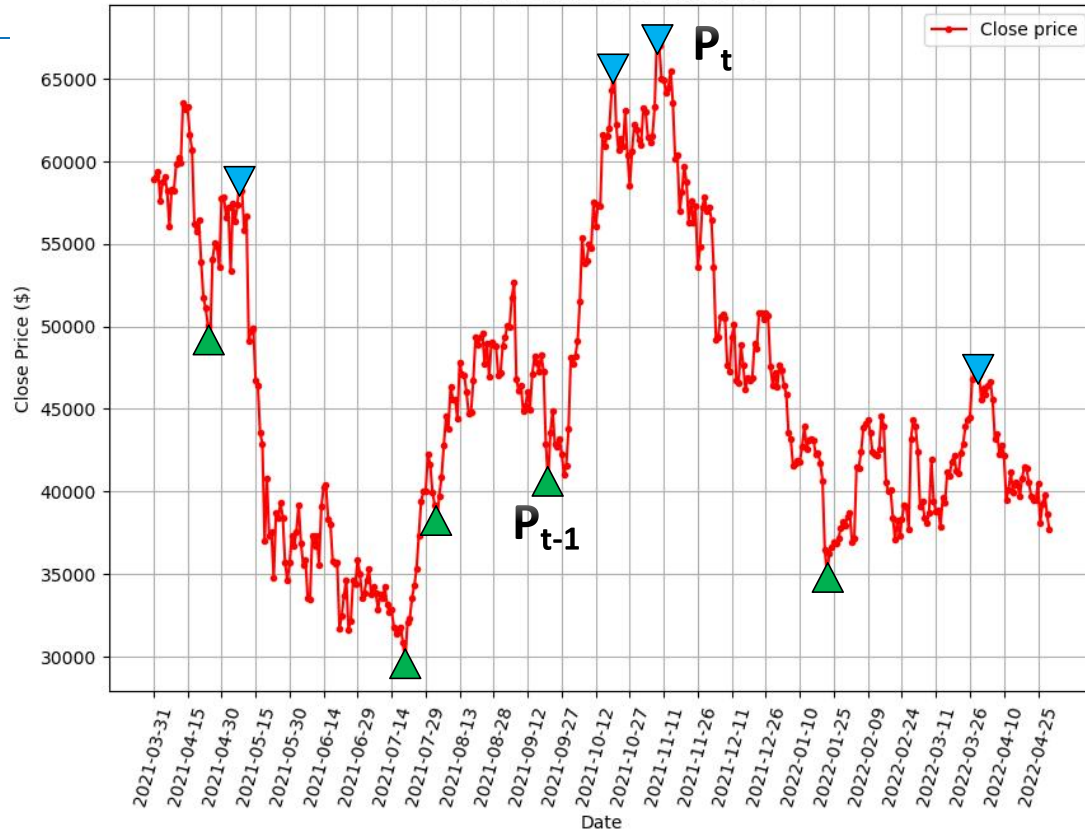
BTC Close price over date priod



Date	News title
2021-3-31	Iphone user loses \$ 600,000 downloading scam bitcoin app apple store
2021-3-31	New app help previously incarcerated individual find job
2021-3-31	Bitcoin ether hit highest since mid - may sentiment warms
2021-4-1	Rise bot trading world cryptocurrency
2021-4-2	Art nft question : next frontier trading , new form tulip ?
.....	.....

# Introduction and Aim

BTC Close price over date period



## Trading signal

An actionable indicator that suggests when to buy, sell, or hold an asset.

▲ Buy signal

▼ Sell signal

## Return

$P_t$  Price at time  $t$

$P_{t-1}$  Price at the previous time step.

$r_t$  Return

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

# Introduction and Aim

---

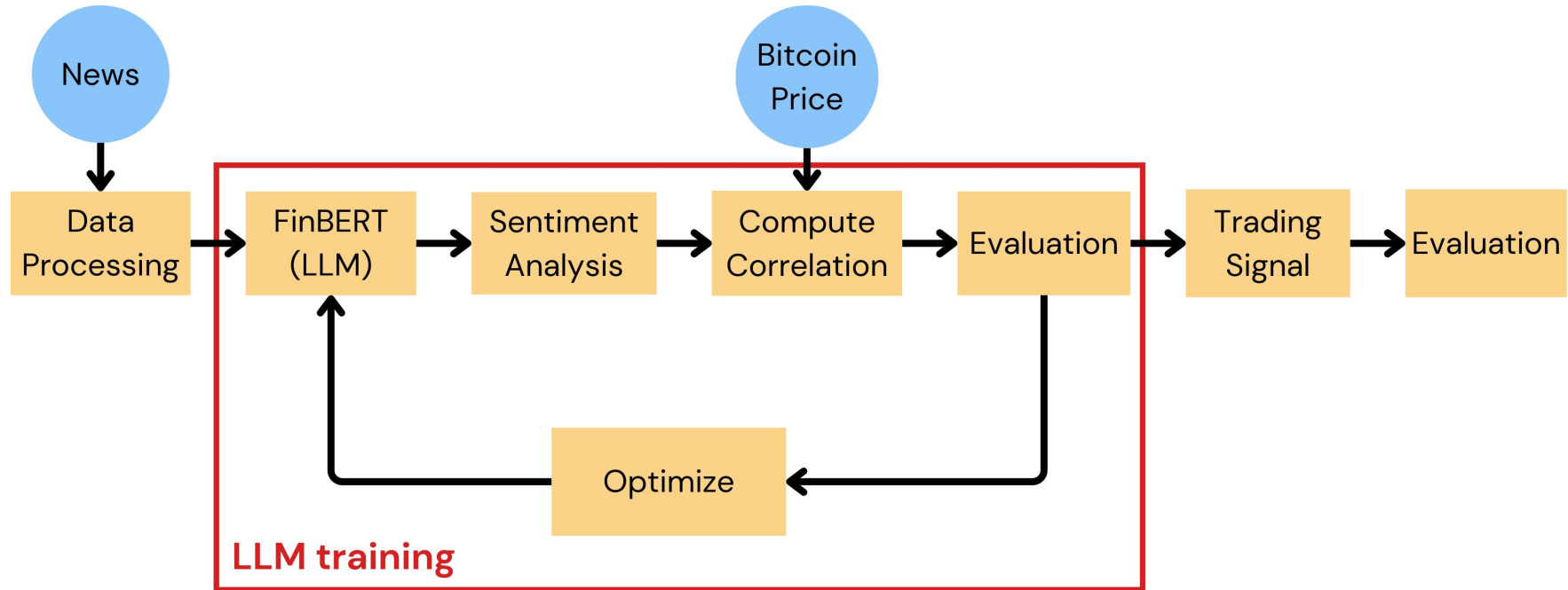
- This work explores the use of financial **news headlines** for **predicting bitcoin price trends** through automated sentiment analysis.
- Sentiment analysis performance is evaluated by generating **trading signals** based on sentiment and measuring the resulting **returns**.
- Integrates machine learning theory with real-world business data
- Better asset management, Algorithmic trading

# Objectives

---

- Review existing literature on LLMs in financial predictions
- Replicate FinBERT(LLM)'s performance on the GDELT news database
- Retrain and implement FinBERT to improve sentiment analysis accuracy
- Evaluate FinBERT's performance using statistical and financial metrics
- Calculate correlation between sentiment and returns
- Develop a trading signal based on sentiment analysis

# Method



# Data

## News

243504 **News** titles were extracted from **GDELT** between March 2021 and April 2022 containing any of the key words: cryptocurrency, cryptocurrencies, CBDC, Bitcoin, Ethereum, Litecoin, BitcoinCash...[1]

**Global Data on Events, Location and Tone (GDELT)** contains over 200 million geolocated events with worldwide coverage from 1979 to the present.

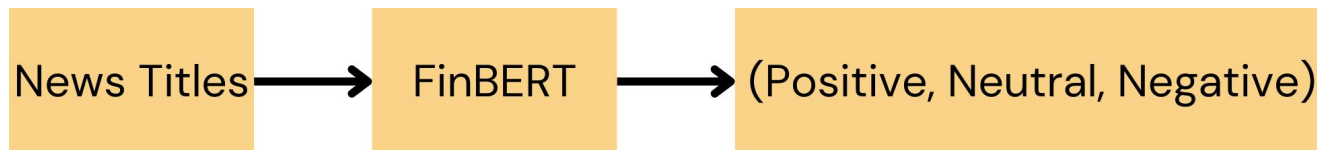
## Bitcoin Price

**Bitcoin price** is downloaded from **CoinMarketCap** covering the same date period.



# FinBERT - Large Language Model

FinBERT is trained over 61 GB of text data, with approximately 12.7 billion words from financial sources and 3.3 billion words from general sources.



ethereum expected hit new high

$$S_{\text{pos}} = 0.98$$

$$S_{\text{neu}} = 0.01$$

$$S_{\text{neg}} = 0.01$$





# Retrain FinBERT

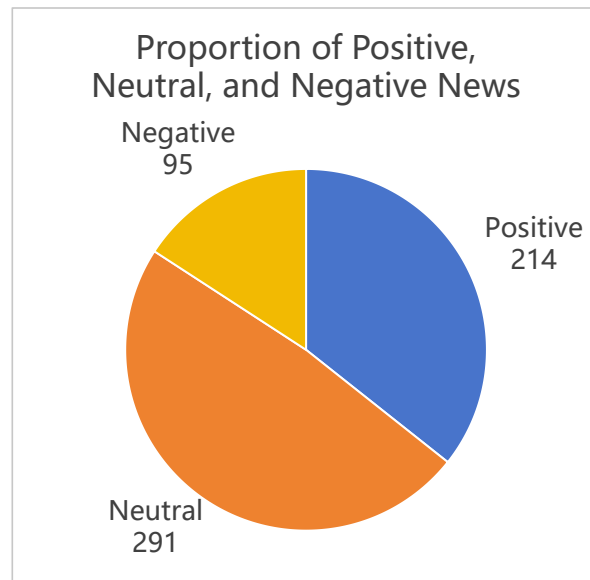
The original FinBERT model has an accuracy of **64.8%** on a manually labeled GDELT news dataset consisting of 600 news articles.

## Retrain:

1. Remove the stop words. ("the", "and", "is")
2. Remove punctuation and special characters
3. Convert all text to lower case
4. lemmatizes each word

Train	70%	Learning rate	2e-5
Validation	20%	Batch size	8
Test	10%	Weight decay	0.01

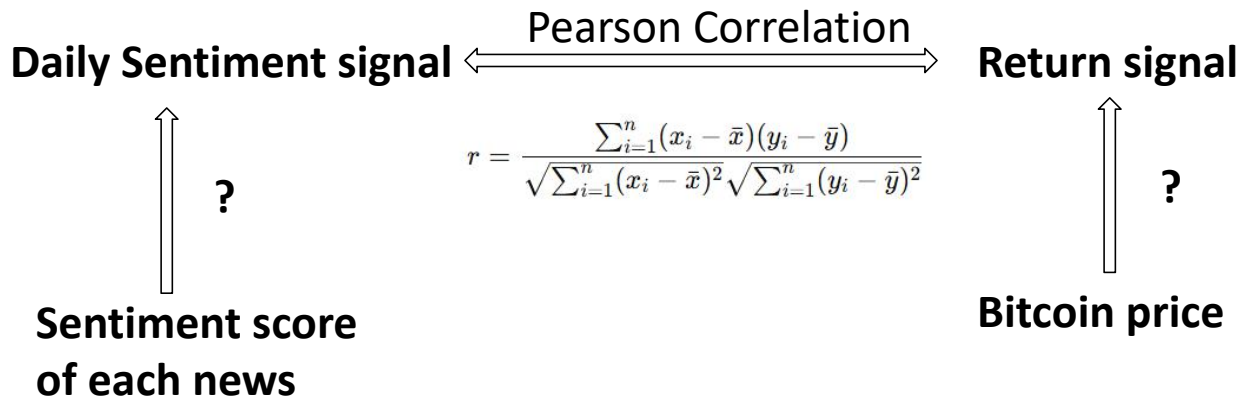
Retrained FinBERT has **85%** accuracy on test set



# Sentiment Analysis

To build the trading signal, we need to understand the relationship between the **Sentiment signal** and the **Return signal**

**Pearson correlation** can represent the strength and direction of the signals linear relationship.



# Sentiment Analysis

To compute daily sentiment, we aggregate individual news sentiment scores by day

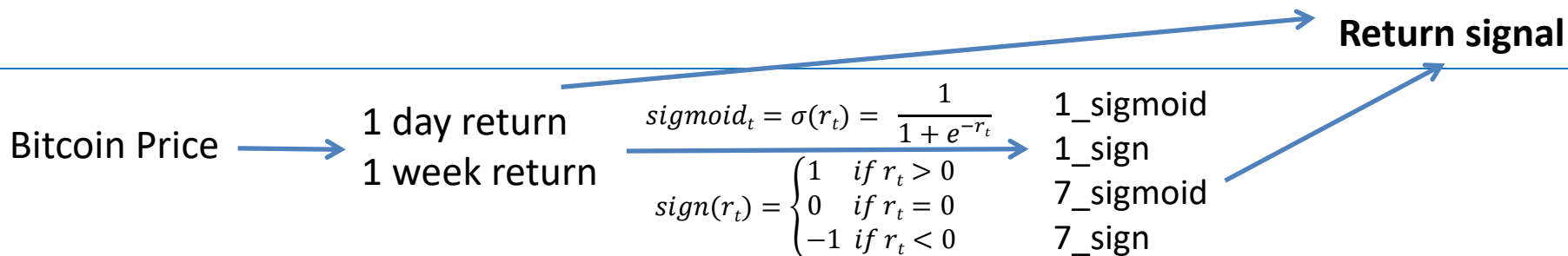
$$\begin{array}{l}
 S_{pos} = 0.98 \\
 S_{neu} = 0.01 \\
 S_{neg} = 0.01
 \end{array}
 \quad
 \begin{array}{l}
 \text{discrete\_score} = \begin{cases} 1 & \text{if } \max(S_{pos}, S_{neu}, S_{neg}) = S_{pos} \\ 0 & \text{if } \max(S_{pos}, S_{neu}, S_{neg}) = S_{neu} \\ -1 & \text{if } \max(S_{pos}, S_{neu}, S_{neg}) = S_{neg} \end{cases} \\
 \text{continuous\_score} = -1 * S_{neg} + 0 * S_{neu} + 1 * S_{pos}
 \end{array}
 \rightarrow
 \begin{array}{l}
 \text{discrete score} = 1 \\
 \text{continuous score} = 0.97
 \end{array}$$

$$\begin{array}{l}
 \text{compute daily sentiment score} \rightarrow \begin{array}{l} \text{discrete\_sum} = -8 \\ \text{discrete\_mean} = -0.44 \\ \text{continuous\_sum} = -7.6 \\ \text{continuous\_mean} = -0.43 \end{array}
 \end{array}
 \quad
 \begin{array}{l}
 \text{sigmoid}_t = \sigma(r_t) = \frac{1}{1 + e^{-r_t}} \\
 \text{sign}(r_t) = \begin{cases} 1 & \text{if } r_t > 0 \\ 0 & \text{if } r_t = 0 \\ -1 & \text{if } r_t < 0 \end{cases}
 \end{array}
 \rightarrow
 \begin{array}{l}
 \text{discrete\_sum\_sign} = -1 \\
 \text{continuous\_sum\_sigmoid} = 0.0005
 \end{array}$$

**Daily sentiment signal**



# Pearson Correlation



Return signal

Daily sentiment signal

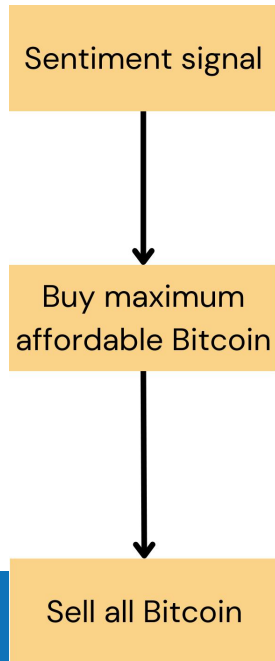
Metric	1-day	1-week	1_sigmoid	7_sigmoid	1_sign	7_sign
discrete_sum	0.142	-0.063	0.070	-0.024	0.072	-0.045
discrete_mean	0.116	-0.059	0.046	0.016	0.052	0.001
continuous_sum	0.135	-0.067	0.069	-0.019	0.077	-0.053
continuous_mean	<b>0.185</b>	-0.053	0.102	0.036	0.101	0.011
discrete_sum_sign	-0.055	-0.121	-0.108	-0.164	-0.100	<u>-0.182</u>
continuous_sum_sigmoid	0.015	-0.063	-0.052	0.027	-0.035	-0.003



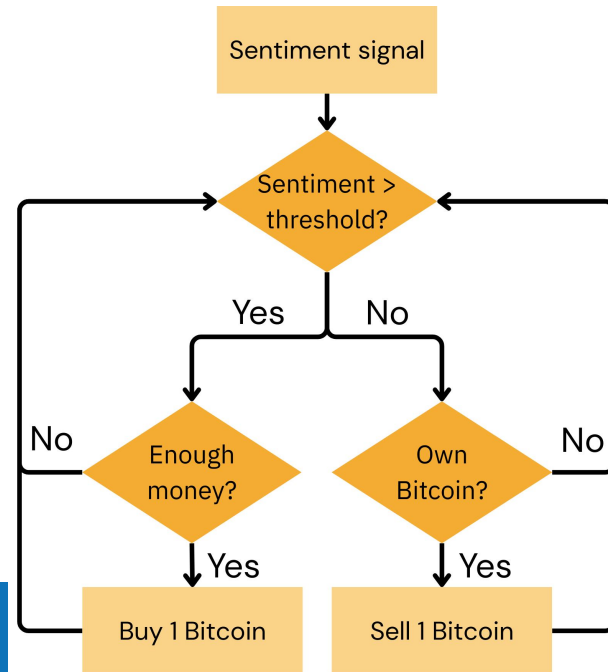
# Development of trading algorithm

*continuous\_mean, Threshold (= 0.5), Start money (= €1,000,000)*

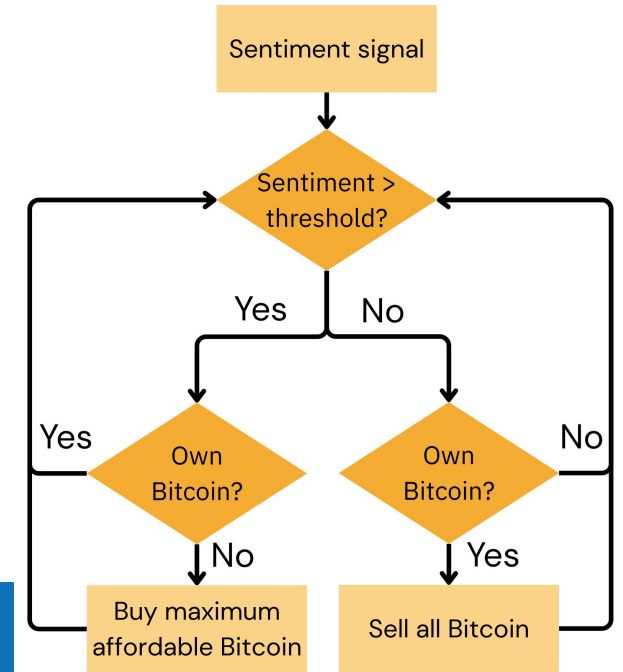
- **Buy-and-hold Strategy**



- **Sentiment Strategy 1**



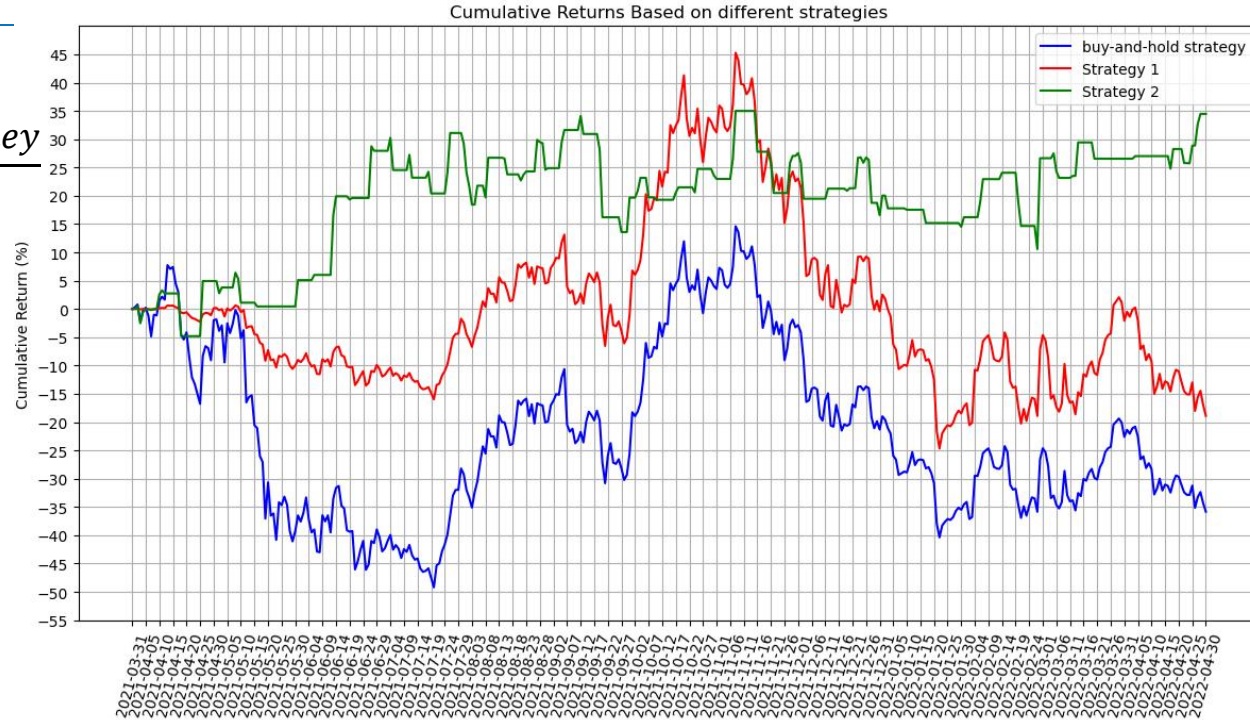
- **Sentiment Strategy 2**



# Performance

$$\text{Return} = \frac{\text{Current value} - \text{Start money}}{\text{Start money}}$$

Current value = Current money +  
BTC valued at current price



# Limitation

- The price signal is highly noisy and influenced by multiple factors beyond sentiment
- The dataset used to retrain FinBERT is imbalanced
- The accuracy of the retrained FinBERT model remains a concern
- Current sentiment strategy could be improved
- Does not account for real-time price changes, transaction fees, and other trading costs.



# Future work

---

- Collect a more balanced and comprehensive news dataset
- Consider other factors such as moving average, interest rate etc.
- Enhance the performance of the retrained FinBERT model
- Optimize the trading signal strategy for better return







**Trinity College Dublin**

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

# Thank You

