

# 인공지능을 이용한 금융 시계열 빅데이터 분석 및 예측 시스템

# Contents

Part 01

Introduction

Part 02

Prototype

Part 03

Conclusion

# Introduction

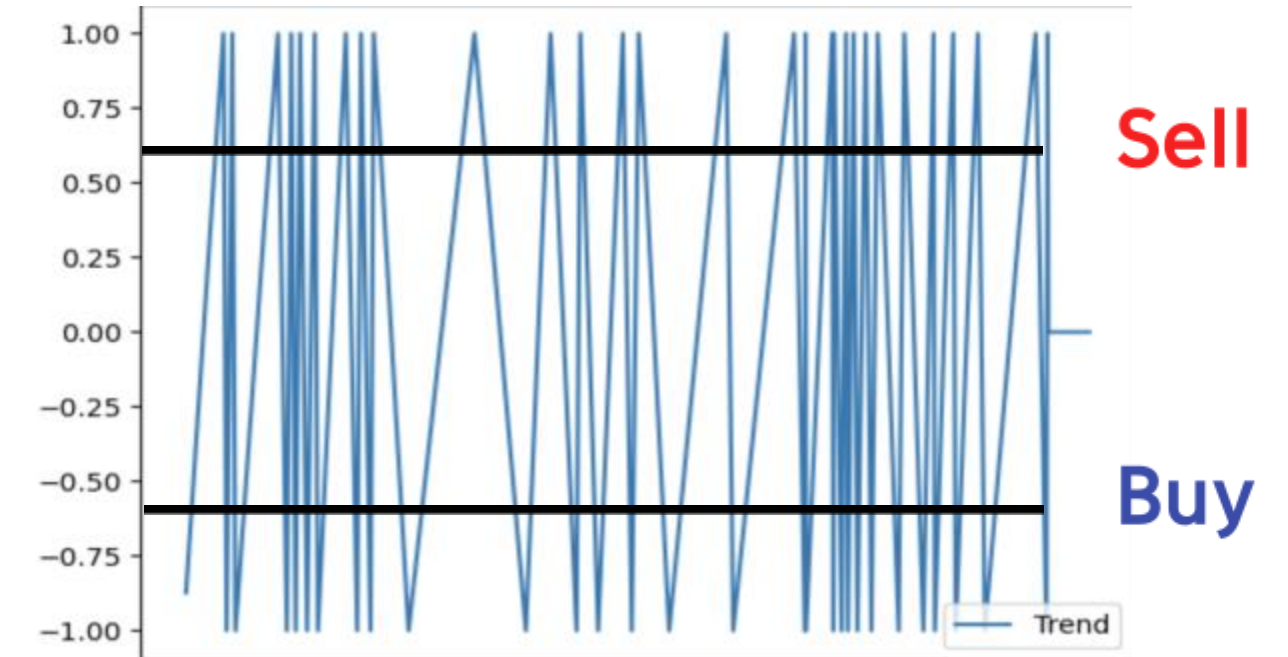
3

## Goal

rsi	sma_cross	ema_cross	roc	stochastic
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0



ML  
predict

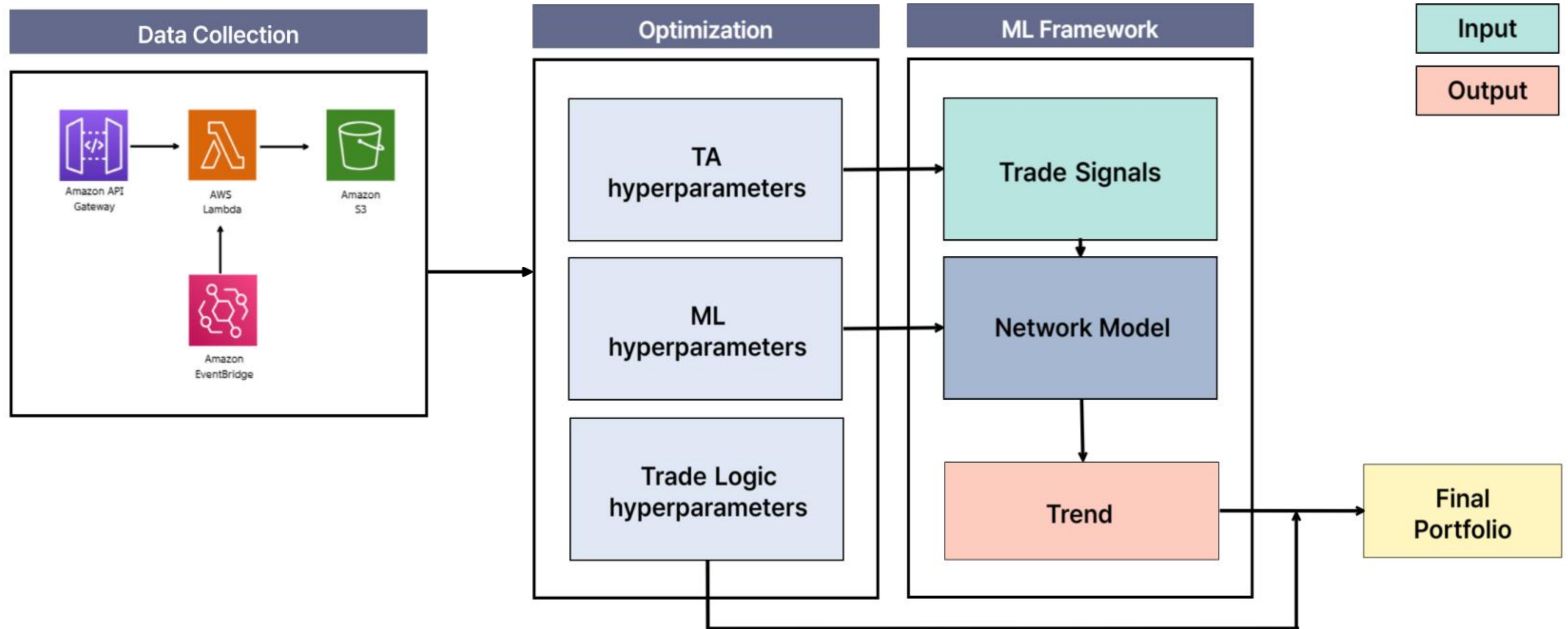


"시가총액의 추세를 예측하여  
매매거래 결정을 지원하는 통합적인 프레임워크 개발"

# Introduction

3

## Flowchart

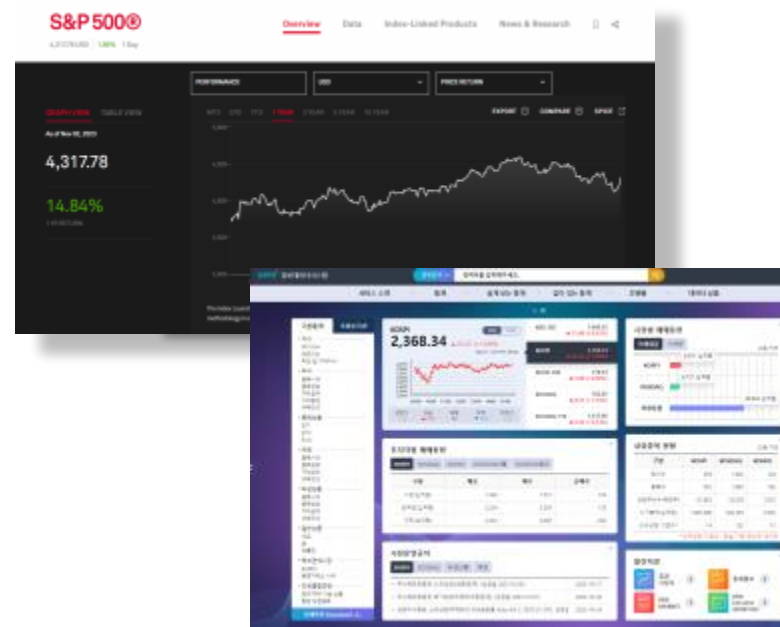


# Data collection

4

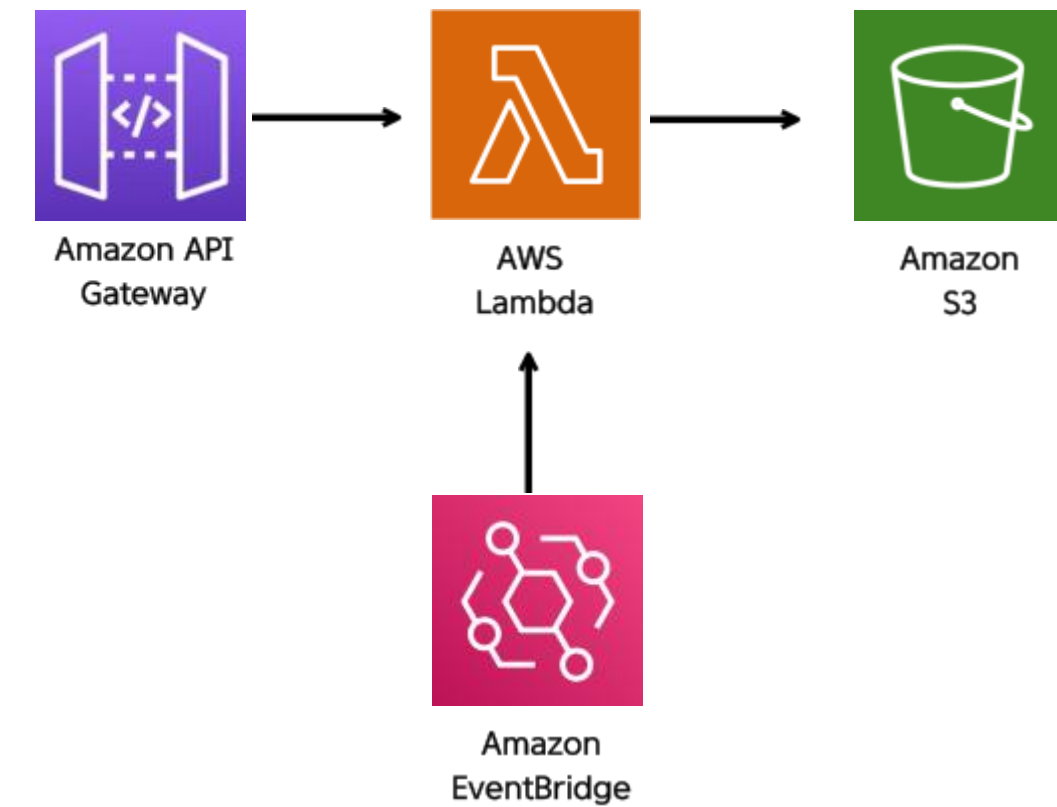
크롤링

데이터 크롤링 시스템 구축



KRX, S&P 500

AWS 클라우드



실시간 데이터 업데이트로 정확하고 신속한 정보 제공 가능

# Data collection


5

크롤링

데이터 크롤링 시스템

총 객체 수 = 데이터 수 + 1 (폴더)


S&P 500

이름	총 객체 수 ▼
 s&p/	504

#		Company	Symbol	Portfolio%	Price	Chg	% Chg
0	1	Microsoft Corp	MSFT	7.32%	381.33	2.72	(0.72%)
1	2	Apple Inc.	AAPL	7.29%	190.02	0.23	(0.12%)
2	3	Amazon.com Inc	AMZN	3.48%	147.02	-0.71	(-0.48%)
3	4	Nvidia Corp	NVDA	3.08%	477.20	-5.22	(-1.08%)
4	5	Alphabet Inc. Class A	GOOGL	2.12%	136.64	0.23	(0.17%)
...	...	...	...	...	...	...	...
498	499	Alaska Air Group, Inc.	ALK	0.01%	36.59	-0.25	(-0.67%)
499	500	Solaredge Technologies, Inc.	SEDG	0.01%	76.78	-0.04	(-0.05%)
500	501	Mohawk Industries, Inc.	MHK	0.01%	85.69	1.41	(1.67%)
501	502	Fox Corporation Class B	FOX	0.01%	28.04	-0.21	(-0.73%)
502	503	News Corporation Class B	NWS	0.01%	23.43	-0.05	(-0.19%)

503 rows × 7 columns

KRX

이름	총 객체 수 ▼
 krx/	2781

	full_code	short_code	codeName	marketCode	marketName	marketEngName	ord
0	KR7060310000	060310	3S	KSQ	코스닥	KOSDAQ	
1	KR7095570008	095570	AJ네트웍스	STK	유가증권	KOSPI	
2	KR7006840003	006840	AK홀딩스	STK	유가증권	KOSPI	
3	KR7054620000	054620	APS	KSQ	코스닥	KOSDAQ	
4	KR7265520007	265520	AP시스템	KSQ	코스닥	KOSDAQ	
...	...	...	...	...	...	...	...
2775	KR7000540005	000540	흥국화재	STK	유가증권	KOSPI	
2776	KR7000541003	000545	흥국화재우	STK	유가증권	KOSPI	
2777	KR7003280005	003280	흥아해운	STK	유가증권	KOSPI	
2778	KR7037440005	037440	희림	KSQ	코스닥	KOSDAQ	
2779	KR7238490007	238490	힘스	KSQ	코스닥	KOSDAQ	

2780 rows × 8 columns

# Data preprocessing

6

Data selection → Screening

Hurst exponent

Hurst exponent 값

=> 0.5 에 가까울수록 예측이 어려움

주가 시계열 데이터의 안정성에 대한 지수 파악



예측 가능성이 높은 데이터 선별 가능

threshold 지정

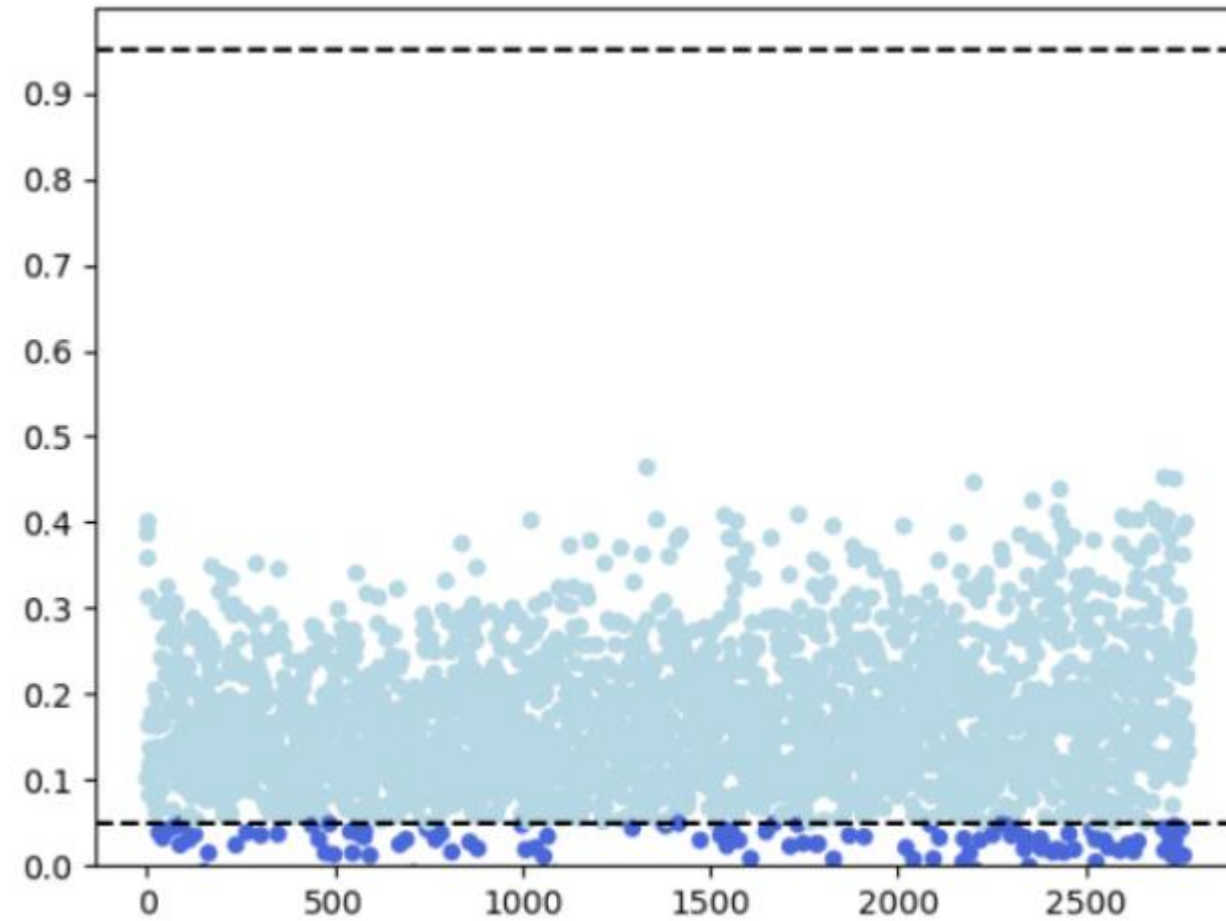


# Data preprocessing

7

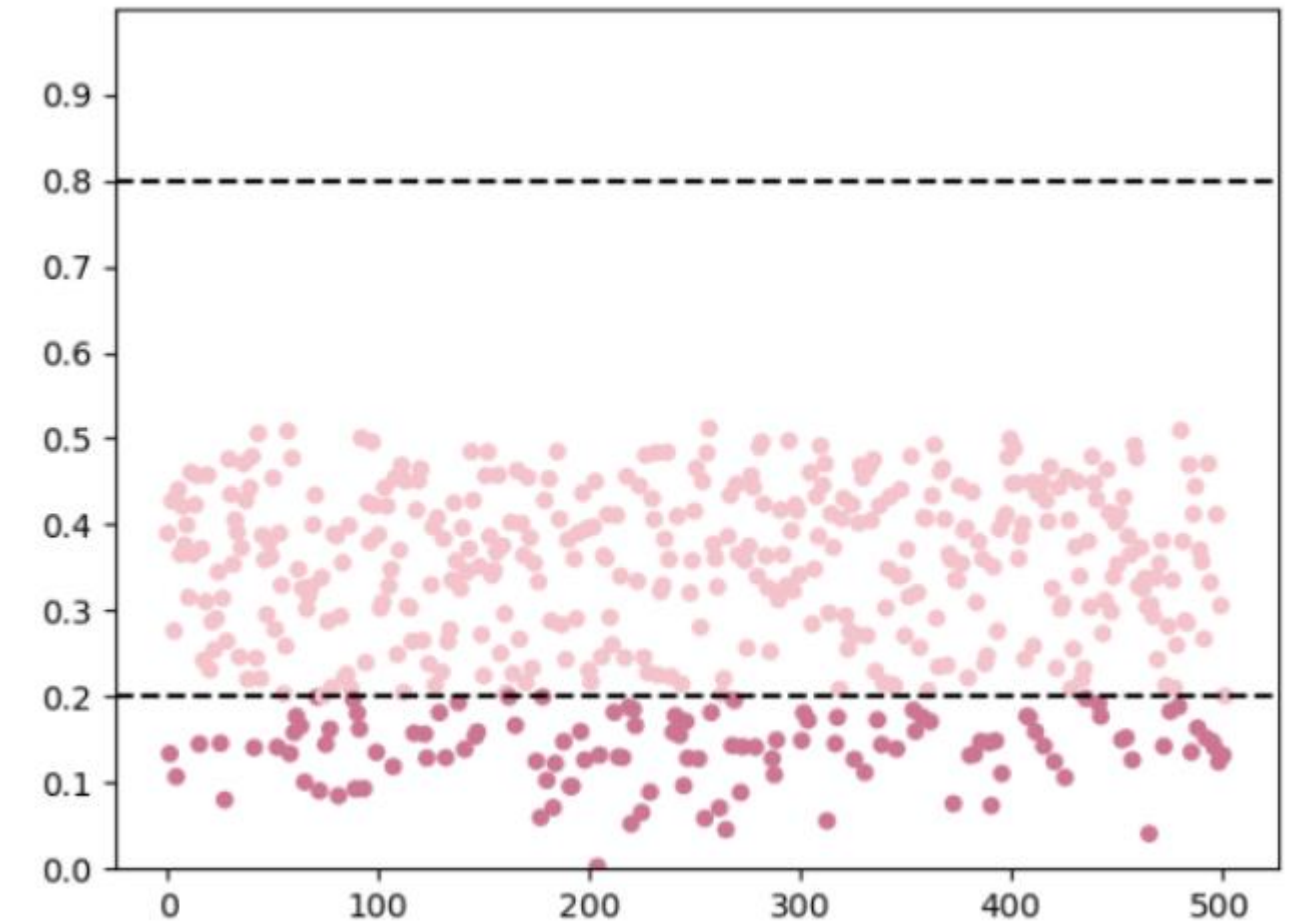
Data selection → Screening

Hurst exponent



KRX

$h < 0.05, h > 0.95$



s&p500

$h < 0.2, h > 0.8$



100

## TA(Technical Analysis) index

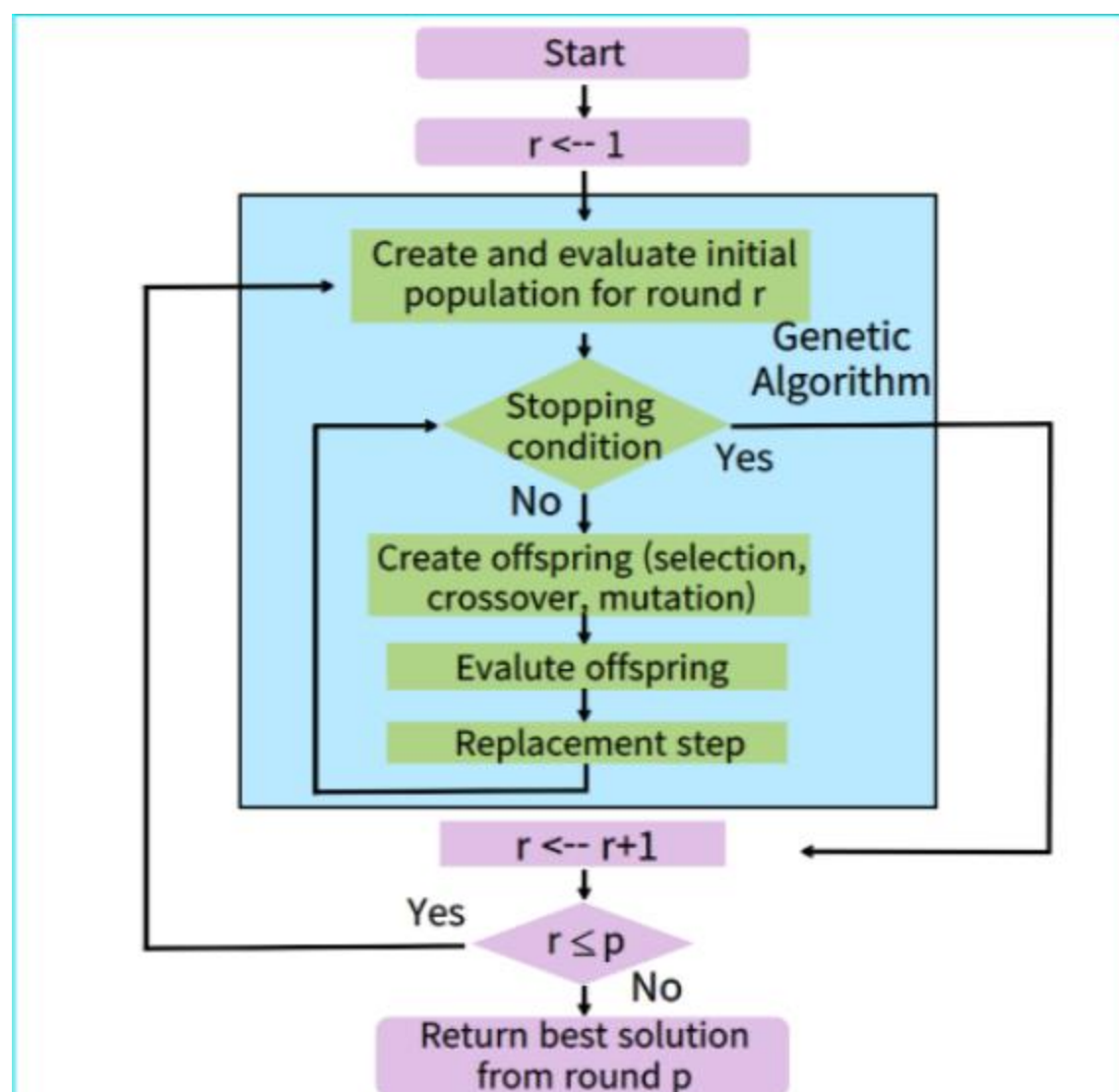
- RSI
- SMA
- EMA
- ROC
- Stochastic

[illegible]

# Data preprocessing

## Optimizer

### GA (Genetic Algorithm)



- 기술적 지표

$$a * \text{profit} + b * 1 / \text{max\_drawdown}$$

→ 사용자의 투자성향 고려

수익률이 높고 MDD가 작아지도록 함

- 각 모델, 매수에 사용할 자본의 비율, 시가총액의 trend data의 threshold

수익률이 높아지도록 함

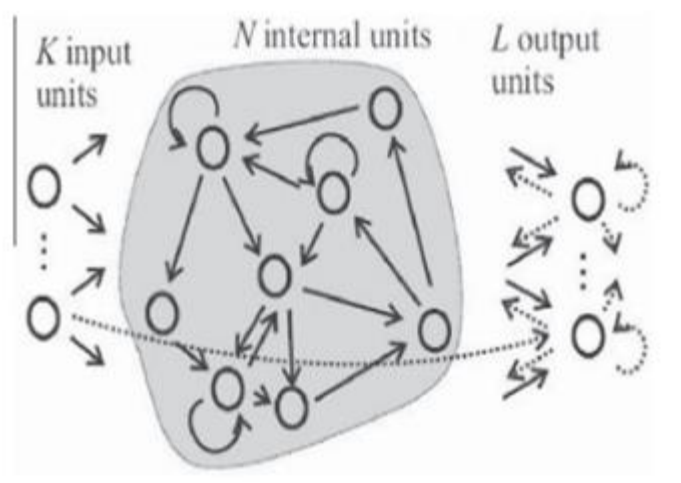
## Backtesting

생성된 매수 및 매도 신호를 이용해 예상 수익 측정 및 투자 전략 신뢰성 검증  
Train / test = 7:3 비율로 진행 후 test에 대한 예측 진행

## 거래 전략

- 매수 시점과 매도 시점의 차이의 비율을 통해 수익률을 측정하고, 거래(전량 매수)
- 매도 시 전량 매도

## Machine Learning Model



ESN

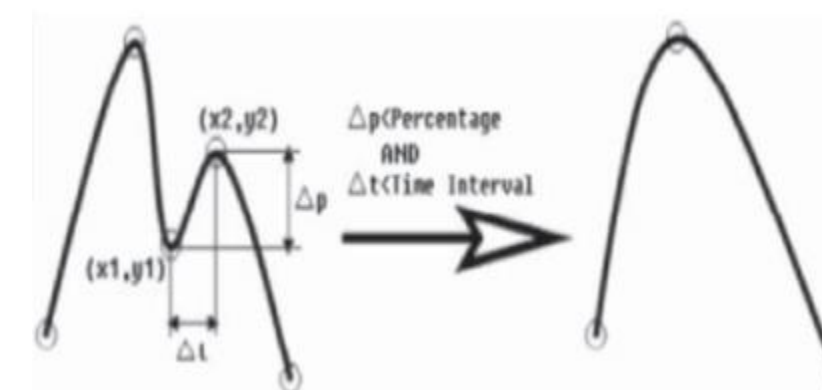
rsi	sma_cross	ema_cross	roc	stochastic
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0
0	-1	-1	-1	0

input



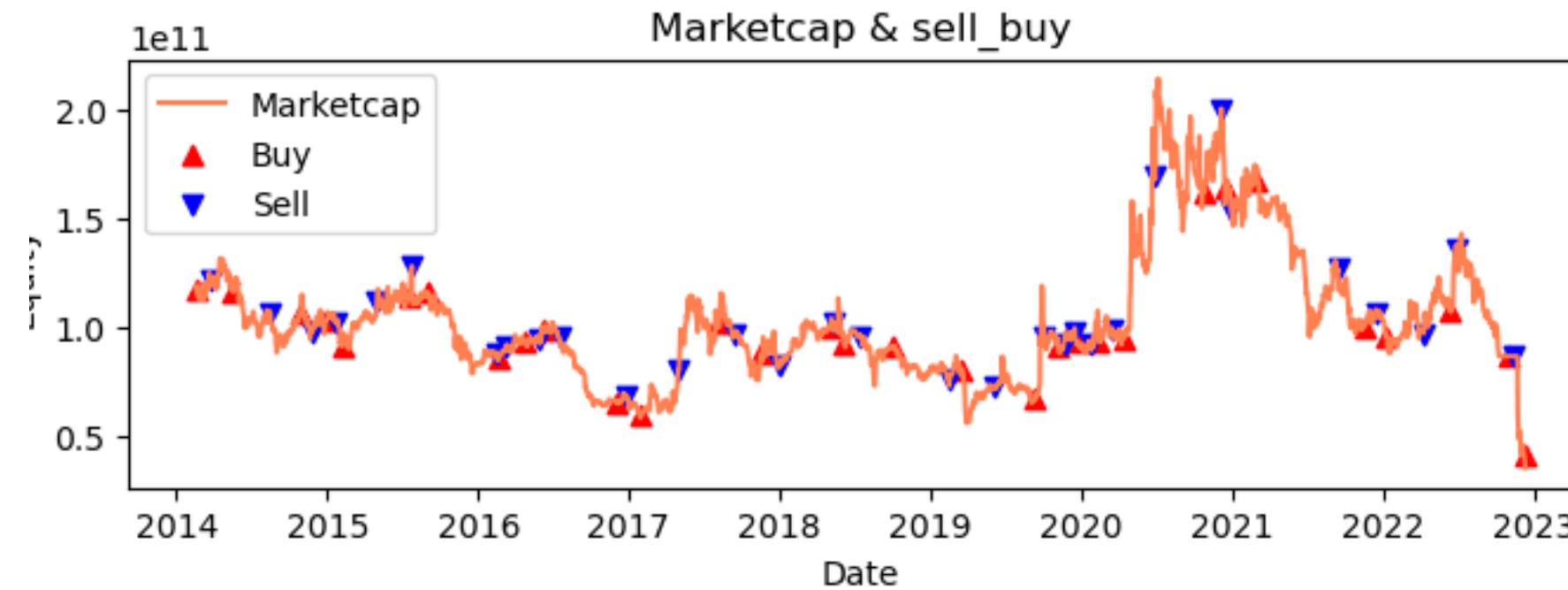
1. 배치모델

2. realtime모델



output

대유폴러스 -- buy\_and\_hold : 710.64%, rsi : 103356.44%, ema : 30433.04%, roc : 841.12%, sto : 36978.43%, sma : 49362.27% 수익률 : 12762.26%



Buy and Hold 대비 return(%): 824.1007332276059

## Buy and hold

-> 시장평균수익률 따르고 시장 변동에 영향을 적게 받아 벤치마크로도 자주 활용

우리 데이터의 수익률과 MDD값 중점적으로 고려 => 결과 비교

**저점에서 매수, 고점에서 매도하는 특성을 잘 나타냄**

1. MDD (Maximum DrawDown) 값 상수 변경

=> 사용자의 거래 성향 반영 가능

2. 다른 도메인으로의 확장 가능성

=> 복잡한 다른 시계열 데이터들에도 적용 할 수 있음



"시가총액의 추세를 예측하여  
매매거래 결정을 지원하는 통합적인 프레임워크 개발"





**THANK YOU**

