

## 캡스톤 디자인

- 인공지능을 이용한 금융시계열 빅데이터 분석 및 예측 시스템 -



팀명	3조
팀원	이름
	201811837 박한벗
	201930109 이태훈

# 목 차

1. 주제선정 .....	
가. 주제선정 .....	
2. 이론적 배경 .....	
가. 단어 및 이론 .....	
3. 제작 계획 및 과정 .....	
가. 제작 계획 .....	
나. 제작 과정 .....	
4. 제작 결과 및 고찰 .....	
가. 제작 결과 .....	
나. 고찰 .....	
5. 참고 문헌 .....	
가. 참고문헌 .....	

## 1. 주제선정

## 가. 주제선정(1주차)

주제는 인공지능을 이용한 금융 시계열 빅데이터 분석 및 예측 시스템으로서 시계열 빅데이터 처리 방법을 이용하여 금융 데이터를 분석 및 예측하는 프로젝트임.

금융 도메인인 만큼 인공지능뿐만 아니라 금융적 지식이 필요함.

이를 위해 1주차에는 금융에 필요한 전문 단어에 대해 공부를 진행.

또한 한국 거래소에서 금융데이터를 매번 직접 구매하면 많은 비용이 발생하고 번거롭기 때문에 한국 거래소 웹사이트에서 크롤링을 통해 데이터를 수집하기 위해 코드를 준비함.

## 2. 이론적 배경

### 가. 단어 및 이론

기초적 지표 정리

- 1.종가 (Closing Price): 주식시장에서 하루의 거래가 마감되었을 때의 주가. 다음 날 시장 개장 전까지 기준.
- 2.시가 (Opening Price): 주식시장에서 하루의 거래가 시작될 때의 주가. 시장 개장 시 첫 거래가 완료된 가격.
- 3.저가 (Low Price): 하루의 거래 기간 동안에 기록된 가장 낮은 주가.
- 4.고가 (High Price): 하루의 거래 기간 동안에 기록된 가장 높은 주가.
- 5.시가총액 (Market Capitalization): 회사의 주식 가치를 측정하는 지표로, 주식 가격과 발행된 주식 수를 곱한 값.
- 6.EPS (Earnings Per Share): 주당순이익으로, 회사의 순이익을 발행 주식수로 나눈 값. 주식의 수익성을 평가하는 지표.
- 7.PER (Price to Earnings Ratio): 주가 수익비율로, 주가를 주당순이익(EPS)으로 나눈 값. 주식의 가치를 평가하는데 사용하고, 낮은 PER은 상대적으로 저평가된 주식으로 평가.
- 8.BPS (Book Value Per Share): 주당순자산가치로, 회사의 순자산을 발행 주식수로 나눈 값.
- 9.PBR (Price to Book Ratio): 주가 순자산비율로, 주가를 주당순자산가치(BPS)로 나눈 값. 낮은 PBR은 상대적으로 저평가된 주식으로 간주.
- 10.주당배당금 (Dividends per Share): 주식회사가 주주들에게 배당하는 이익의 일부분으로, 주당 배당금은 주식 하나당 배당되는 금액.
- 11.배당수익률 (Dividend Yield): 주식의 배당금액을 주가로 나눈 값으로, 주식 투자로 얻을 수 있는 배당 수익의 비율. 높은 배당수익률은 높은 수익을 의미
- 12.ROE (Return on Equity): 자기자본이익률로, 회사의 순이익을 자기자본으로 나눈 값. ROE는 기업의 경영효율성과 자본 수익률을 평가하는 지표로 사용하고 높은 ROE

는 효율적인 자본 운용을 의미

13.ROA (Return on Assets): 총자산이익률로, 회사의 순이익을 총 자산으로 나눈 값. ROA는 기업의 자산 운용 효율성을 평가하는 지표로 사용되며, 높은 ROA는 경영의 효율성을 의미.

14.EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization): 이자, 세금, 감가상각비 및 무형자산상각비를 제외한 순이익으로, 기업의 영업이익을 나타내는 지표.

15.EV (Enterprise Value): 기업 가치로, 시가총액, 부채, 현금 및 현금성자산을 고려한 기업의 총 가치를 나타내는 지표.

16.EV/EBITDA (Enterprise Value to Earnings Before Interest, Taxes, Depreciation, and Amortization): 기업 가치 대비 EBITDA 비율로, 기업의 가치를 평가하는 지표로 사용. 낮은 EV/EBITDA 비율은 상대적으로 저평가된 기업으로 간주.

17.PEG (Price to Earnings Growth) Ratio: 주가 수익 성장비율로, PER을 회사의 예상 성장률로 나눈 값. 낮은 PEG 비율은 높은 성장 기대와 저평가된 주식으로 간주.

18.이동평균지수 (Moving Average, MA): 이동평균은 주가의 평균값을 계산하여 시계열 데이터의 잡음을 줄이고 추세를 파악하는 데 도움을 주는 지표. 일반적으로 단기 이동평균 (5일, 10일, 20일)과 장기 이동평균(60일, 120일, 240일)을 사용.

19.골든 크로스 (Golden Cross): 골든 크로스는 단기 이동평균선이 장기 이동평균선을 상향 돌파하는 시점. 주로 강세장 진입 신호로 해석하며, 주가의 상승세를 예측하는 데 사용.

20.데드 크로스 (Dead Cross): 데드 크로스는 단기 이동평균선이 장기 이동평균선을 하향 돌파하는 시점. 약세장 진입 신호로 해석하며, 주가의 하락세를 예측하는 데 사용.

21.RSI (Relative Strength Index): 상대강도지수는 주식의 가격 움직임을 분석하여 과매수 또는 과매도 상황을 판단하는 지표. RSI는 0에서 100 사이의 값으로 표시되며, 일반적으로 30 이하의 값은 과매도, 70 이상의 값은 과매수 상황으로 간주. 과매도 상황에서는 주가의 반등 가능성이 높고, 과매수 상황에서는 주가의 조정 가능성이 높다고 해석.

22.MACD (Moving Average Convergence Divergence): 이동평균 수렴 발산 지표로, 단기 이동평균과 장기 이동평균의 차이를 통해 추세 전환 시점을 파악하는 데 사용.

23.볼린저 밴드 (Bollinger Bands): 주가의 변동성을 측정하는 지표로, 이동평균선을 중심으로 일정한 표준편차 범위를 설정하여 밴드를 그린 것. 볼린저 밴드는 주가가 밴드의 상단을 벗어나면 과매수, 하단을 벗어나면 과매도 상황으로 간주.

24.스토캐스틱 오실레이터 (Stochastic Oscillator): 과매수 또는 과매도 상황을 판단하기 위해 사용되는 모멘텀 지표. 스토캐스틱 오실레이터는 0에서 100 사이의 값으로 표시하고, 일반적으로 20 이하의 값은 과매도, 80 이상의 값은 과매수 상황으로 간주.

25.ADX (Average Directional Index): 추세의 강도를 측정하는 지표로, ADX 값이 높을수록 강한 추세를, 낮을수록 약한 추세를 의미. 일반적으로 ADX 값이 25 이상이면 강한 추세로 간주.

26.거래량 (Volume): 주식 거래의 양을 나타내는 지표로, 거래량 분석은 주가 움직임에 대한 시장 참여자들의 관심도를 평가하는 데 사용. 거래량이 증가하면 주가 움직임에 강한 신호, 반대로 거래량이 감소하면 약한 신호로 해석.

27.피보나치 수열 (Fibonacci Retracement): 주가의 트렌드 반전 지점을 찾기 위해 사용되는 기술적 분석 도구. 주가의 상승 또는 하락 구간에서 일정한 비율 23.6%, 38.2%, 50%, 61.8%, 78.6% 로 가격이 반전될 것이라 보는 지표.

28.일목균형표 (Ichimoku Cloud): 일본에서 개발 일정 기간 동안의 가격 평균을 시각화하여 추세와 지지 및 저항 구간을 파악하는 데 사용

29.RVI (Relative Vigor Index): 가격 변동과 거래 활동 간의 상관관계를 측정하는 모멘텀 지표, 가격의 상승세와 하락세를 비교하여 시장의 힘을 평가하는 데 사용. 과매수, 과매도 시점을 확인할때 도움.

30.CCI (Commodity Channel Index): 상품 가격의 일정 기간 동안의 평균 가격과 현재 가격 간의 차이를 측정하는 지표, 과매수 또는 과매도 상황을 확인할때 도움. 일반적으로 CCI 값이 100 이상이면 과매수, 100 이하면 과매도 상황.

31.CAGR (Compound Annual Growth Rate): 연복리 성장률로, 일정 기간 동안의 투자 수익률을 나타내는 지표.

32.MDD (Maximum Drawdown): 최대 낙폭으로, 투자 기간 동안 투자자산의 최고점에서 최저점까지의 최대 손실률.

33.Sharpe Ratio: 샤프 비율은 투자의 위험 대비 수익률을 나타내는 지표로, 투자 수익률과 무위험 수익률의 차이를 투자의 표준편차로 나눈 값. 샤프 비율이 높을수록 위험 대비 높은 수익률을 기록한 것으로 평가.

34.Rebalancing: 리밸런싱은 투자 포트폴리오의 자산 비중을 일정한 목표 비중으로 유지하기 위해 주기적으로 조정하는 과정.

35.Back Testing: 백테스팅은 과거 데이터를 활용하여 투자 전략의 성과를 평가하는 과정.

36.Alpha (알파): 초과 수익률로, 벤치마크 대비 포트폴리오의 성과를 나타냄. 알파가 양수 -> 벤치마크에 비해 높은 수익률.

37.Beta (베타): 주식이나 포트폴리오의 시장 리스크를 측정하는 지표로, 주가의 변동성을 시장 지수와 비교하여 평가. 베타가 1초과 -> 시장보다 더 큰 변동성을 가진 것, 베타가 1미만-> 시장보다 낮은 변동성을 가진 것. 베타가 1 -> 주식이나 포트폴리오의 변동성이 시장과 동일.

38.Sortino Ratio (소르티노 비율): 투자의 위험 대비 수익률을 나타내는 지표. 다운사이드 리스크(하락 시의 리스크)만 생각. 소르티노 비율이 높을수록 위험 대비 높은 수익률을 기록한 것.

39.AUM (Assets Under Management): 자산운용사가 관리하는 자산의 총액.

40.Quantitative Trading (퀀트 트레이딩): 수량적인 데이터와 수학적, 통계적 기법을 활용하여 투자 전략을 세우고 실행하는 투자 방식.

### 3. 제작계획 및 과정

#### 가. 제작계획

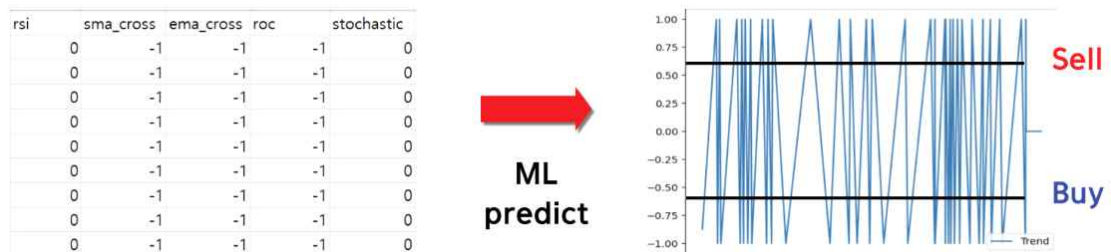


그림 1. 시가총액의 추세를 예측하여 매매 거래 결정을 지원하는 통합적인 프레임워크 개발을 목적으로 함. 이때 높은 수익률과 낮은 MDD(최대 낙폭)를 가질 수 있도록 고안함.

#### Flowchart

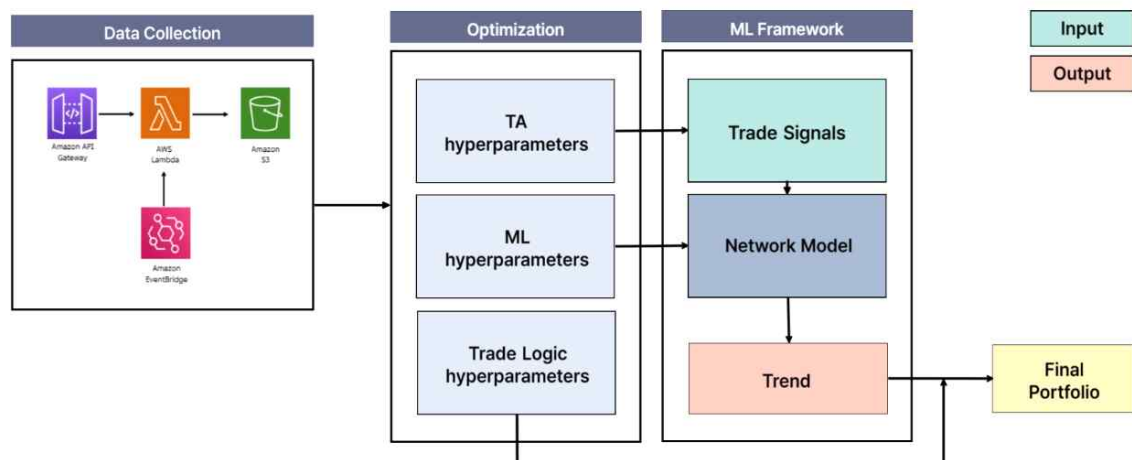


그림 2. 구체적인 흐름도로 AWS를 사용해 데이터를 크롤링하고 데이터를 바탕으로 TA 지수, ML모델의 매개변수, 거래전략의 매개변수를 GA(유전 알고리즘)을 통해 높은 수익률과 낮은 MDD(최대 낙폭)를 가질 수 있도록 최적화하고, 이를 바탕으로 시가총액의 추세를 예측하여 거래를 진행함.

#### 제작 순서

1. 데이터 수집
2. 데이터 스크리닝
3. 데이터 전처리
4. 최적화(기계학습, 거래전략)
5. 성능평가

#### 나. 제작 과정

## 1. 데이터 수집

한국 거래소에서 AWS 클라우드를 통해, 크롤링 시스템을 구축하였고, 이를 통해 매일 모든 종목의 데이터를 수집할 수 있도록 설계하였음.

### 크롤링 데이터 크롤링 시스템 구축

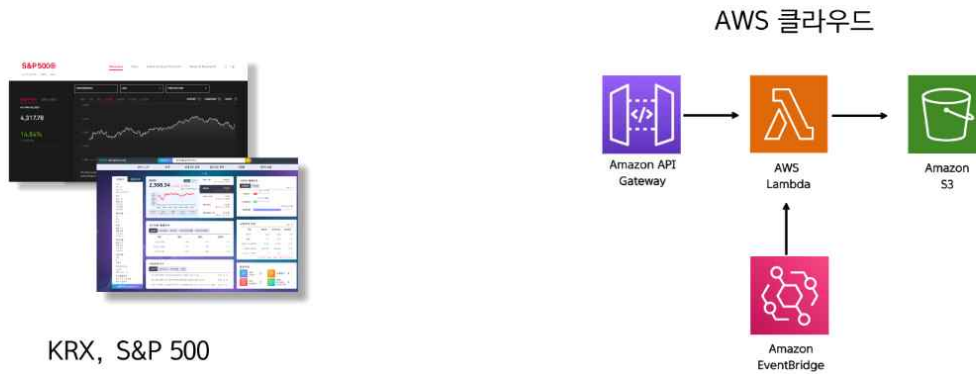


그림 3. AWS를 통해 데이터 크롤링 시스템을 구축하여, 매일 데이터를 업데이트로 정확하고 신속한 정보 제공 가능하도록 설계

	full_code	short_code	codeName	marketCode	marketName	marketEngName	ord1	ord2
0	KR7060310000	060310	3S	KSQ	코스닥	KOSDAQ	16	
1	KR7095570008	095570	AJ네트웍스	STK	유가증권	KOSPI	16	
2	KR7006840003	006840	AK홀딩스	STK	유가증권	KOSPI	16	
3	KR7054620000	054620	APS홀딩스	KSQ	코스닥	KOSDAQ	16	
4	KR7265520007	265520	AP시스템	KSQ	코스닥	KOSDAQ	16	
...	...	...	...	...	...	...	...	...
2703	KR7000542001	000547	한국화재2우B	STK	유가증권	KOSPI	16	
2704	KR7000541003	000545	한국화재우	STK	유가증권	KOSPI	16	
2705	KR7003280005	003280	한국해운	STK	유가증권	KOSPI	16	
2706	KR7037440005	037440	회림	KSQ	코스닥	KOSDAQ	16	
2707	KR7238490007	238490	힘스	KSQ	코스닥	KOSDAQ	16	

그림 4. 크롤링을 통해 모든 주식 종목의 fullcode, short code, codename 등의 정보를 확인 가능



	TRD_DD	TDD_CLSPRC	FLUC_TP_CD	CMPPREVDD_PRC	FLUC_RT	TDD_OPNPRC	TDD_HGPRC	TDD_LWPRC	ACC_TRDVOL	ACC_TRDVAL	MKTCAI
0	2023/03/24	63,000	1	700	1.12	62,700	63,300	62,300	18,278,602	1,147,804,584,500	376,096,300,650,00
1	2023/03/23	62,300	1	1,200	1.96	60,600	62,300	60,600	15,381,057	946,393,607,300	371,917,452,865,00
2	2023/03/22	61,100	1	800	1.33	61,000	61,200	60,500	8,978,591	546,787,085,426	364,753,713,805,00
3	2023/03/21	60,300	1	100	0.17	60,500	60,700	60,100	8,318,514	501,917,706,700	359,977,887,765,00
4	2023/03/20	60,200	2	-1,100	-1.79	61,100	61,200	60,200	9,618,009	582,445,819,345	359,380,909,510,00
...	...	...	...	...	...	...	...	...	...	...	...
298	2022/01/07	78,300	1	1,400	1.82	78,100	78,400	77,400	15,163,757	1,184,236,552,700	467,433,973,665,00
299	2022/01/06	76,900	2	-500	-0.65	76,700	77,600	76,600	12,931,954	996,378,432,570	459,076,278,095,00
300	2022/01/05	77,400	2	-1,300	-1.65	78,800	79,000	76,400	25,470,640	1,967,841,509,800	462,061,169,370,00
301	2022/01/04	78,700	1	100	0.13	78,800	79,200	78,300	12,427,416	977,990,613,000	469,821,886,685,00
302	2022/01/03	78,600	1	300	0.38	79,400	79,800	78,200	13,502,112	1,066,006,837,750	469,224,908,430,00

303 rows × 12 columns

그림 5. 종목의 fullcode를 통해 각 종목의 시가, 종가, 저가, 고가, 시가총액 등의 데이터를 수집할 수 있음

## 2. 데이터 스크리닝

Hurst Exponent를 통해 예측이 가능한 종목을 선별

### Data selection → Screening

#### Hurst exponent

Hurst exponent 값

=> 0.5에 가까울수록 예측이 어려움

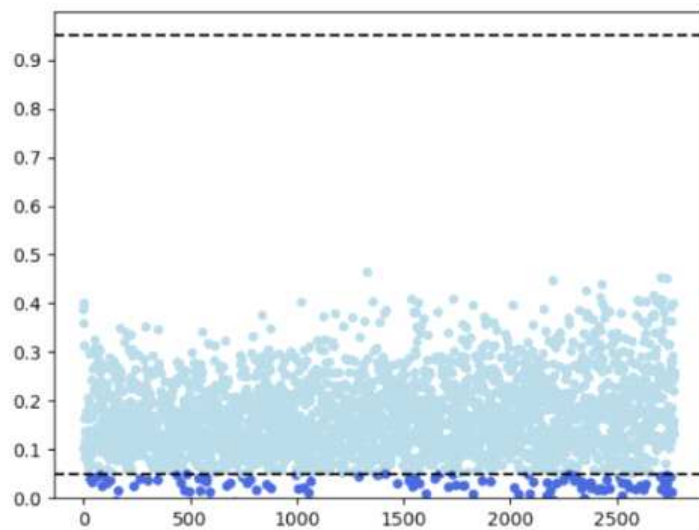
주가 시계열 데이터의 안정성에 대한 지수 파악



예측 가능성이 높은 데이터 선별 가능

threshold 지정

그림 6. Hurst exponent에 대한 설명으로 0.5에 가까울수록 무작위적인 특성을 가지기 때문에 예측을 하는 것이 어려움. 따라서 예측이 가능하도록, 기준선을 정해서 hurst exponent 값이 0과 1에 가까운 종목만을 선별함.



KRX

$h < 0.05, h > 0.95$

그림 7. KRX 종목들을 hurst exponent를 통해 계산하고 잘 예측되는 종목을 선별하기 위해 0.05보다 작고 0.95보다 큰 종목만을 선별하였음.

### 3. 데이터 전처리

#### TA(Technical Analysis) index

RSI  
SMA  
EMA  
ROC  
Stochastic



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

그림 8. 기술 지표인 TA index 중, RSI, SMA, EMA, ROC, Stochastic을 통해 선별한 주식 종목의 매매 신호를 생성하였음. 이때 sell은 -1, stay는 0, buy는 1로 생성하였음.

## • 기술적 지표

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

-> 사용자의 투자성향 고려

수익률이 높고 MDD가 작아지도록 함

그림 9. 기술적 지표를 최적화 할 때, 목적함수로 다음과 같은 식을 통해 사용자의 투자성향을 고려할 수 있음.

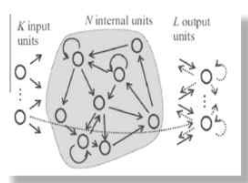
매매 신호를 생성한 이후 자연 기반의 최적화기인 GA(유전 알고리즘)를 통해 그림 9에서 확인할 수 있듯이 투자자의 성향을 반영하여 수익률은 높고, MDD는 낮아지도록 최적화함.

이때 수익률은 backtest를 통해 확인할 수 있음. 수익률은 매수 금액과 매도 금액의 비율을 통해 측정하였고, 복리 계산을 통해 수익률을 더욱 증가시킬 수 있었음.

거래 시 발생할 수 있는 수수료 및 슬리피지를 선정하여 더욱 현실에 가까운 거래가 될 수 있도록 알고리즘을 설계함.

## 4. 최적화(기계학습, 거래전략)

### Machine Learning Model



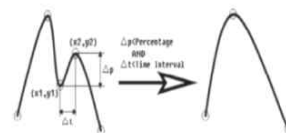
ESN

1. 배치모델

2. realtime모델

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

input



output

그림 10. 기계학습 모델로 복잡한 비선형 시계열 데이터 예측에 적합한 ESN(Echo State Network)를 선정하였고, 이를 배치 모델과 실시간 모델을 만들어 필요에 따라 사용할 수 있도록 하였음.

선별한 주식 종목의 시가총액을 추세를 구할 수 있는 알고리즘을 통하여 생성하고, TA index를 통해 생성한 매매 신호를 통해 추세 데이터를 예측할 수 있도록 기계학습 모델을 학습하였음.

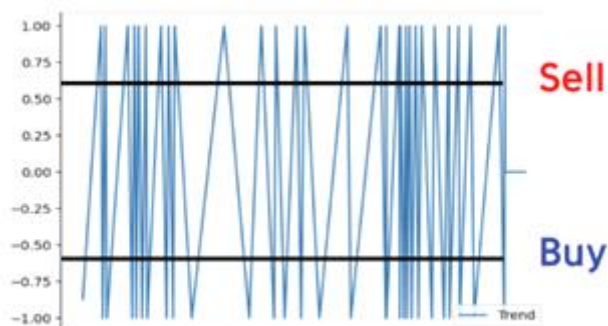


그림 11. 예측한 시가총액에서 기준선을 정하여 매매 신호를 생성하는데, 이때 기준선 또한 최적화를 통해 정하였음.

Train, Valid, Test를 5:2:3의 비율로 하여, Valid 데이터에서 높은 수익률과 낮은 MDD를 가질 수 있도록, ESN의 매개변수와 거래전략의 매개변수를 최적화하였음.

#### 4. 성능평가

수익률, MDD 그리고 거래 시점을 통해 기존 전략에 비해 나은지 판별함. 이때 기존 전략은 해당 종목의 시가총액의 변화량을 통해 비교하였음. 이는 사용자가 장기투자를 하는 유형과 매우 유사함.

대유플러스 -- buy\_and\_hold : 710.64%, rsi : 103355.44%, ewa : 30433.04%, roc : 841.12%, sto : 36978.43%, sma : 49362.27% 수익률 : 12762.26%



Buy and Hold 대비 return(%): 824.1007332276059

그림 12. 해당 종목은 대유플러스라는 종목으로 GA를 통해 최적화하였을 때 RSI에서 buy and hold 대비 12762% 수익률을 얻을 수 있고, 최적화된 TA로 생성한 매매 신호를 통해 기계학습으로 예측하여 매매를 진행했을 때, 위의 그림과 같은 매매 시점을 확인할 수 있음. 또한 buy and hold 대비 824%의 수익률로 장기투자 했을 때와 비교하여 훨씬 높은 수익률을 얻을 수 있음.

본 프로젝트의 성능평가는 수익률과 MDD를 통하여 확인할 수 있었음. 기존 전략인 buy and hold에 비해 수익률은 높고 MDD는 낮아졌기 때문에, 높은 성능을 가지는 거래 전략을 생성하였다고 할 수 있음.

## 4. 제작결과 및 고찰

### 가. 제작결과

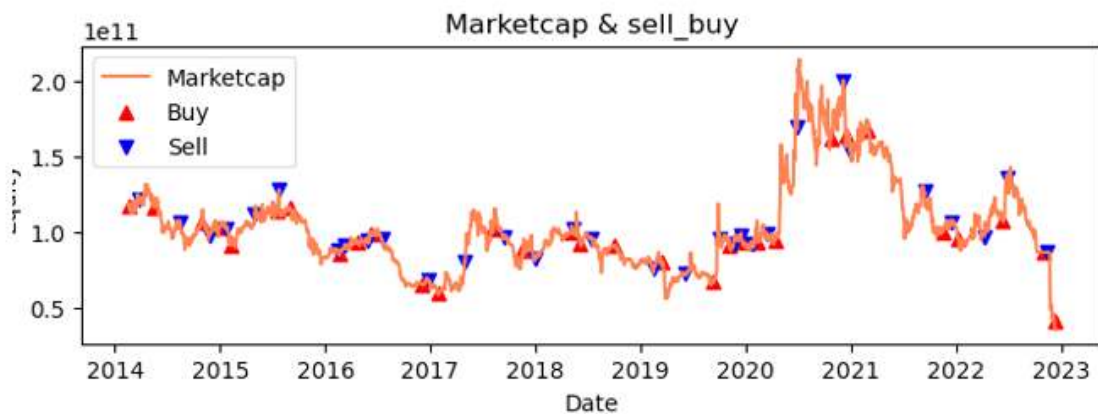


그림 13. 본 프로젝트에서 만든 프레임워크를 통해 매매를 진행했을 때의 거래 시점을 나타내며 붉은색 세모는 buy, 푸른색 세모는 sell을 나타냄. 확인할 수 있듯이 고점에서 판매하고 저점에서 구매하는 양상을 보임.

그림 12, 그림 13을 통해 최대 낙폭이 낮아 이용자들이 심리적으로 안정적으로 거래를 할 수 있으며, 수익률 또한 높게 나와 충분히 실제 거래에서도 사용할 수 있는 프레임 워크를 개발하였다고 여겨짐.

해당 종목 뿐만 아니라 Hurst exponent로부터 선별한 모든 종목에서 buy and hold 대비 높은 수익률과 낮은 MDD를 확인할 수 있었음.

### 나. 고찰

1. 실제 거래를 현실적으로 반영하기 위해서는 거래 플랫폼의 수수료와 실제 상황을 잘 고려할 수 있는 슬리피지 값을 반영해야함.
2. 이를 통해 매일 거래 뿐만 아니라 실시간 거래에도 비슷한 유형의 방식으로 접근할 수 있을 것으로 예상함.

## 5. 참고 문헌

가. 참고문헌

「Lin XYang ZSong Y, Intelligent stock trading system based on improved technical analysis and Echo State network」