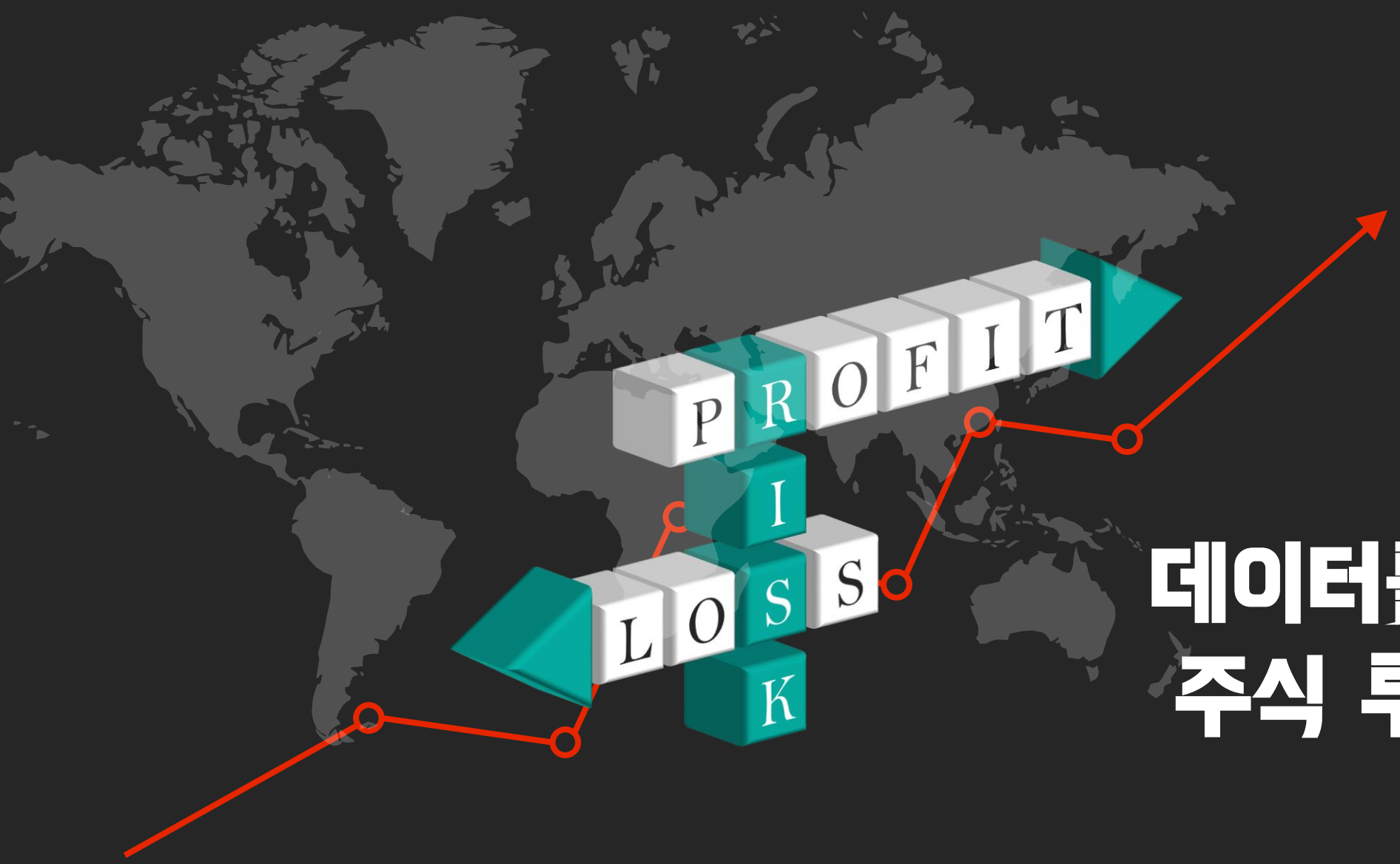



# 데이터를 활용한 주식 투자 전략

팀명 : 떡상 가즈아~!!

# Our Team Member

- 이일섭 멘토
- 신한카드 데이터 분석

**Jo Sanghun**

- 데이터 수집
- ARIMA
- 재무제표 분석

**Jeong Daun**

- 데이터 수집
- 기술적 지표 분석
- 모델링(Pycaret)

**Yun Songyi**

- 데이터 수집
- ARIMA
- 재무제표 분석

**Bang Eunho**

- 데이터 수집
- 기술적 지표 분석
- 모델링(Backtesting)

# Table of contents


01 선정 배경

주제 및 데이터 선정 배경

02 시계열 분석

ARIMA 분석

03 기술적 분석

RSI, MACD, 정분류율, 모멘텀

Pycaret, Backtesting

04 기본적 분석 및 결론

재무제표 분석

결론

# 선정 배경

# 주제 선정 배경

**물가 오른만큼 월급이 안 오른다…실질임금 0%**

입력 2022-01-27 14:52:09 수정 2022.01.27 14:52:09 세종=양종곤 기자

MZ세대 선호 재테크 수단 1위는 '부동산'… 2위는 주식

머니 경제

'고위험·고수익' 추구하는 MZ…자산관리도 주식·펀드 등 투자상품

2022.01.13 금융·증권 > 금융

"5% 은행 이자는 별로"… 군인들도 적금 대신 주식·코인한다

'장병내일준비적금' 가입자 감소세
최근 주식·가상화폐 시장 변동성 커 투자 주의해야

- 사회생활을 하게 되면 개인의 자산을 관리하는 경우 발생
- 적금은 자산운용에 있어서 더 이상 효율적인 방법이 아님
- 하지만 정보도 없이 투자에 매일 매달리는 것은 수익이 나도 장기적인 관점에서는 삶의 도움이 안됨

어떤 투자방법을 선택하느냐도 중요하지만, 이에 앞서, 나의 현재 재무상태, 위험에 대한 감내 수준, 기대수익률 등을 체크해보는 시간이 먼저 필요합니다. 이러한 부분들을 고려하여 목표를 세우고, 상품에 대한 고민을 해야 합니다. 또한 상품에 대한 지식은 인터넷이나 금융기관을 통해 쉽게 얻을수 있지만, 먼저 경제에 대한 공부를 통해 필요한 정보와 잘못된 정보를 선별할 수 있는 안목을 길러야 합니다. 이를 통해 자신에게 맞는 상품을 선택한다면, 최선의 투자라고 생각합니다.

**너무 잦은 매매가 아닌**
**투자 유형을 수립**

# 데이터 선정 배경

- 사용 데이터

  KOSPI 200 Data

Why??

- 미국 주식, 코인보다 편한 투자 시장
- 코스닥 보다 시장변동성이 적음
- 수많은 레퍼런스와 데이터 활용 용이

- 데이터 종류(2019 -2021)
  - ✓ 코로나 시기가 포함된 기간의 주가 동향을 분석

< 한국 거래소 데이터 >

지수명 코스피 200 조회

조회기간 20180703 ~ 20220104 1일 1개월 6개월 1년

Close ×

2022.03.30 AM 11:06:53 (20분 지연 정보) 전주 백만원

| 일자 | 종가 | 대비 | | 등락률 | 시가 | 고가 | 저가 | 거래량 | 거래대금 | 상장시가총액 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2022/01/04 | 395.40 | ▼ | 0.11 | -0.03 | 396.01 | 396.71 | 393.35 | 117,161 | 7,335,075 | 1,908,027,275 |
| 2022/01/03 | 395.51 | ▲ | 1.32 | +0.33 | 397.16 | 399.44 | 394.26 | 97,581 | 5,916,888 | 1,909,559,749 |
| 2021/12/30 | 394.19 | ▼ | 2.53 | -0.64 | 397.13 | 398.59 | 393.89 | 103,887 | 6,441,432 | 1,902,371,491 |
| 2021/12/29 | 396.72 | ▼ | 4.49 | -1.12 | 398.74 | 399.16 | 395.85 | 169,048 | 8,442,540 | 1,914,640,390 |
| 2021/12/28 | 401.21 | ▲ | 2.60 | +0.65 | 399.56 | 401.21 | 397.24 | 169,712 | 9,338,598 | 1,935,521,460 |
| 2021/12/27 | 398.61 | ▼ | 1.92 | -0.48 | 400.61 | 401.07 | 398.31 | 103,636 | 5,653,275 | 1,920,708,427 |
| 2021/12/24 | 400.53 | ▲ | 2.38 | +0.60 | 399.72 | 402.46 | 399.66 | 102,503 | 6,252,102 | 1,929,070,500 |
| 2021/12/23 | 398.15 | ▲ | 2.08 | +0.53 | 397.98 | 398.36 | 395.71 | 105,282 | 6,647,241 | 1,919,620,421 |
| 2021/12/22 | 396.07 | ▲ | 2.12 | +0.54 | 396.82 | 398.06 | 394.77 | 100,313 | 6,446,064 | 1,910,087,957 |
| 2021/12/21 | 393.95 | ▲ | 2.58 | +0.66 | 394.35 | 395.28 | 391.19 | 105,998 | 6,533,208 | 1,904,137,929 |
| 2021/12/20 | 391.37 | ▼ | 7.59 | -1.90 | 396.42 | 396.61 | 390.92 | 97,270 | 6,575,366 | 1,895,359,388 |
| 2021/12/17 | 398.96 | ▲ | 1.28 | +0.32 | 394.46 | 399.02 | 394.26 | 134,698 | 8,722,582 | 1,932,001,853 |

< Yahoo Finance 데이터 >

| Date | Open | High | Low | Close | Volume | Change |
|---|---|---|---|---|---|---|
| 2018-12-03 | 42750 | 43400 | 42400 | 43250 | 12110702 | 0.033453 |
| 2018-12-04 | 42650 | 42900 | 41900 | 42150 | 14347746 | -0.025434 |
| 2018-12-05 | 40900 | 41750 | 40850 | 41450 | 12631983 | -0.016607 |
| 2018-12-06 | 40600 | 41100 | 40450 | 40500 | 14251826 | -0.022919 |
| 2018-12-07 | 40900 | 41400 | 40850 | 40950 | 11433083 | 0.011111 |

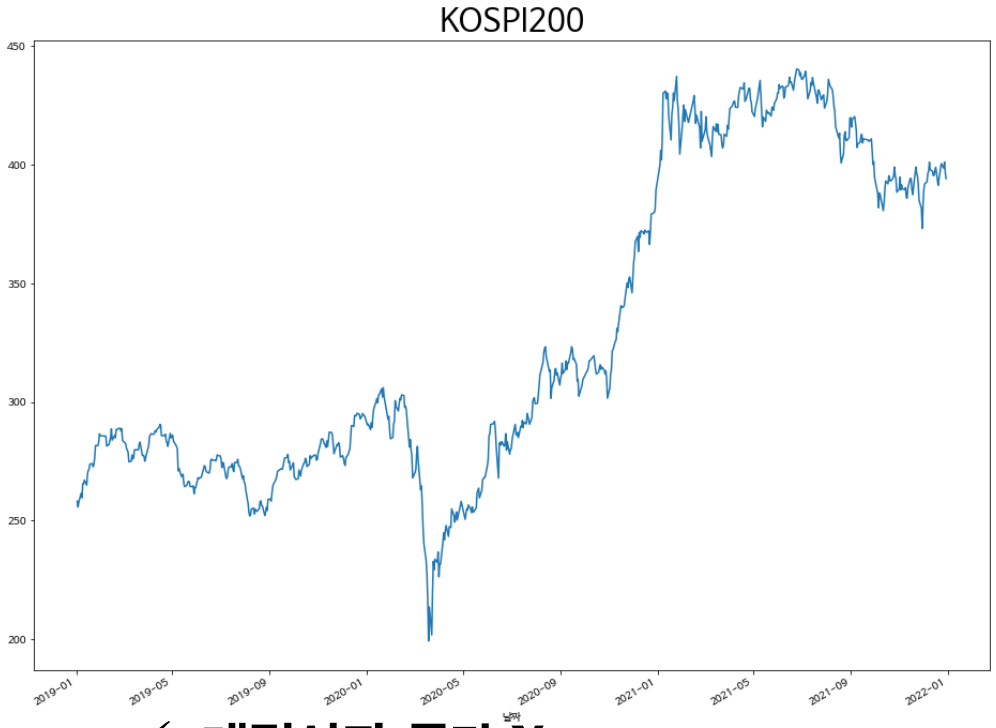
# 시계열 분석

# 시계열 분석

- **정의**
  - 시계열 데이터에 바탕을 둔 분석 방법
  - 시계열 데이터는 통계 숫자를 시간의 흐름에 따라 일정한 간격마다 기록한 통계 계열 데이터
- **적용 모델**
  - ARIMA(자기 회귀 누적 이동 평균)

# 시계열 분석

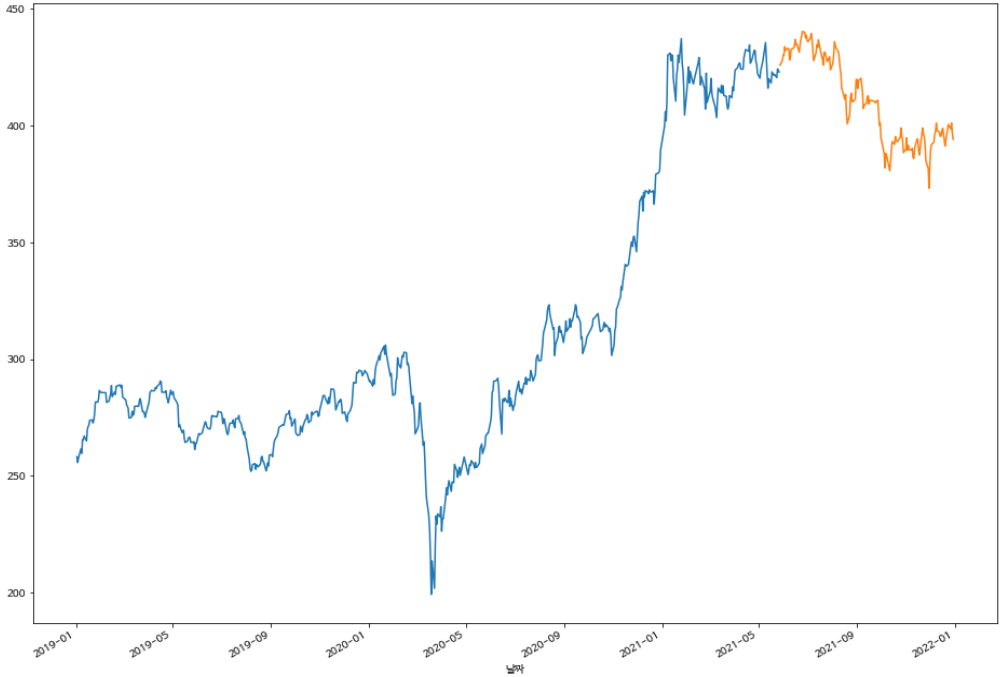
- **ARIMA**
  - ➢ **KOSPI200 지수**
    - **종가 시각화**




✓ **계절성과 주기 X**

✓ **대체적으로 상승 추세**

- **데이터 나누기**



# 시계열 분석

- **ARIMA**
  - ➢ **KOSPI200 지수**

- **최적 ARIMA 모형**

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=3361.466, Time=0.24 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=3360.247, Time=0.03 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=3361.760, Time=0.10 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=3361.849, Time=0.12 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=3360.943, Time=0.03 sec

Best model:  ARIMA(0,1,0)(0,0,0)[0] intercept
Total fit time: 0.516 seconds
```

# 시계열 분석

- **ARIMA**
  - ➢ **KOSPI200 지수**
    - **잔차 검정**

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  593
Model:               SARIMAX(0, 1, 0)   Log Likelihood               -1678.123
Date:                Wed, 30 Mar 2022   AIC                           3360.247
Time:                        09:12:00   BIC                           3369.014
Sample:                             0   HQIC                          3363.662
                                - 593
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.2783      0.169      1.644      0.100      -0.054       0.610
sigma2        16.9702      0.637     26.639      0.000      15.722      18.219
===================================================================================
Ljung-Box (L1) (Q):                   0.49   Jarque-Bera (JB):               192.6
Prob(Q):                              0.48   Prob(JB):                         0.0
Heteroskedasticity (H):               4.18   Skew:                             0.0
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.7
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

✓ **Jarque-Bera**

➢ 잔차가 정규성을 따름

✓ **Ljung-Box**

➢ 잔차는 자기상관성이 존재하지 않음(귀무 가설)

➢ 귀무 가설 채택

잔차는 **자기상관성이 존재하지 않지만**
정규성을 따르지 않음을 확인

# 시계열 분석

## • ARIMA

### ➢ KOSPI200 지수

- 예측 시각화

✓ 한 번에 예측하는 것이 아니라, 일자를 기준으로 하나씩 예측하고, 관측치(Close)를 업데이트

- 모델 평가(MAPE)

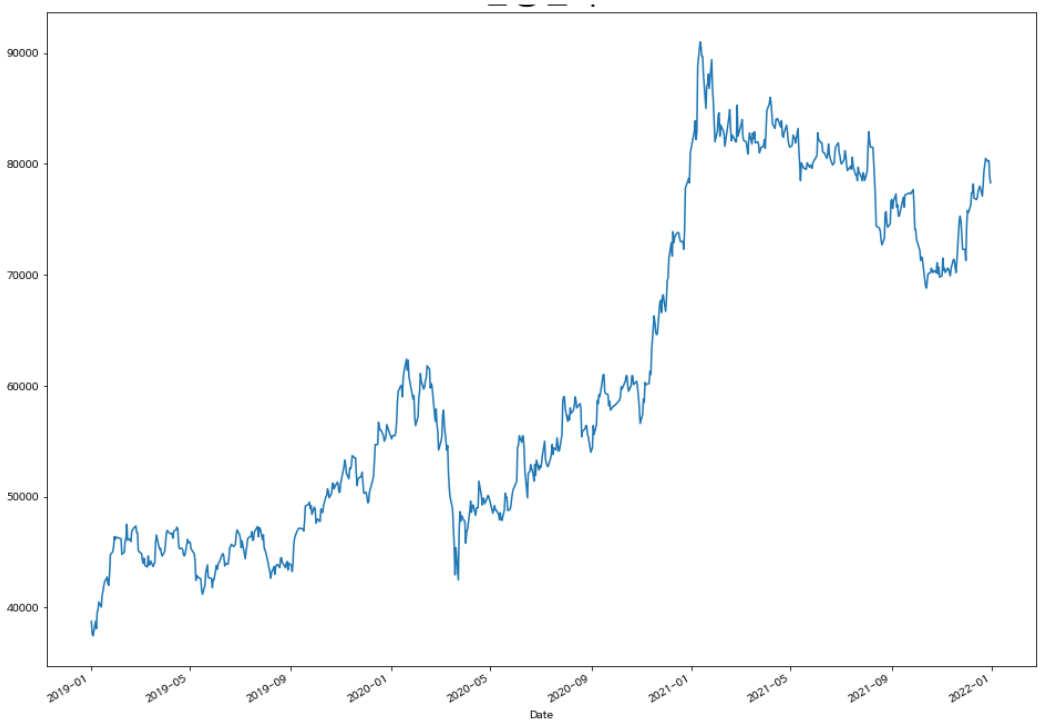
✓ MAPE(Mean Absolute Percentage Error)

✓ [ |(예측값 - 실제값)| / 실제값] * 100

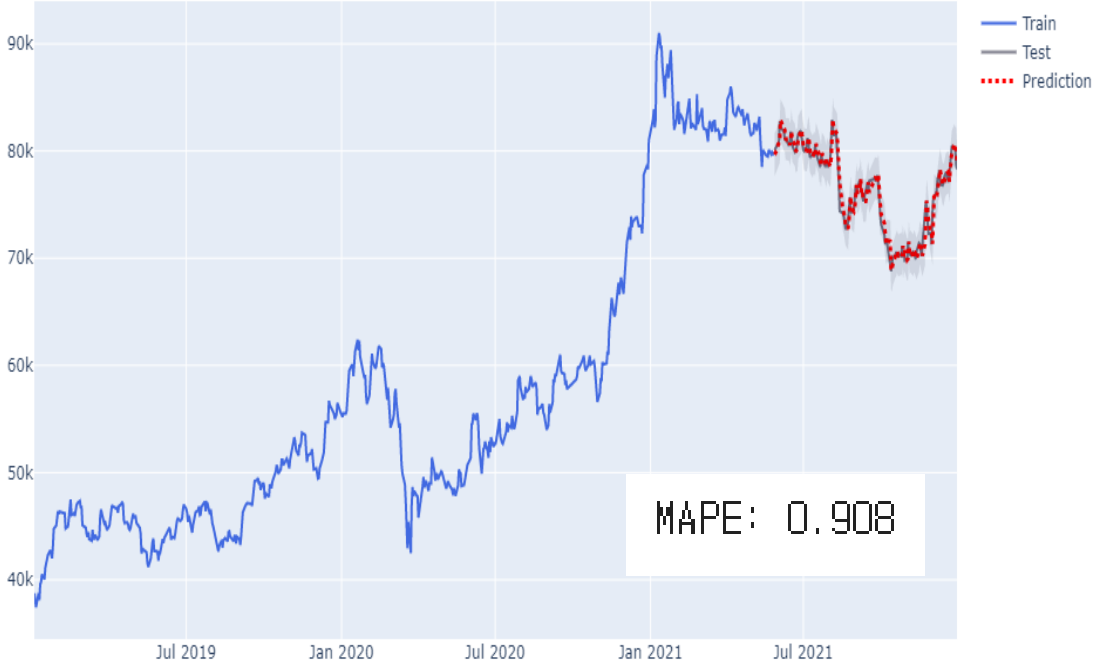
➢ 잔차가 실제값의 0.722% 차지

# 시계열 분석

- **ARIMA**
  - ➢ **삼성전자**
    - **종가 시각화**




- **예측 시각화**




➢ **잔차가 실제값의 0.908% 차지**

# 기술적 분석

# 기술적 분석 지표(평균회귀 기준)

- ✓ 가격 데이터를 사용한 분석으로 각 지표는 특정 기간의 일자를 사용한 투자 방식
- ✓ 기술적 지표에는 대칭/Ascending/Descending Triangle, Bollinger Band, ADX, 일목균형표, MACD, RSI, MA Crossover, Momentum 등 존재
- ✓ 본 프로젝트는 데이터를 지표에 잘 적용할 수 있는 MACD, RSI, MA Crossover, Momentum을 적용

# 기술적 분석 지표(평균회귀 기준)

- 이동평균선(MA) : N일 동안의 종가를 N으로 나눈 평균값 예) 5, 20, 60, 120
- 보류(= 0) : 다음날에도 상승, 하락 추세가 유지 될 경우(공통적인 조건)

## MA Crossover

- 이동평균선(MA)의 차
- 5일 이동 평균선이 20,60,120일 이동평균선을 상향 돌파하면 매수(= 1), 하향 돌파하면 매도(-1), 보류하는 경우(= 0)

## Momentum

- (N일 전 주가/ 현재 주가 - N일 전 주가) * 100
- 5일(Signal), 20일 Momentum 지표 사용
- 20일 Momentum이 음수이면서 Signal선이 20일 Momentum을 상향 돌파하면 매수(= 1), 반대의 경우(= -1), 보류하는 경우(= 0)

## MACD

- 12이동평균선 - 26일 이동평균선
- MACD Signal = 9일 이동 평균선
- MACD가 양수이거나 음수에서 양수로 변하는 경우 MACD - MACD Signal이 양수면 매수(= 1), 반대의 경우 매도(= -1), 보류하는 경우(= 0)

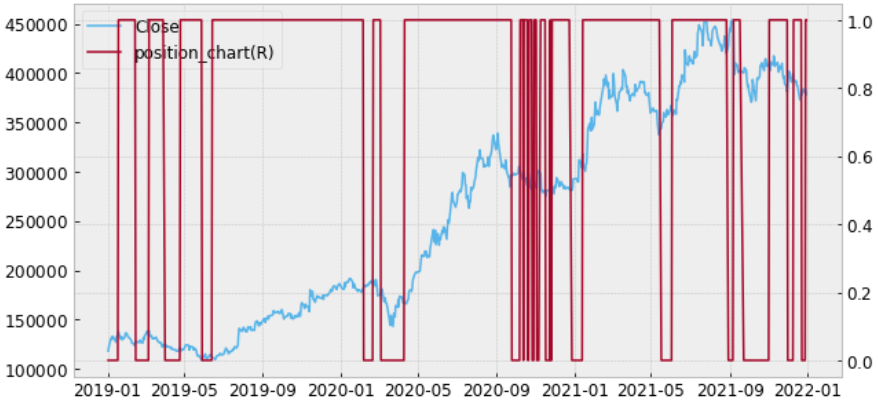
## RSI

- [(14일 종가 상승폭 합계) / (14일 종가 상승폭 합계 + 14일 종가 하락 폭 합계)] * 100
- RSI 지수가 30 이하면 매수(= 1), 70이상이면 매도(= -1), 보류하는 경우(= 0)

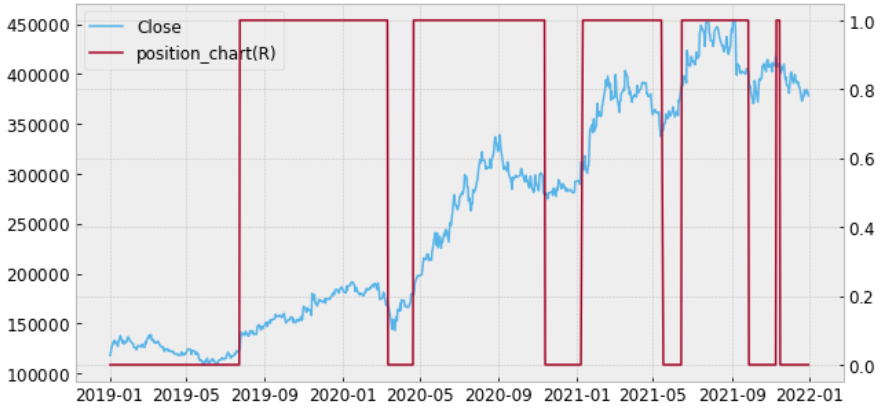
# Backtesting

## ➢ NAVER(상승 추세)

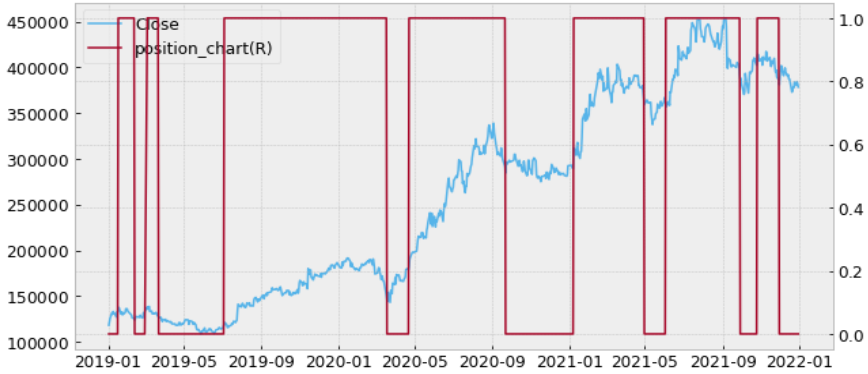
### < MACD >



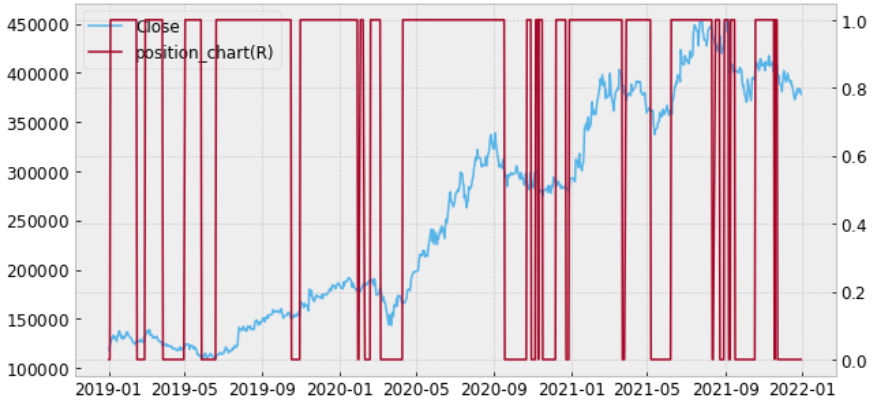

### < MA Minus >




### < RSI >




### < Momentum >




## ➢ 수익률(%)

- MACD : 71.56
- MA Minus : 134.31
- RSI : 83.77
- Momentum : 226.36

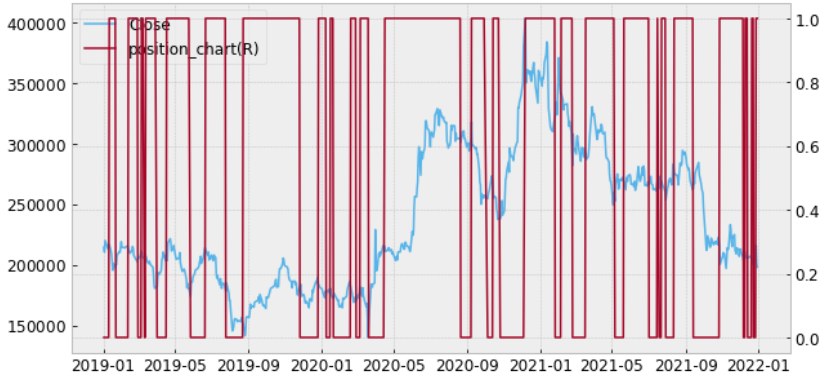
| 벤치마크 | |
|---|---|
| 수익률 | MDD |
| 220.76 | -25.33 |

**상승 추세에서는 기술적 지표가 벤치마크보다 수익률이 낮음**

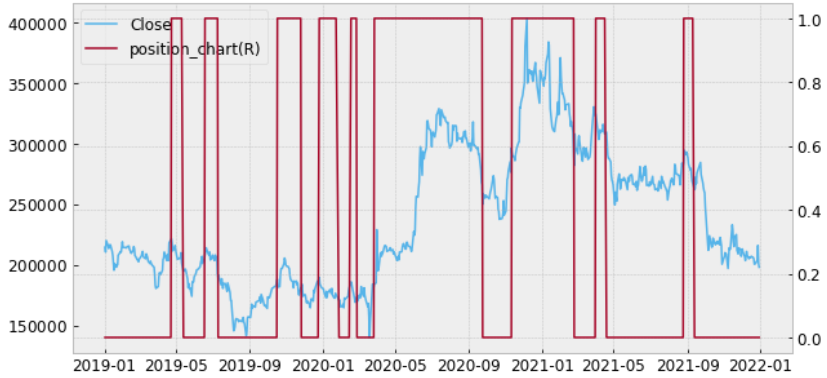
# Backtesting

## ➢ POSCO(횡보 추세)

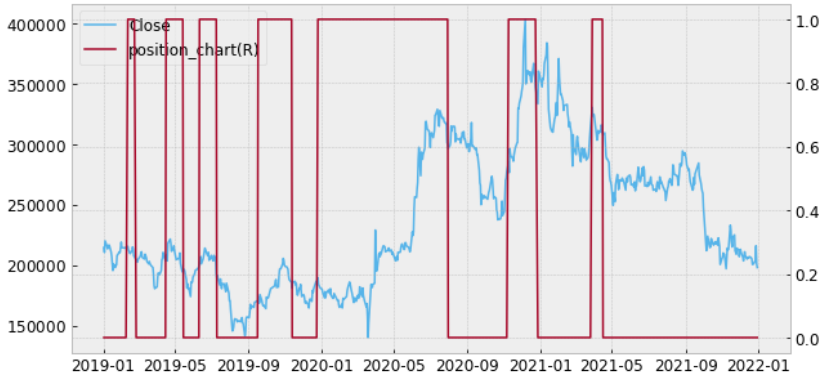
### < MACD >



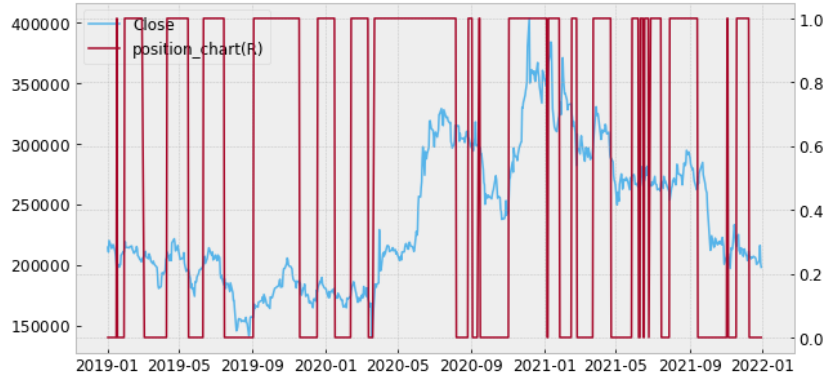

### < MA Minus >




### < RSI >




### < Momentum >




## ➢ 수익률(%)

- MACD : 372.88
- MA Minus : -27.45
- RSI : 50.63
- Momentum : 79.04

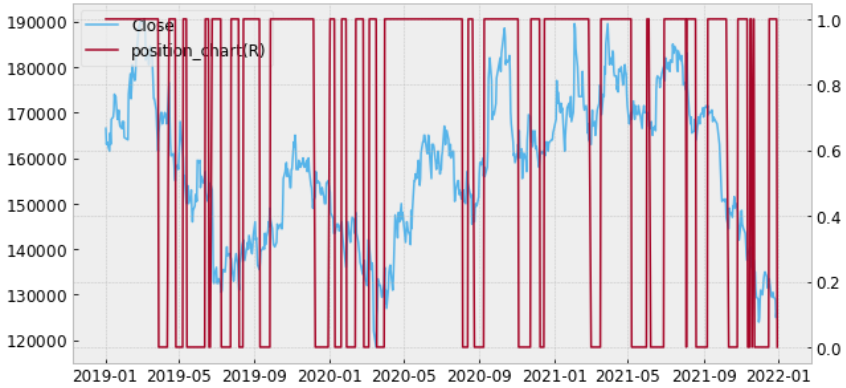
| 벤치마크 | |
|---|---|
| 수익률 | MDD |
| -7.69 | -51.18 |

**횡보 추세에서는 기술적 지표가 벤치마크보다 수익률 높음**

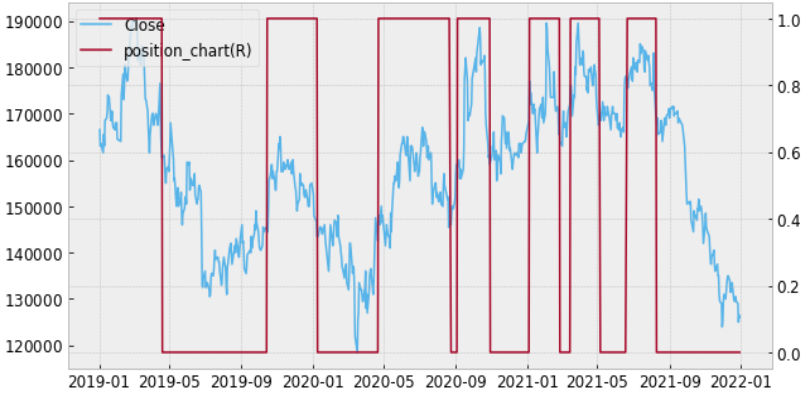
# Backtesting

## ➢ CJ대한통운(하락 추세)

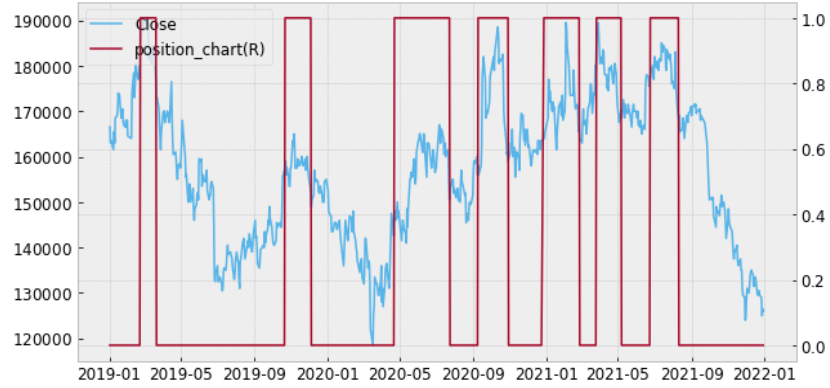
### < MACD >



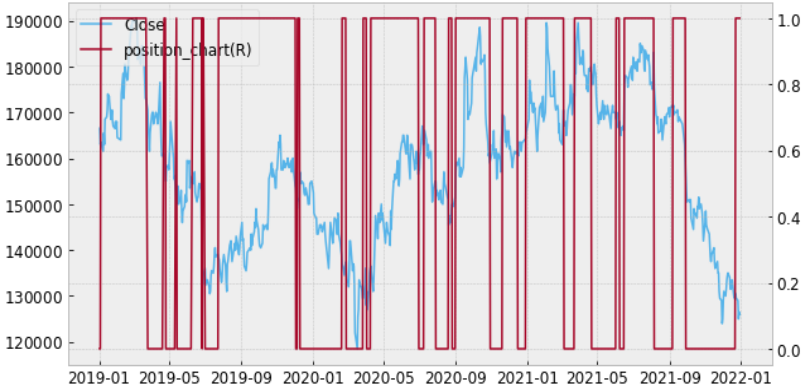

### < MA Minus >




### < RSI >




### < Momentum >




## ➢ 수익률(%)

- MACD : 142.3
- MA Minus : -17.34
- RSI : -13.39
- Momentum : 4.22

| 벤치마크 | |
|---|---|
| 수익률 | MDD |
| -24.32 | -37.8 |

**하락 추세에서는 기술적 지표가 벤치마크보다 손실률이 낮음**

# Backtesting

- KOSPI 200 기업의 각 평가 지표의 평균을 계산
- 단순한 평가지표인 이동평균선을 제외하고는 벤치마크 수익율보다 높음
- 절대적인 수치라고 볼 수는 없지만 Data 분석 결과 사고 파는 전략 스탠스를 세울 때 수익이 더 높다고 추론할 수 있음

| 종목명 | 기술적지표 | | | |
|---|---|---|---|---|
| | MA Minus 수익률 | MACD 수익률 | RSI 수익률 | Momentum 수익률 |
| 합계 | 2381.35 | 19769.67 | 4194.71 | 11093.46 |
| 평균 | 28.34940476 | 235.3532143 | 49.93702381 | 132.065 |

| 벤치마크 | |
|---|---|
| 수익률 | MDD |
| 4493.53 | -4352.78 |
| 53.49440476 | -51.81880952 |

# Pycaret

- 평가 기준 : 종가 차(매수 : 1, 매도 : 0), Classification 접근

➢ NAVER(상승 추세)

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|---|---|---|---|---|---|---|---|---|---|
| dummy | Dummy Classifie | 0.5336 | 0.5000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.009 |
| lr | Logistic Regressior | 0.5168 | 0.5301 | 0.4352 | 0.4736 | 0.4411 | 0.0234 | 0.0216 | 0.028 |
| svm | SVM - Linear Kerne | 0.5151 | 0.0000 | 0.4673 | 0.4816 | 0.4723 | 0.0246 | 0.0246 | 0.013 |
| ridge | Ridge Classifie | 0.5151 | 0.0000 | 0.3634 | 0.4787 | 0.4098 | 0.0118 | 0.0135 | 0.018 |
| lda | Linear Discriminant Analysi | 0.5101 | 0.5036 | 0.3598 | 0.4713 | 0.4052 | 0.0018 | 0.0030 | 0.023 |
| qda | Quadratic Discriminant Analysi | 0.5051 | 0.4985 | 0.5148 | 0.4845 | 0.4277 | 0.0128 | 0.0183 | 0.014 |
| nb | Naive Baye | 0.5035 | 0.4341 | 0.1148 | 0.4237 | 0.1693 | -0.0443 | -0.0510 | 0.015 |
| xgboost | Extreme Gradient Boosting | 0.5019 | 0.4808 | 0.4683 | 0.4633 | 0.4641 | -0.0010 | -0.0013 | 0.298 |
| catboost | CatBoost Classifie | 0.4984 | 0.4758 | 0.4028 | 0.4620 | 0.4287 | -0.0151 | -0.0143 | 2.126 |
| lightgbm | Light Gradient Boosting Machine | 0.4951 | 0.4716 | 0.4249 | 0.4578 | 0.4392 | -0.0186 | -0.0187 | 0.086 |
| rf | Random Forest Classifie | 0.4932 | 0.4727 | 0.3705 | 0.4543 | 0.4070 | -0.0288 | -0.0279 | 1.011 |
| gbc | Gradient Boosting Classifie | 0.4917 | 0.4851 | 0.4106 | 0.4484 | 0.4259 | -0.0274 | -0.0286 | 0.236 |
| dt | Decision Tree Classifie | 0.4914 | 0.4890 | 0.4565 | 0.4612 | 0.4573 | -0.0212 | -0.0208 | 0.019 |
| knn | K Neighbors Classifie | 0.4848 | 0.4773 | 0.4172 | 0.4434 | 0.4276 | -0.0388 | -0.0393 | 0.208 |
| ada | Ada Boost Classifie | 0.4751 | 0.4606 | 0.3779 | 0.4302 | 0.4008 | -0.0624 | -0.0632 | 0.157 |
| et | Extra Trees Classifie | 0.4396 | 0.4362 | 0.3635 | 0.3909 | 0.3746 | -0.1305 | -0.1313 | 1.042 |

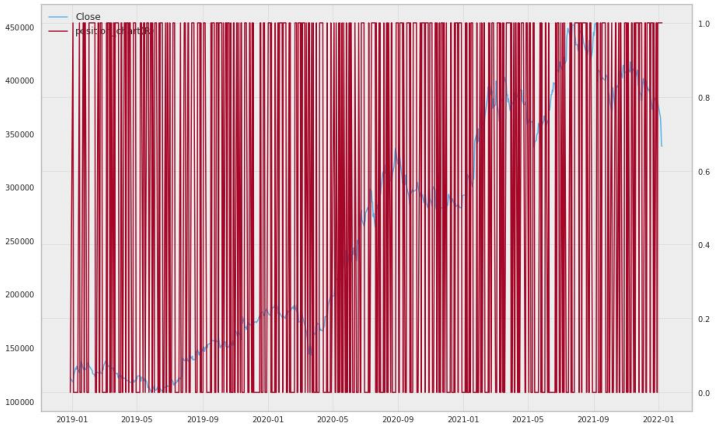
# Pycaret

## ➢ NAVER(상승 추세)

### < Dummy Classification >

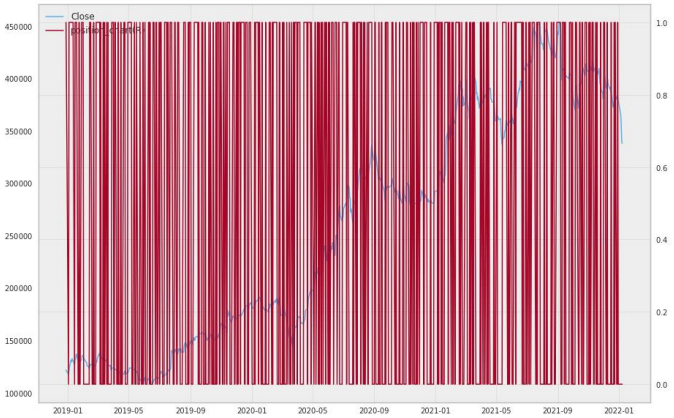
```
CAGR: 0.00%
Accumulated return: 0.00%
Average return: nan%
Benchmark return : 175.92%
Number of trades: 0
Number of win: 0
Hit ratio: 0.00%
Investment period: 3.0yrs
Sharpe ratio: 0.00
MDD: 0.00%
Benchmark MDD: -26.52%
```

### < Logistic Regression >




```
CAGR: -45.16%
Accumulated return: -83.80%
Average return: -0.91%
Benchmark return : 175.92%
Number of trades: 180
Number of win: 63
Hit ratio: 35.00%
Investment period: 3.0yrs
Sharpe ratio: -1.68
MDD: -83.80%
Benchmark MDD: -26.52%
```

### < SVM >




```
CAGR: 101.37%
Accumulated return: 733.94%
Average return: 1.25%
Benchmark return : 175.92%
Number of trades: 179
Number of win: 113
Hit ratio: 63.13%
Investment period: 3.0yrs
Sharpe ratio: 2.89
MDD: -30.15%
Benchmark MDD: -26.52%
```

# Pycaret

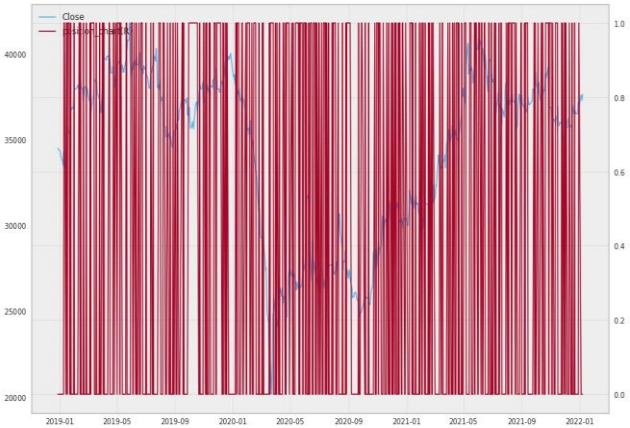
## ➢ POSCO(횡보 추세)

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|---|---|---|---|---|---|---|---|---|---|
| knn | K Neighbors Classifier | 0.5637 | 0.5716 | 0.5365 | 0.5485 | 0.5415 | 0.1254 | 0.1258 | 0.165 |
| et | Extra Trees Classifier | 0.5570 | 0.5608 | 0.5504 | 0.5376 | 0.5428 | 0.1134 | 0.1138 | 0.807 |
| lightgbm | Light Gradient Boosting Machine | 0.5500 | 0.5459 | 0.5326 | 0.5381 | 0.5329 | 0.0990 | 0.0996 | 0.078 |
| dt | Decision Tree Classifier | 0.5468 | 0.5468 | 0.5469 | 0.5362 | 0.5385 | 0.0940 | 0.0950 | 0.017 |
| catboost | CatBoost Classifier | 0.5434 | 0.5457 | 0.5294 | 0.5284 | 0.5282 | 0.0861 | 0.0865 | 2.603 |
| rf | Random Forest Classifier | 0.5416 | 0.5383 | 0.4978 | 0.5228 | 0.5088 | 0.0799 | 0.0799 | 0.801 |
| xgboost | Extreme Gradient Boosting | 0.5318 | 0.5387 | 0.5055 | 0.5142 | 0.5069 | 0.0620 | 0.0627 | 0.261 |
| ridge | Ridge Classifier | 0.5252 | 0.0000 | 0.4771 | 0.5092 | 0.4910 | 0.0469 | 0.0474 | 0.015 |
| qda | Quadratic Discriminant Analysis | 0.5252 | 0.5222 | 0.6474 | 0.5045 | 0.5636 | 0.0579 | 0.0633 | 0.011 |
| lda | Linear Discriminant Analysis | 0.5251 | 0.5284 | 0.4771 | 0.5090 | 0.4910 | 0.0469 | 0.0474 | 0.019 |
| dummy | Dummy Classifier | 0.5185 | 0.5000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.006 |
| svm | SVM - Linear Kernel | 0.5081 | 0.0000 | 0.5012 | 0.4874 | 0.4932 | 0.0159 | 0.0162 | 0.010 |
| gbc | Gradient Boosting Classifier | 0.5065 | 0.4908 | 0.4941 | 0.4879 | 0.4898 | 0.0118 | 0.0119 | 0.298 |
| lr | Logistic Regression | 0.4950 | 0.4652 | 0.1415 | 0.3434 | 0.1882 | -0.0378 | -0.0575 | 0.030 |
| ada | Ada Boost Classifier | 0.4913 | 0.4863 | 0.4765 | 0.4707 | 0.4718 | -0.0188 | -0.0193 | 0.181 |
| nb | Naive Bayes | 0.4613 | 0.4525 | 0.4331 | 0.4343 | 0.4249 | -0.0782 | -0.0832 | 0.010 |

# Pycaret

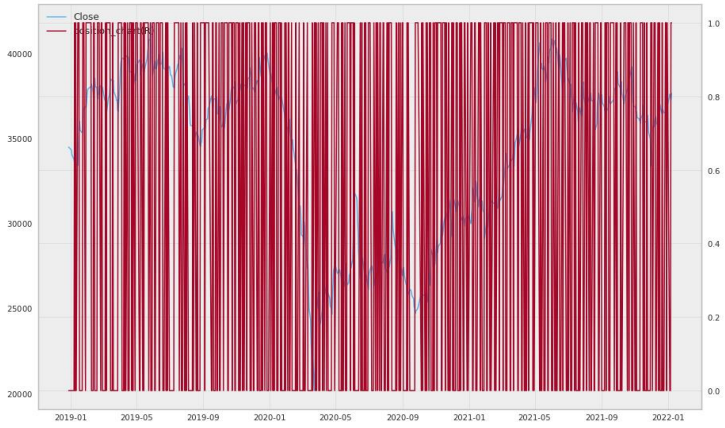
## ➢ POSCO(횡보 추세)

### < K-Neighbors >



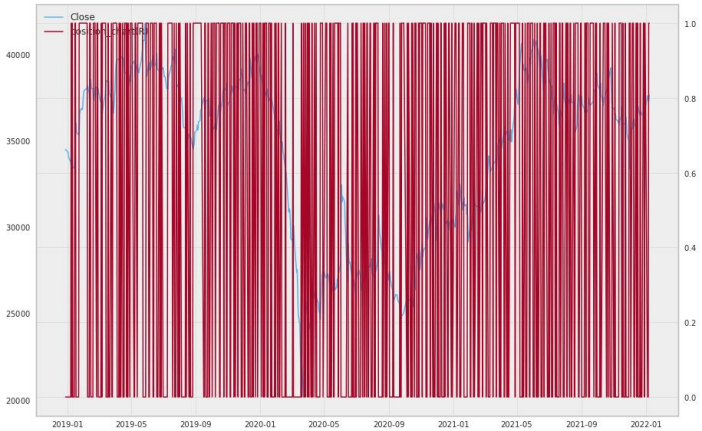

```
CAGR: -0.72%
Accumulated return: -2.16%
Average return: 0.03%
Benchmark return : -4.73%
Number of trades: 168
Number of win: 76
Hit ratio: 45.24%
Investment period: 3.0yrs
Sharpe ratio: -0.11
MDD: -37.23%
Benchmark MDD: -52.84%
```

### < Extra Trees >




```
CAGR: 26.77%
Accumulated return: 105.18%
Average return: 0.42%
Benchmark return : -4.73%
Number of trades: 191
Number of win: 92
Hit ratio: 48.17%
Investment period: 3.0yrs
Sharpe ratio: 0.89
MDD: -22.30%
Benchmark MDD: -52.84%
```

### < LGBM >




```
CAGR: 12.08%
Accumulated return: 41.26%
Average return: 0.22%
Benchmark return : -4.73%
Number of trades: 183
Number of win: 87
Hit ratio: 47.54%
Investment period: 3.0yrs
Sharpe ratio: 0.42
MDD: -21.57%
Benchmark MDD: -52.84%
```

# Pycaret

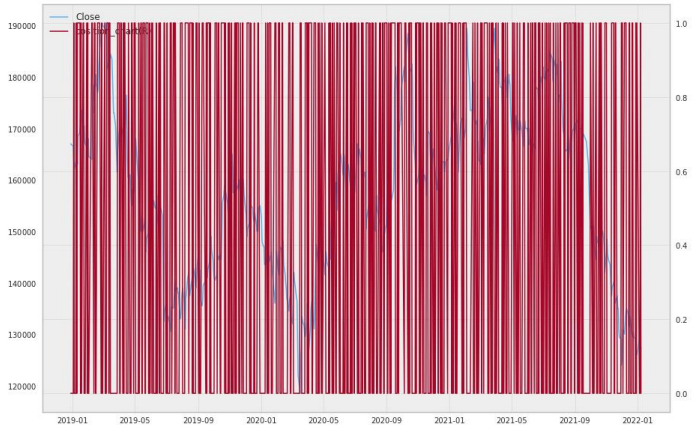
## ➢ CJ대한통운(하락 추세)

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|---|---|---|---|---|---|---|---|---|---|
| gbc | Gradient Boosting Classifier | 0.5553 | 0.5297 | 0.4222 | 0.5186 | 0.4601 | 0.0891 | 0.0931 | 0.296 |
| dummy | Dummy Classifier | 0.5469 | 0.5000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.007 |
| nb | Naive Bayes | 0.5385 | 0.5075 | 0.0593 | 0.3764 | 0.0986 | -0.0052 | 0.0097 | 0.010 |
| ridge | Ridge Classifier | 0.5367 | 0.0000 | 0.3370 | 0.4920 | 0.3964 | 0.0411 | 0.0452 | 0.012 |
| lda | Linear Discriminant Analysis | 0.5367 | 0.5320 | 0.3407 | 0.4929 | 0.3989 | 0.0419 | 0.0460 | 0.018 |
| catboost | CatBoost Classifier | 0.5351 | 0.5335 | 0.3741 | 0.4884 | 0.4214 | 0.0437 | 0.0460 | 2.225 |
| qda | Quadratic Discriminant Analysis | 0.5301 | 0.5124 | 0.2444 | 0.4554 | 0.3162 | 0.0115 | 0.0097 | 0.010 |
| rf | Random Forest Classifier | 0.5285 | 0.5224 | 0.3926 | 0.4753 | 0.4267 | 0.0335 | 0.0345 | 0.879 |
| svm | SVM - Linear Kernel | 0.5251 | 0.0000 | 0.5185 | 0.4754 | 0.4949 | 0.0484 | 0.0492 | 0.010 |
| et | Extra Trees Classifier | 0.5236 | 0.5168 | 0.4111 | 0.4738 | 0.4386 | 0.0288 | 0.0295 | 0.808 |
| xgboost | Extreme Gradient Boosting | 0.5168 | 0.5275 | 0.4074 | 0.4666 | 0.4321 | 0.0150 | 0.0161 | 0.253 |
| lightgbm | Light Gradient Boosting Machine | 0.5152 | 0.5494 | 0.4333 | 0.4620 | 0.4430 | 0.0158 | 0.0162 | 0.069 |
| ada | Ada Boost Classifier | 0.5115 | 0.4895 | 0.4037 | 0.4576 | 0.4259 | 0.0042 | 0.0041 | 0.136 |
| lr | Logistic Regression | 0.5068 | 0.4981 | 0.4519 | 0.4083 | 0.4284 | 0.0031 | 0.0034 | 0.020 |
| knn | K Neighbors Classifier | 0.4968 | 0.4658 | 0.3926 | 0.4405 | 0.4128 | -0.0245 | -0.0244 | 0.156 |
| dt | Decision Tree Classifier | 0.4899 | 0.4843 | 0.4259 | 0.4354 | 0.4285 | -0.0314 | -0.0318 | 0.018 |

# Pycaret

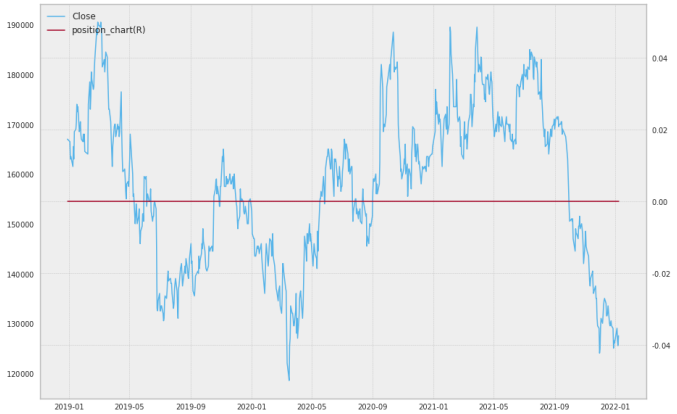
## ➢ CJ대한통운(하락 추세)

### < Gradient Boosting >



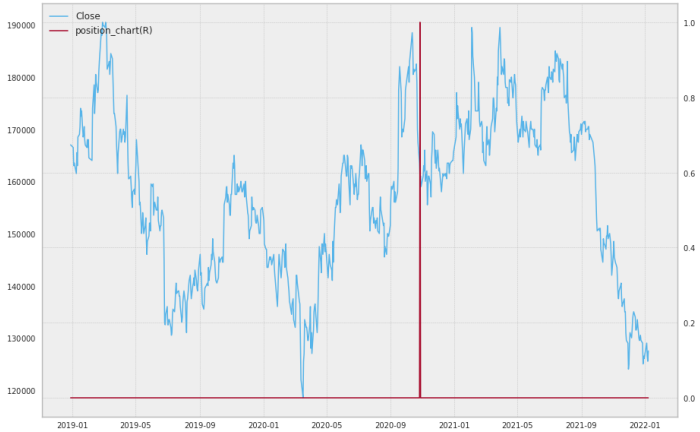

```
CAGR: -2.38%
Accumulated return: -7.05%
Average return: 0.00%
Benchmark return : -25.96%
Number of trades: 193
Number of win: 84
Hit ratio: 43.52%
Investment period: 3.0yrs
Sharpe ratio: -0.17
MDD: -42.15%
Benchmark MDD: -40.68%
```

### < Dummy Classification >




```
CAGR: 0.00%
Accumulated return: 0.00%
Average return: nan%
Benchmark return : -25.96%
Number of trades: 0
Number of win: 0
Hit ratio: 0.00%
Investment period: 3.0yrs
Sharpe ratio: 0.00
MDD: 0.00%
Benchmark MDD: -40.68%
```

### < Naïve Bayes >




```
CAGR: -1.09%
Accumulated return: -3.27%
Average return: -3.27%
Benchmark return : -25.96%
Number of trades: 1
Number of win: 0
Hit ratio: 0.00%
Investment period: 3.0yrs
Sharpe ratio: -1.35
MDD: -3.27%
Benchmark MDD: -40.68%
```

# 기본적 분석 및 결론

# 기본적 분석

- **정의**
  - ➢ 기업의 내재가치를 분석하여 주가의 흐름을 파악
  - ➢ 주로 기업의 재무제표를 활용(PER, PBR, 배당수익률, 영업 이익 등)

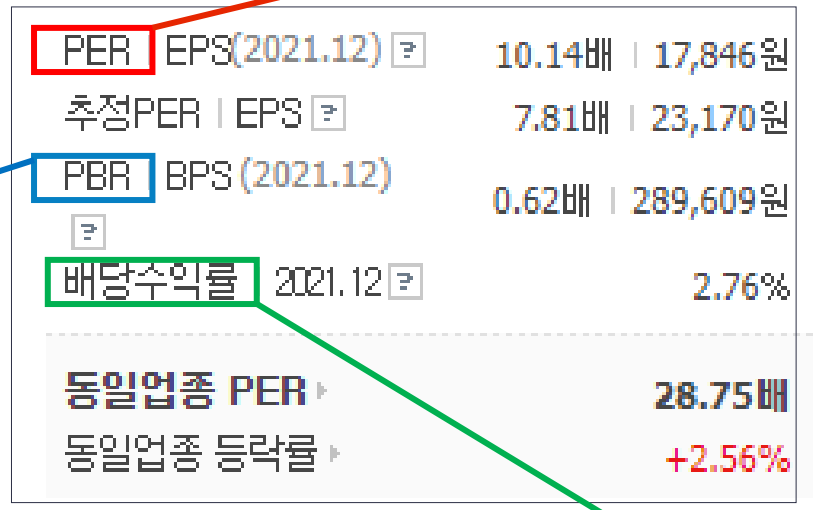
- **적용 방법**
  - ➢ PER, PBR, 영업이익, 배당수익률과 주가수익률 간의 상관관계 분석

# 기본적 분석

- PER, PBR, 배당수익률

- **PER**
  - ➢ 시가총액/당기순이익
  - ➢ PER이 높을수록 주가의 가치가 고평가

- **PBR**
  - ➢ 시가총액/자기자본(순자산)
  - ➢ PBR이 높을수록 주가의 가치가 고평가




- **배당수익률**
  - ➢ 배당(주식배당,현금배당)
  - ➢ 배당수익률이 2% 주식 5만원에 구매시 1000원 배당 소득

# 기본적 분석

- 영업 이익
  - ➢ 매출액 - 매출원가
  - ➢ 회사가 가진 비즈니스모델을 통해서 거두어들이는 이윤

| 주요재무정보 | 최근 연간 실적 | | | |
|---|---|---|---|---|
| | 2019.12 | 2020.12 | 2021.12 | 2022.12(E) |
| | IFRS 연결 | IFRS 연결 | IFRS 연결 | IFRS 연결 |
| 매출액(억원) | 1,057,464 | 1,039,976 | 1,176,106 | 1,300,742 |
| 영업이익(억원) | 36,055 | 23,947 | 66,789 | 76,636 |
| 당기순이익(억원) | 31,856 | 19,246 | 56,931 | 71,405 |

# 기본적 분석

- **PER, PBR, 배당수익률과 주가수익률**
  - ➢ 독립변수 : PER, PBR, 배당수익률
  - ➢ 종속변수 : 해당 연간수익률

▪ 2011 ~ 2021 회귀분석(Excel)

| 연도 | 결정계수 | F값 | P-값(Y 절편) | P-값(PER) | P-값(PBR) | P-값(배당수익률) |
|---|---|---|---|---|---|---|
| 2011 | 0.004923613 | 0.522508865 | 0.002880866 | 0.83103111 | 0.252294248 | 0.495272679 |
| 2012 | 0.00729799 | 0.290158291 | 0.071557994 | 0.564736906 | 0.29779913 | 0.238993825 |
| 2013 | 0.001453303 | 0.854528533 | 0.005626573 | 0.625070192 | 0.493292973 | 0.839713086 |
| 2014 | 0.003979908 | 0.556849871 | 5.10E-07 | 0.925646928 | 0.153978204 | 0.67966476 |
| 2015 | 0.026592084 | 0.0030922 | 1.53E-08 | 0.587690965 | 0.846570033 | 0.000400748 |
| 2016 | 0.057536053 | 5.47E-07 | 0.996732531 | 0.081985855 | 6.43E-06 | 0.029478367 |
| 2017 | 0.001352457 | 0.857740059 | 0.692456599 | 0.810144526 | 0.989842439 | 0.424726734 |
| 2018 | 0.007437159 | 0.212501871 | 0.01246015 | 0.677109906 | 0.037473262 | 0.634634342 |
| 2019 | 0.001412579 | 0.835369656 | 0.667911307 | 0.688498224 | 0.554243025 | 0.810848367 |
| 2020 | 0.022997387 | 0.003582304 | 1.81E-07 | 0.246492234 | 0.76114947 | 5.78E-05 |

▪ 2011 ~ 2021년 상관분석

| | PER | PBR | 배당수익률_x | 수익률 |
|---|---|---|---|---|
| PER | 1.000000 | 0.453094 | -0.181191 | -0.106563 |
| PBR | 0.453094 | 1.000000 | -0.270474 | -0.114479 |
| 배당수익률_x | -0.181191 | -0.270474 | 1.000000 | 0.053682 |
| 수익률 | -0.106563 | -0.114479 | 0.053682 | 1.000000 |

☑ 수익률과 독립변수들 간의 **약한 상관관계** 지님

# 기본적 분석

## • 영업 이익과 주가수익률

- ➢ 독립변수 : 영업 이익
- ➢ 종속변수 : 해당 연간수익률
- ➢ 자료에 결측값이 존재하면 제외

### < 하락 추세 >

| 연도 | 기준 | 기업 수(개) | 평균수익률(%) |
|---|---|---|---|
| 2017(14-16) | 영업이익 감소 | 17 | 24.85 |
| | 나머지 | 167 | 16.24 |
| 2018(15-17) | 영업이익 감소 | 16 | -27.2 |
| | 나머지 | 178 | -18.23 |
| 2019(16-18) | 영업이익 감소 | 28 | -7.69 |
| | 나머지 | 171 | -3.05 |
| 2020(17-19) | 영업이익 감소 | 29 | 15.68 |
| | 나머지 | 140 | 22.93 |

### < 증가 추세 >

| 연 도 | 기 준 | 기업 수(개) | 평균수익률(%) |
|---|---|---|---|
| 2017(14-16) | 영업 이익 증가 | 50 | 17.22 |
| | 나머지 | 135 | 13.85 |
| 2018(15-17) | 영업 이익 증가 | 40 | -19.2 |
| | 나머지 | 154 | -18.92 |
| 2019(16-18) | 영업 이익 증가 | 34 | 1.67 |
| | 나머지 | 166 | -5.09 |
| 2020(17-19) | 영업 이익 증가 | 34 | 10.27 |
| | 나머지 | 139 | 29 |
| 2021(18-20) | 영업 이익 증가 | 27 | 9.59 |
| | 나머지 | 146 | 7.89 |

# 결론

- 저점에서 매수를 하고 고점해서 매수하는 방식이 Best
- 하지만 상승, 하락을 예측하고 투자하는 것을 불가능
- 기술적, 기본적 지표를 기반으로 하는 본인만의 스탠스를 세운다면 Worst는 피할 수 있다고 추론

## ✓ 아쉬운 점

- ➢ 딥러닝 방식인 LSTM의 결과가 궁금
- ➢ 많은 재무제표 기준을 활용한 전략을 수립하지 못해 아쉬움
- ➢ 경기 순환을 고려하지 못한 프로젝트
- ➢ 현업과 본 프로젝트의 접근 차이점 궁금

# THANK YOU