HMM-BASED
TRADING MASTER

PRESENTED BY: EXPLAINABLE AI BOY BAND

MARKET DATA

Dataset

Scraping from yahoo finance API

Data preparation

Add column to check if in each day **close price is more than open price**

Date	Close	Open	X_t
#########	1.592667	1.266667	1
#########	1.588667	1.719333	0
1/7/2010	1.464	1.666667	0
2/7/2010	1.28	1.533333	0
6/7/2010	1.074	1.333333	0

We use stock price ‘UP’ or ‘DOWN’ as state variable (X_t)

TWEET DATA

Dataset

Stock tweets for sentiment Analysis and Prediction from **Kaggle**

Data preparation

We use tweet sentiment ‘Positive’ , ‘Negative’ , ‘Neutral as evidence variable (E_t)

Tweet [Full text of tweet] → **Sentiment Analysis model** → **classified tweet into**
- **positive**
- **negative**
- **neutral**

cardiffnlp/twitter-roberta-base-sentiment-latest

Data preparation

Problem

Due to the less amount of negative tweet , if we choose the sentiment to represent in a day with majority vote, it will not have the negative tweet in the state variable

Result

We scale the negative tweet with weighted sum by

1. Check the average number of positive tweets thats cause the stock price to go UP **5** days in a row named with **nPos_streak**

2. Check the average number of negative tweets thats cause the stock price to go DOWN **5** days in a row named with **nNeg_streak**

3. Calculate Majority vote =
MAX(num_positive_tweets , num_negative_tweets * **nPos_streak/nNeg_streak** , num_neutral_tweets)

HMM
FILTERING / PREDICTION
INFERENCES

TERMINOLOGY

x_t ; {"up", "down"}

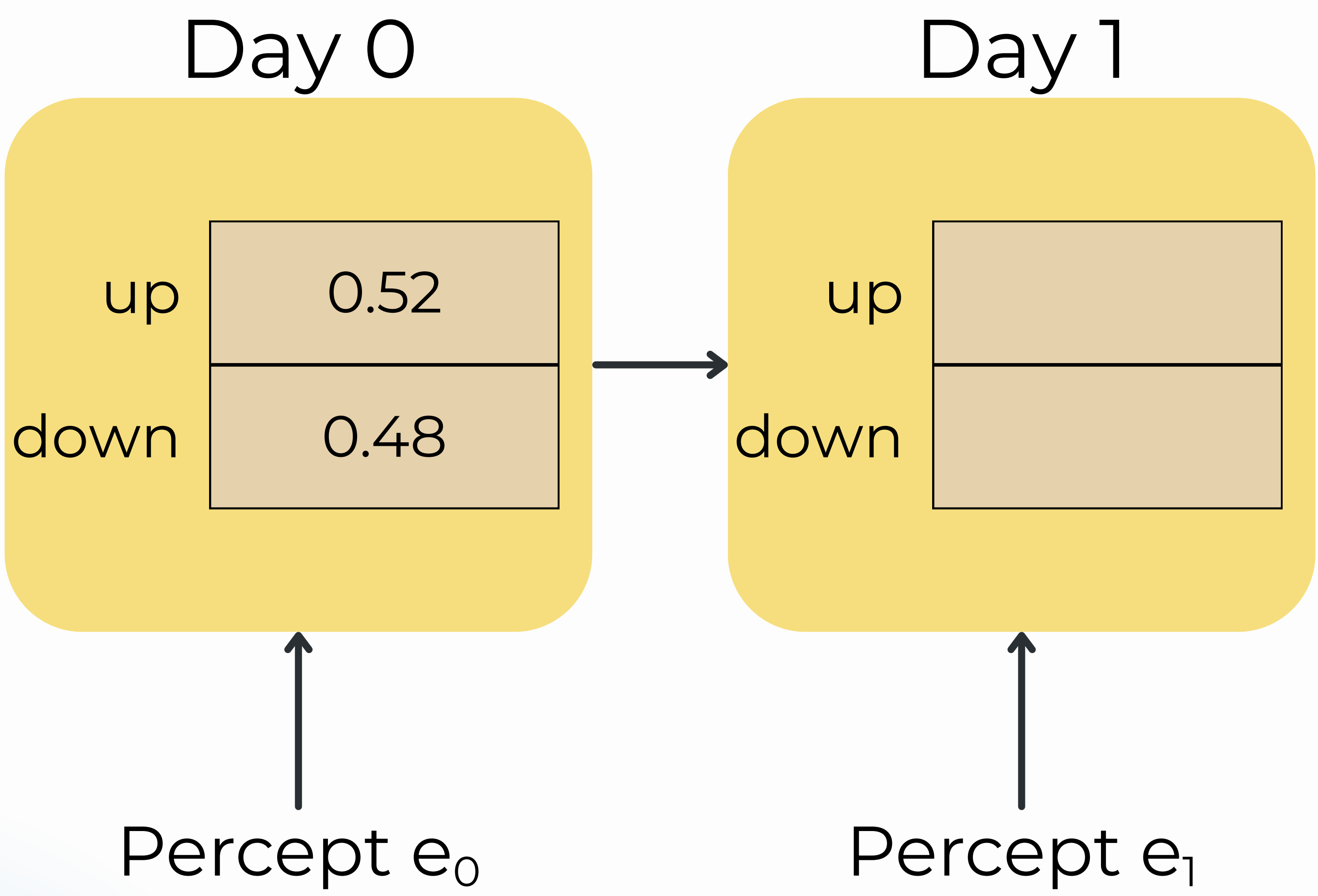
- the stock price is up in day t*

e_t : {"positive", "negative", "neutral"}

- the major** sentiment of tweet in day t

*price is up → Close price > Open price, vise versa

FILTERING

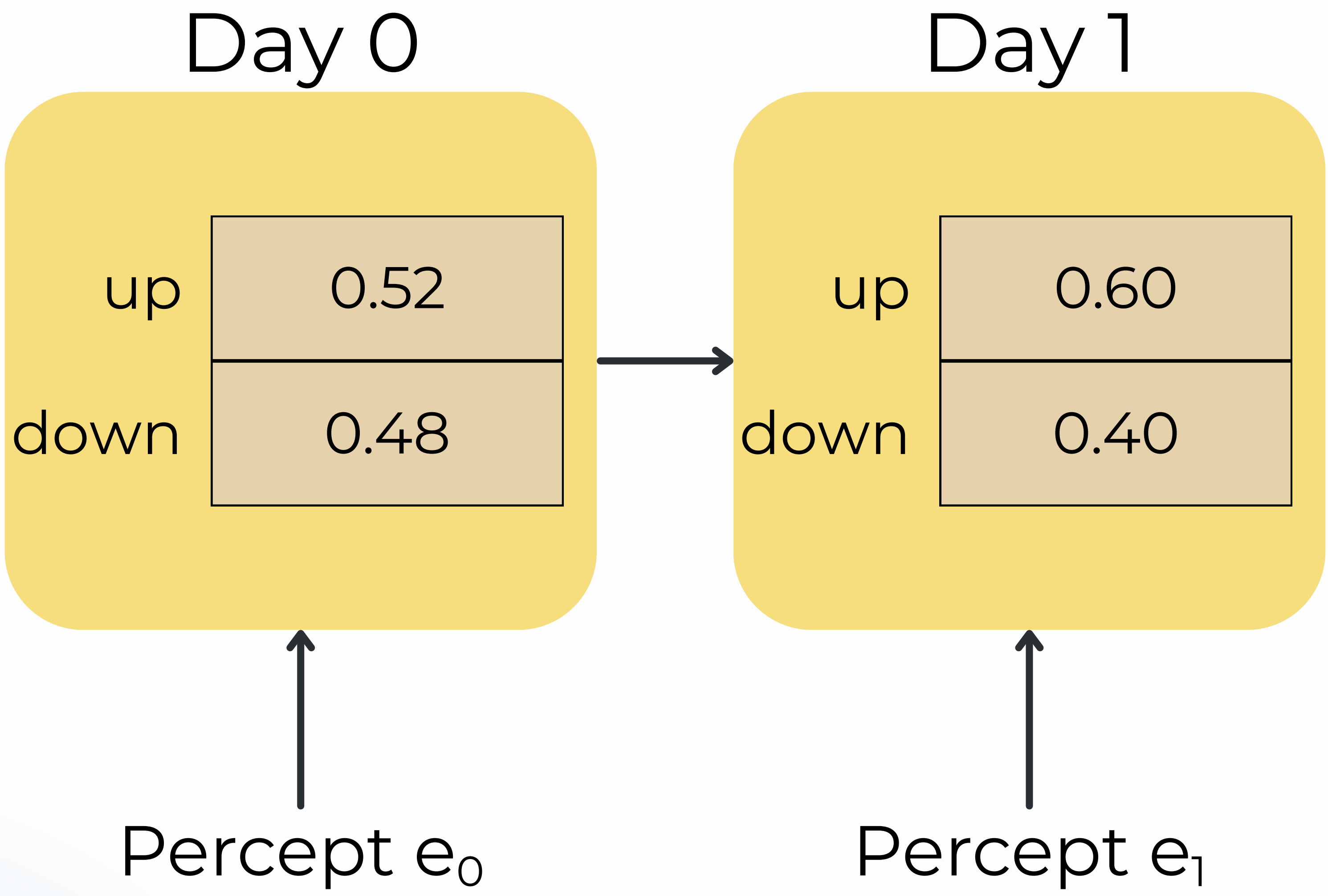


Use $\mathbf{P(x_0)}$ at day 0 and $\mathbf{e_0}$ to predict $\mathbf{P(x_1)}$

$P(x_1) = \sum P(x_0) * P(e_1 \mid x_1) * P(x_1 \mid x_0)$

for x_1 = {"up", "down"}

FILTERING (CONT.)

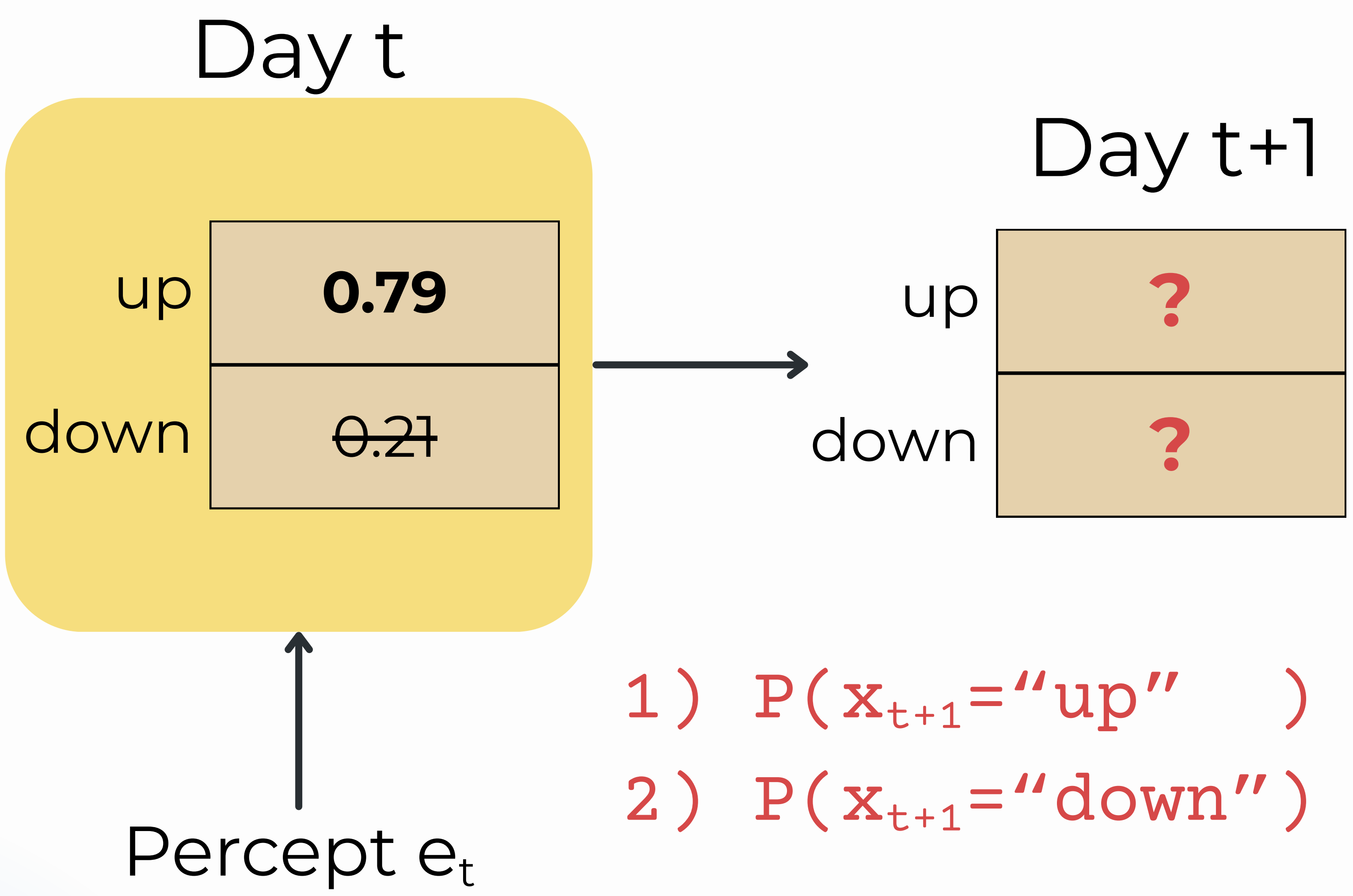


Use **$P(x_0)$** at day 0 and **e_0** to predict **$P(x_1)$**

$P(x_1) = \sum P(x_0) * P(e_1 \mid x_1) * P(x_1 \mid x_0)$

for x_1 = {"up", "down"}

PREDICTION

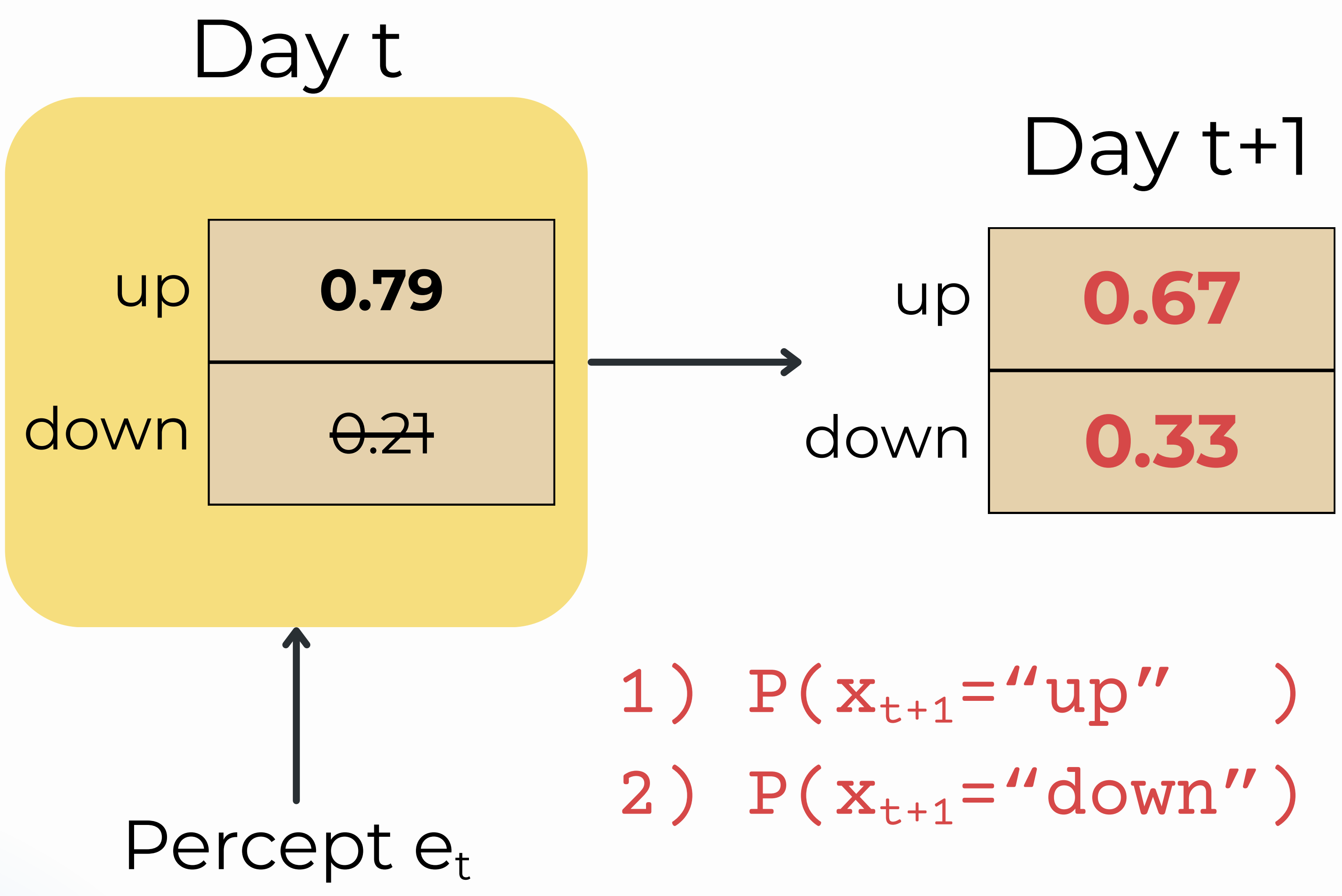


We select case that has the most likelihood ("up" in case) and times with transition prob.

1) $P(x_{t+1}=\text{"up"}) = P(x_{t+1}=\text{"up"} \mid x_t=\text{"up"}) * P(x_t=\text{"up"})$

2) $P(x_{t+1}=\text{"down"}) = P(x_{t+1}=\text{"down"} \mid x_t=\text{"up"}) * P(x_t=\text{"up"})$

PREDICTION (CONT.)



We select case that has the most likelihood ("up" in case) and times with transition prob.

1) $P(x_{t+1}=\text{"up"}) = P(x_{t+1}=\text{"up"} \mid x_t=\text{"up"}) * P(x_t=\text{"up"})$

2) $P(x_{t+1}=\text{"down"}) = P(x_{t+1}=\text{"down"} \mid x_t=\text{"up"}) * P(x_t=\text{"up"})$

RESULT & EVALUATION

Probability $P(x_t)$

x_t	**Probability**
Up	0.52
Down	0.48

Probability $P(x_t \mid x_{t-1})$

$x_t \setminus x_{t-1}$	**Up**	**Down**
Up	0.55	0.49
Down	0.45	0.51

Probability $P(e_t)$

e_t	Probability
Positive	0.31
Neutral	0.42
Negative	0.27

Probability $P(e_t \mid x_t)$

$e_t \setminus x_t$	Up	Down
Positive	0.38	0.22
Neutral	0.45	0.39
Negative	0.17	0.39

Model setting

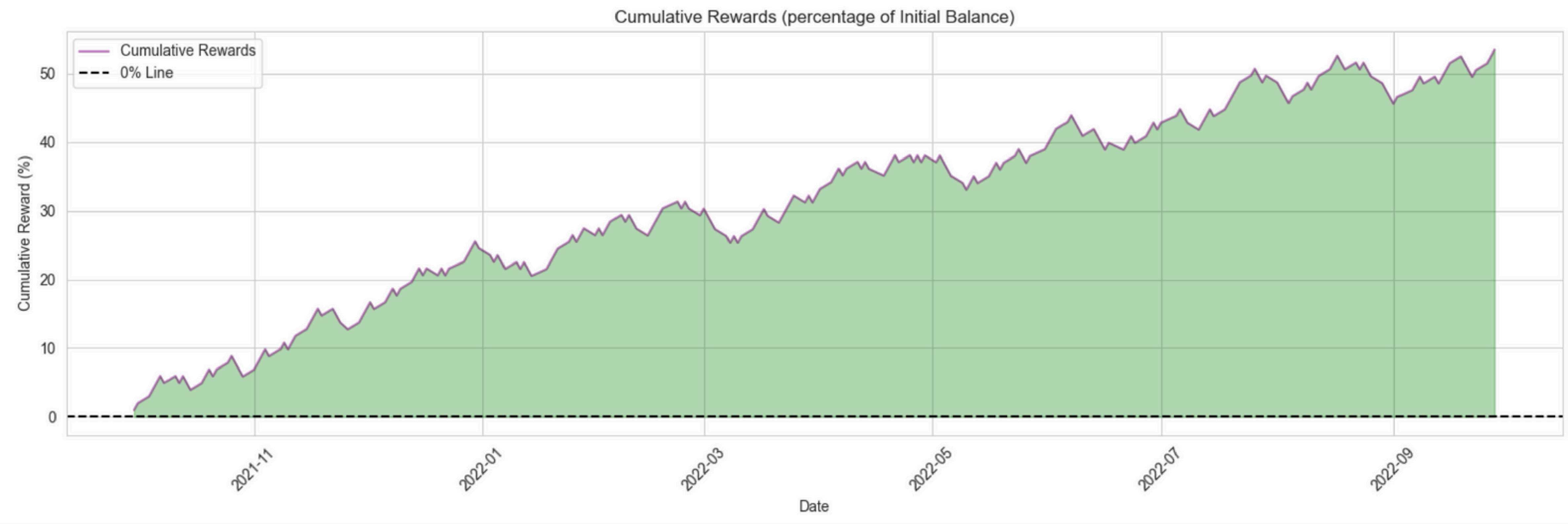
There are 4 models

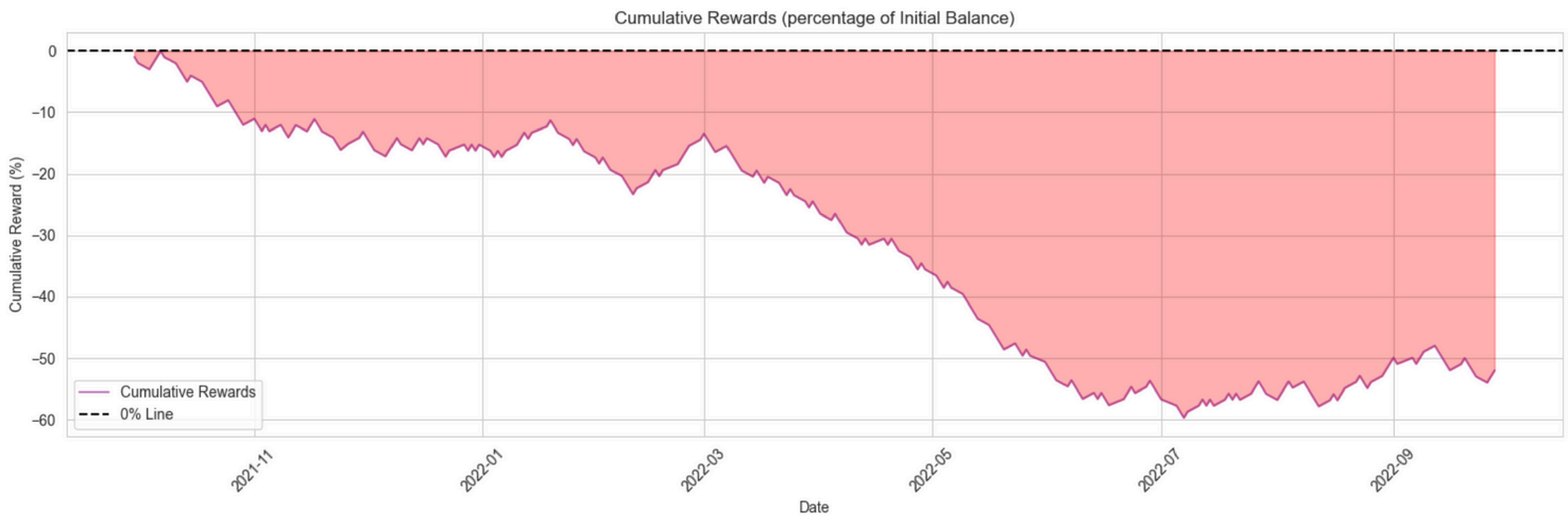
1. Random trade model
2. Only buy trade model
3. Only sell trade model
4. HMM trade model

How to trade

1. For Random, trade 3,000 time on the same period & stock
2. Has 1% commission fee

Random trade model

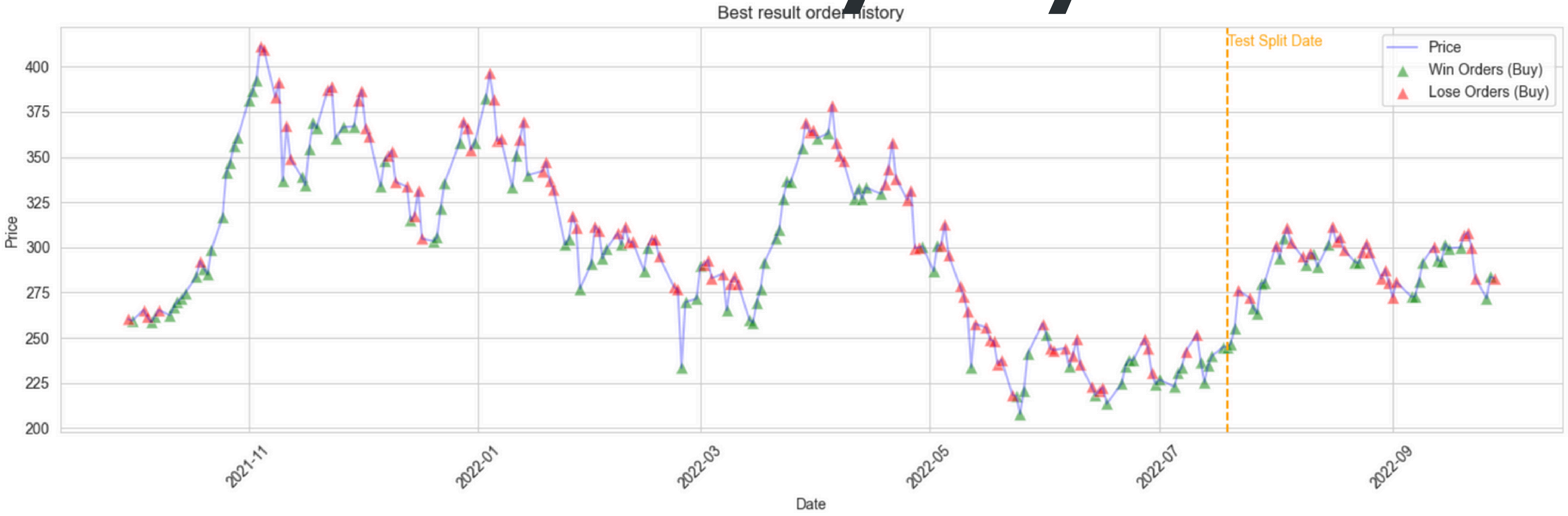


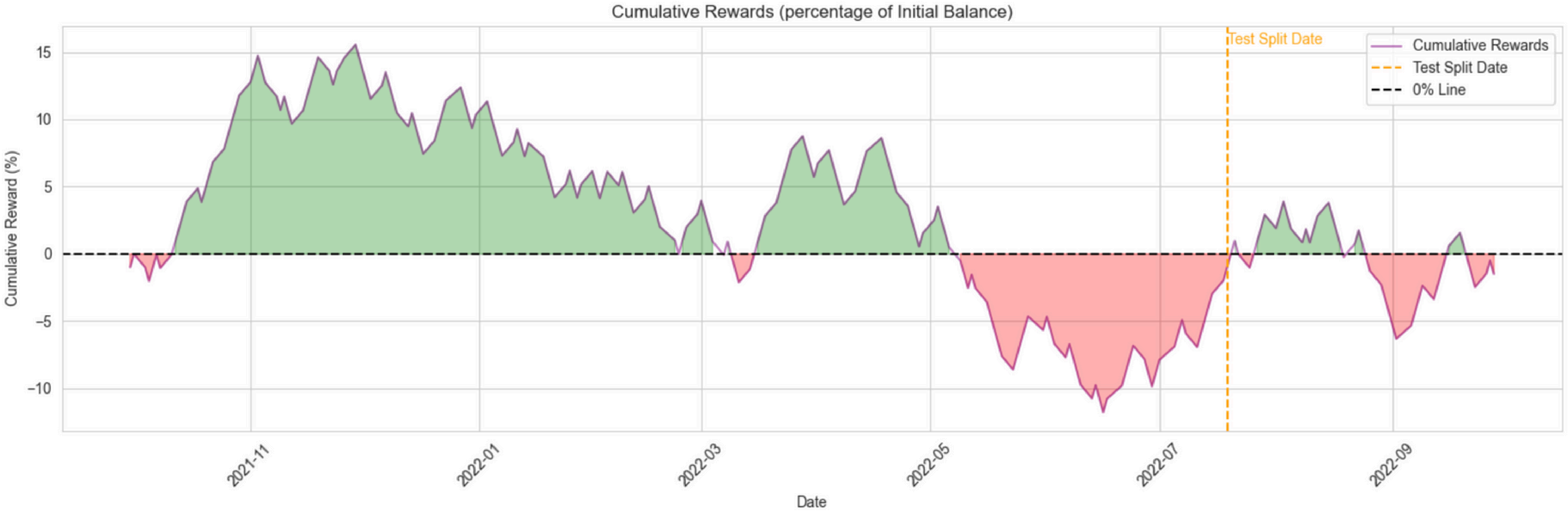


With non-deterministic behavior, we got

Average profit → -1.8% 🤨
SD → 15.86 %

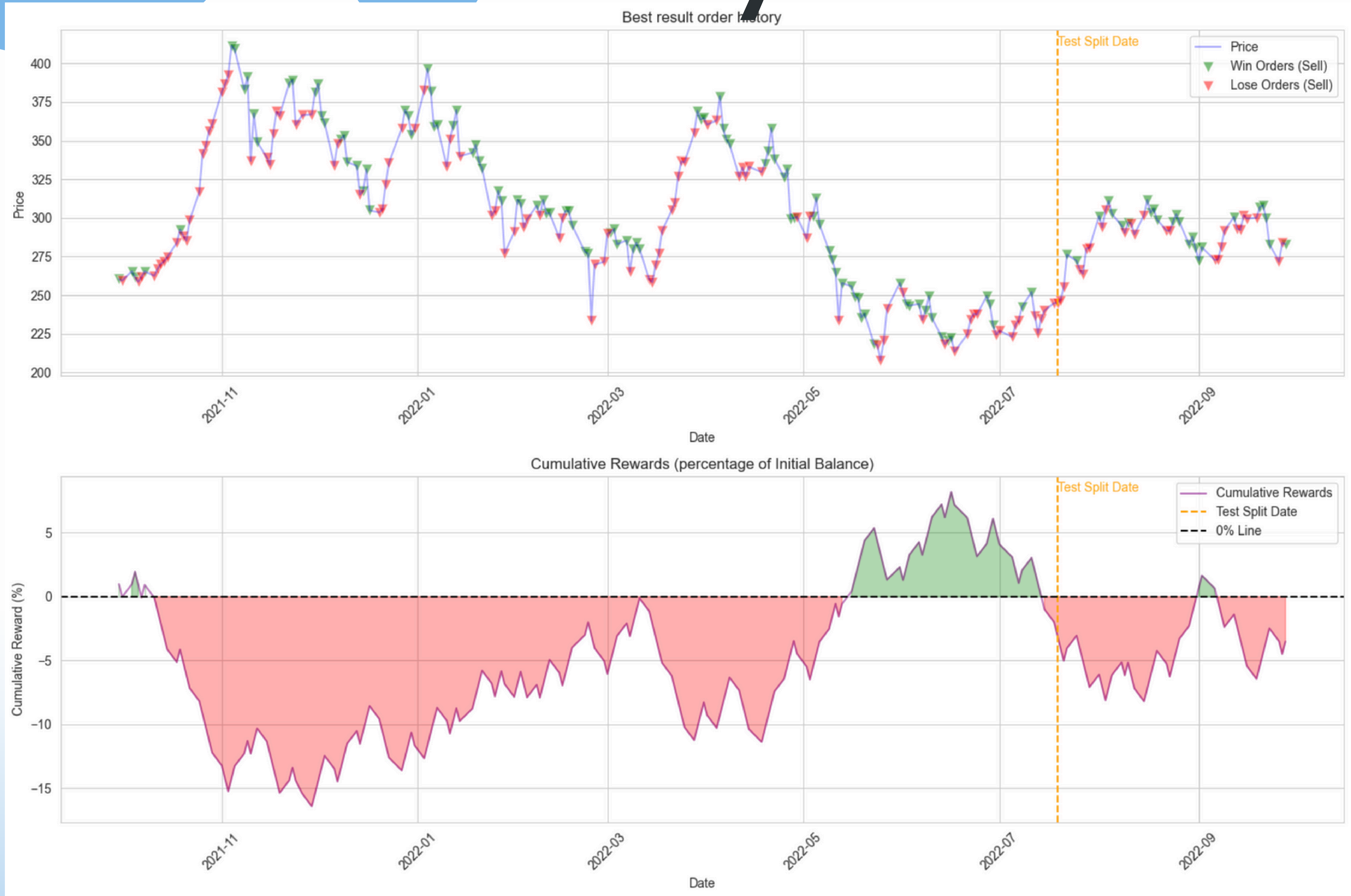
Only buy trade model





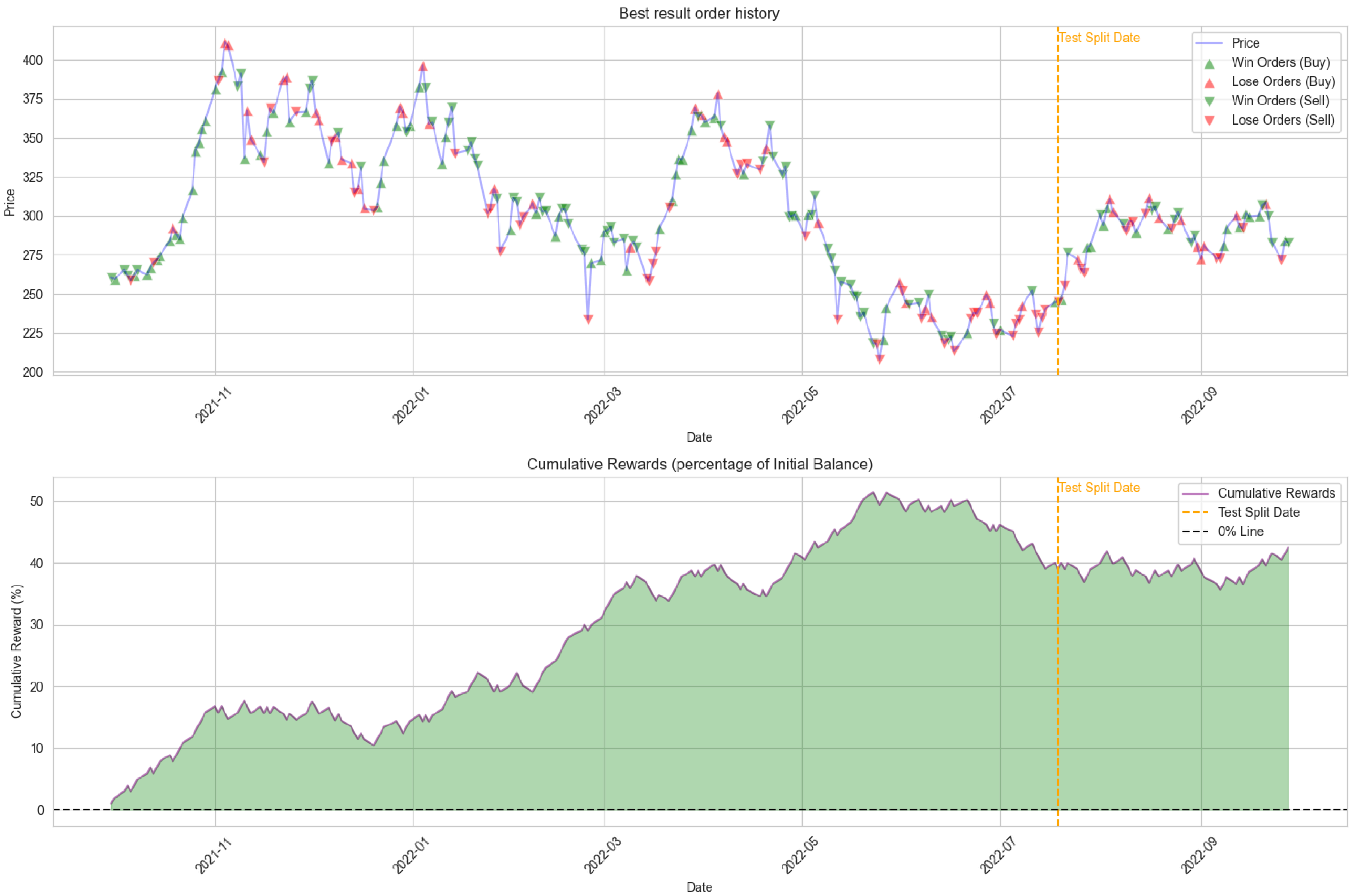
Profit rate = -1.51% 🙅

Only sell trade model



Profit rate = -3.51% 😰

HMM trade model

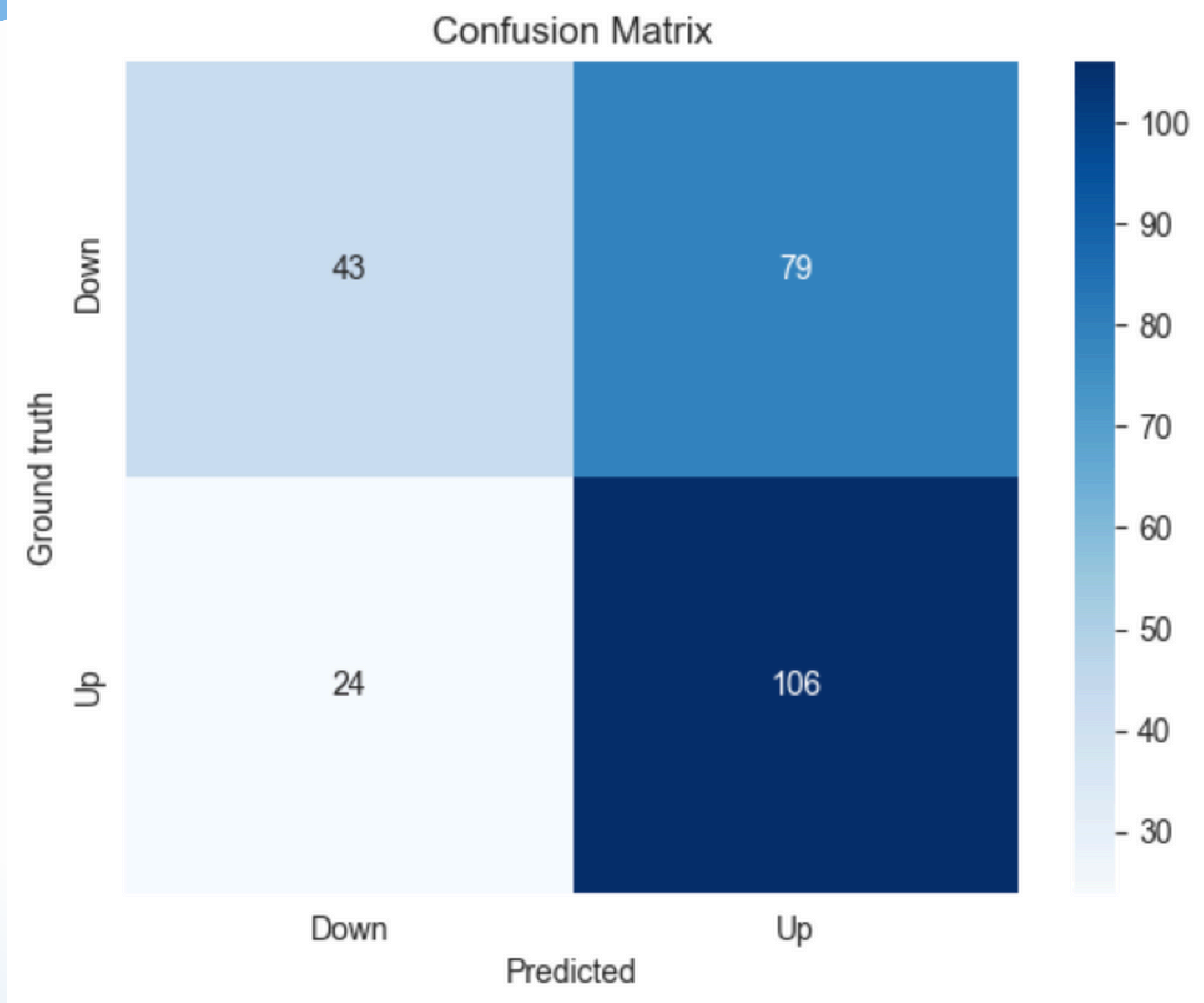


Now, we have deterministic output

So, we got

Profit rate = 42.49% 🎉

HMM trade model

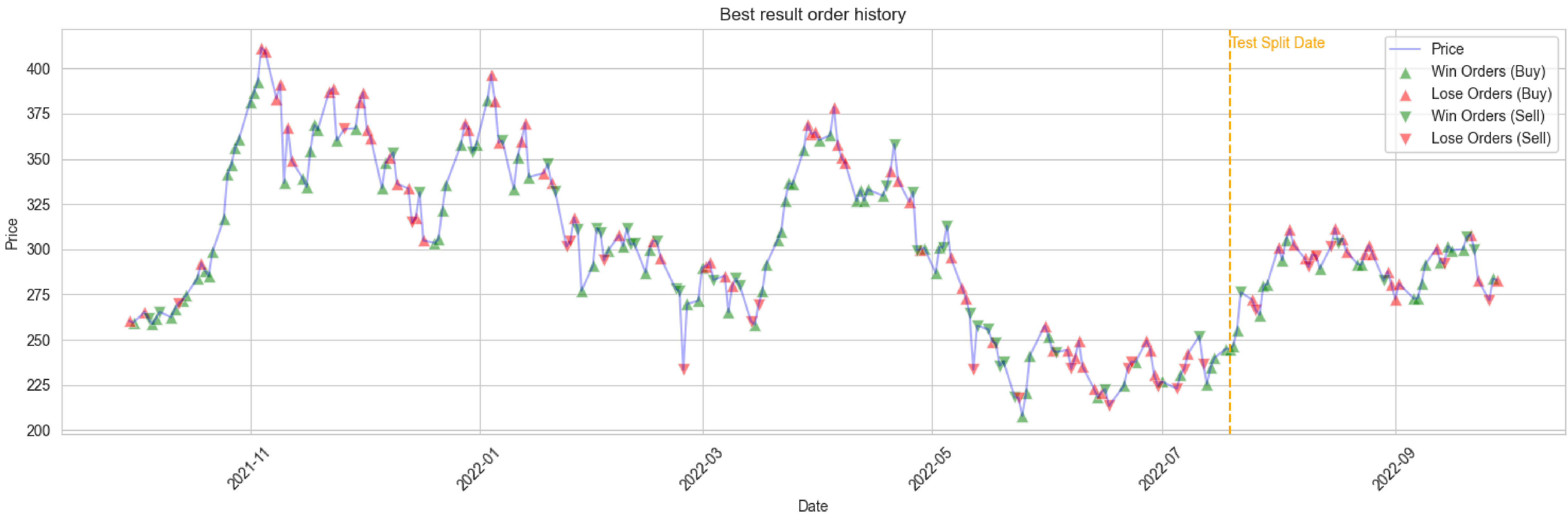


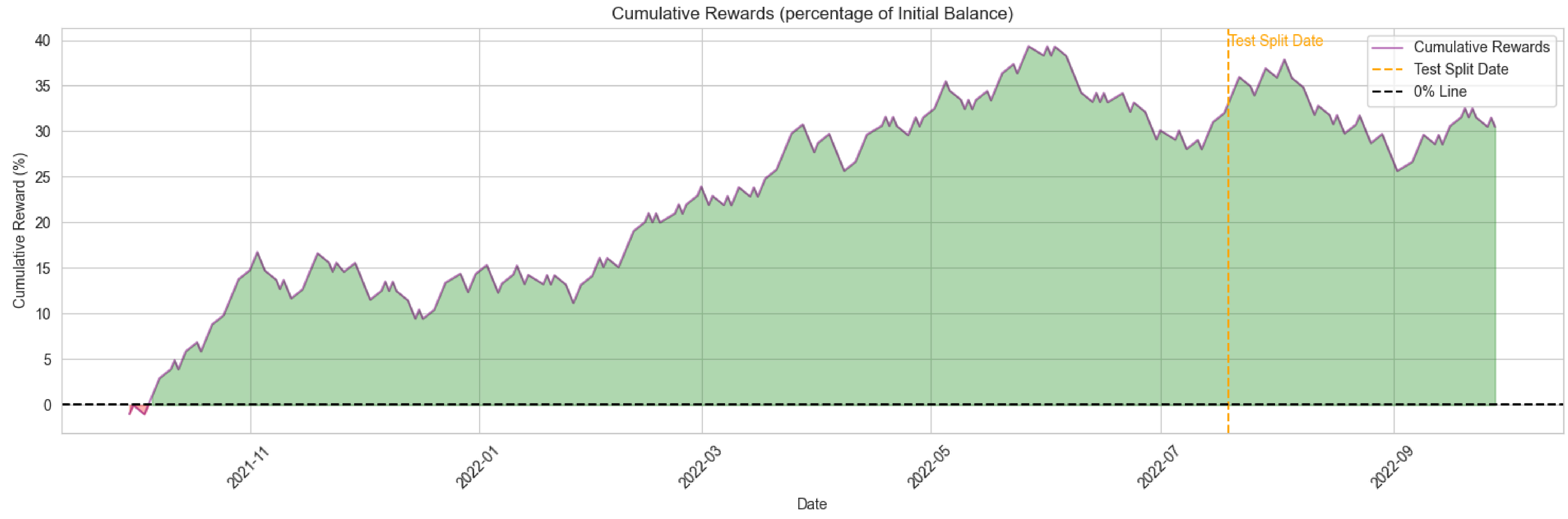
We use predicted x_t from our model to compare with ground truth.

	precision	recall	f1-score	support
Down	0.64	0.35	0.46	122
Up	0.57	$\boxed{0.82}$	0.67	130
accuracy			$\boxed{0.59}$	252
macro avg	0.61	0.58	0.56	252
weighted avg	0.61	0.59	0.57	252

APPENDIX

HMM trade model





In data preprocessing, if we use 3 consecutive day → lower profit rate

So, we got

Profit rate = 30.49% (-12%)