

재무 정보를 바탕으로 선별한 종목의 주가 예측

2015104214 장유진

목차

1. 프로젝트 소개
2. 관련 연구
3. 제안 방법
4. 검증
5. 결론
6. 시연

프로젝트 소개

주가 지수 조회의 불편함 → 투자 종목 선택이 오래 걸림

NAVER (035420) | KOSPI

358,000 0 (0.00%)

| | | | | |
|-------------|----|---------|-----------|-----------------|
| 시가 | | 361,500 | 거래량(주) | 439,112 |
| 고가 | | 361,500 | 거래대금(원) | 157,438,201,500 |
| 저가 | | 355,500 | 시가총액(백만원) | 58,806,295 |
| 52주 (종가) | 최고 | 403,500 | 외국인비율 | 56.49 |
| | 최저 | 225,500 | PER/PBR | 52.06/7.98 |
| 대용가 | | 279,240 | 배당수익률 | 0.11 |

한국거래소 (KRX, 대한민국의 금융거래소)의 조회 서비스

| 투자정보 | 호가 10단계 |
|----------------------|-------------------------|
| 시가총액 | 58조 8,063억원 |
| 시가총액순위 ▶ | 코스피 4위 |
| 상장주식수 | 164,263,395 |
| 액면가 매매단위 | 100원 1주 |
| 외국인한도주식수(A) | 164,263,395 |
| 외국인보유주식수(B) | 92,793,490 |
| 외국인소진율(B/A) ? | 56.49% |
| 투자의견 목표주가 | 4.00매수 493,478 |
| 52주최고 최저 ? | 410,000 225,000 |
| PER EPS(2021.03) ? | 3.65배 98,182원 |
| 추정PER EPS ? | 3.84배 93,187원 |
| PBR BPS(2021.03) ? | 2.40배 149,109원 |
| 배당수익률 2020.12 ? | 0.11% |
| 동일업종 PER ▶ | 100.77배 |
| 동일업종 등락률 ▶ | -1.09% |

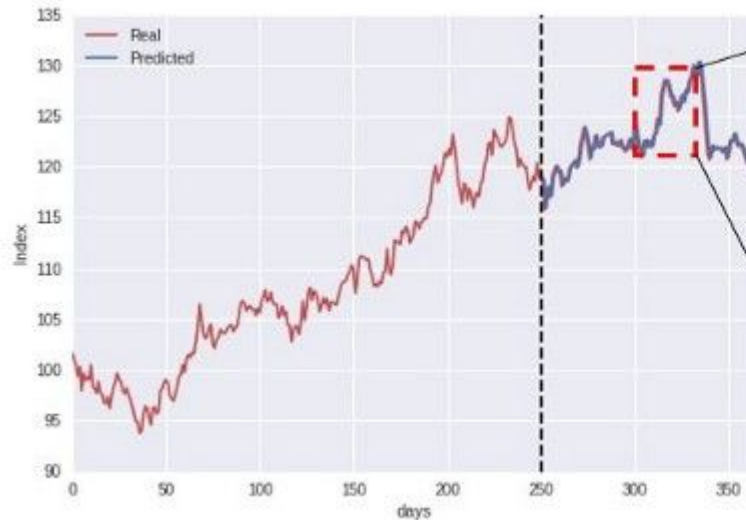
네이버 금융의 조회 서비스

- PER, PBR만 조회 가능. **ROE, PSR, PCR, 부채비율, 영업이익률, 순이익률** 조회 불가능
- 주가 지수로 종목 정렬 불가능
- 심지어 같은 날 조회한 **NAVER** 주가 정보인데
공식 사이트인 금융거래소의 자료보다 네이버 금융의 자료가 최신인 아이러니한 상황
(PER, PBR 값)

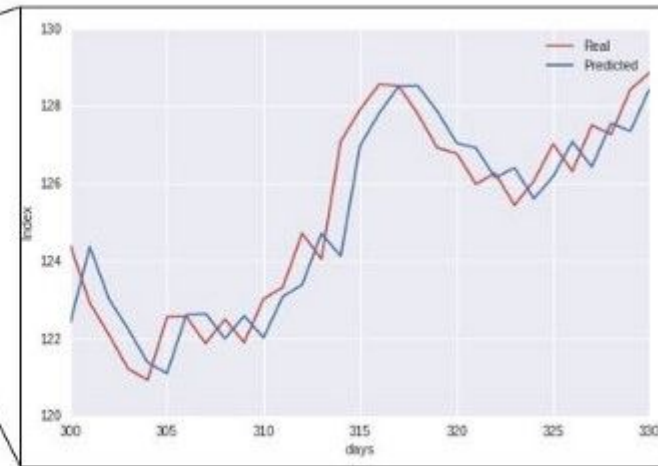
프로젝트 소개

실효성 없는 주가 예측 모델

흔하게 보이는 딥러닝 주가 예측 예제



전체 기간의 예측값, 실제값 그래프



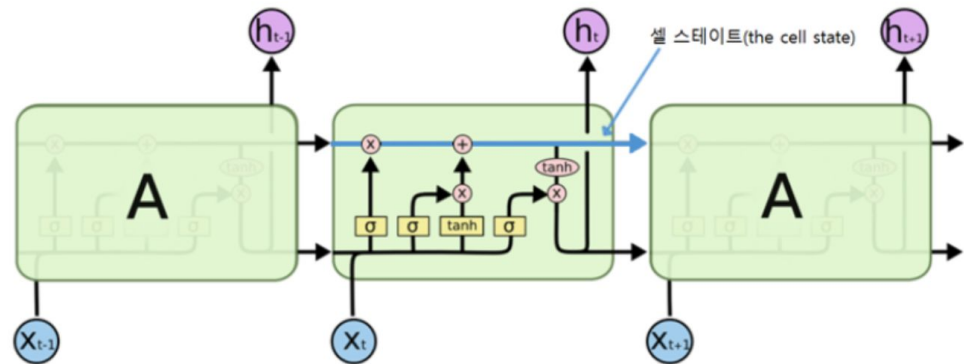
부분 확대했을 때 실상

매우 정확히 예측하는 것처럼 보이지만, 실제로는 예측값이 실제값을 하루씩 늦게 따라가는 모양

대부분 과거 n 일의 값을 보고 다음날(1일)의 값을 예측하는 코드

이는 다음날의 주가를 예측하는 것이 아니라 과거 마지막 날의 값을 따라간다.

LSTM



- 시계열 데이터 학습에 사용
- 셀 스테이트를 사용하여 거리가 먼 데이터 값을 사용하지 못하는 RNN의 단점 보완

양방향 LSTM 순환신경망 기반 주가예측모델

PDF 원문보기

인용

Stock Prediction Model based on Bidirectional LSTM Recurrent Neural Network

한국정보전자통신기술학회논문지 = Journal of Korea institute of information, electronics, and communication technology v.11

no.2 , 2018년, pp.204 - 208

주일택 (Department of Computer Science, Dongshin University) ; 최승호 (Department of Computer Science, Dongshin University)

2.4 제안 방법과 관련한 내용

본 논문에서 주가 예측을 위한 딥러닝 모델은 양방향 순환신경망에 LSTM을 사용하였다. 네트워크 전체적인 구조는 주식의 일별 시가, 고가, 저가, 종가, 거래량을 입력으로 입력계층은 5개로 구성되어 있으며 20개의 LSTM 셀로 구성된 은닉계층과 주가 예측을 위한 1개 출력계층으로 이루어진다.

딥러닝의 최적화기(Optimizer)로 아담(Adam) 최적화기를 이용하여, 과거 및 현재의 주가와 거래량을 이용하여 미래의 주가를 학습한다.

↑ 앞서 말한 잘못된 예측 방법

표 1. 평균 제곱근 오차(RMSE) 및 예측률 결과

Table 1. Results of RMSE and Prediction rate

| | 단방향 LSTM | 양방향 LSTM |
|---------|----------|----------|
| 최소 RMSE | 0.0209 | 0.015 |
| 예측률 | 83% | 84.9% |

양방향 LSTM 순환신경망의 최소 RMSE가 0.015이며, 단방향 LSTM 순환신경망보다 작은 오차 값을 보였고 주가 예측률도 단방향 LSTM 모델보다 향상된 성능을 보이고 있다.

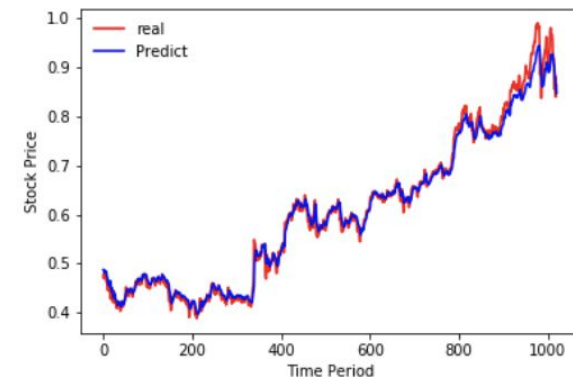


그림 5. 실제 주가와 예측 주가의 그래프

Fig. 5. Graph of real stock price and predicted stock price

↑ 80% 이상의 예측률과 예측 그래프를 보면 실제 투자에 사용하지 않을 이유가 없다. 즉, 잘못된 예측이란 것

제안 방법

사용할 종목들의 재무 정보 조사 후 DB에 저장

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|-----|-----|-------------------------|---------|------------|---------|---------|------------|------------|-----------|------------|-------|-------|---------------|-------|
| 1 | 순번 | 종목명 | 종목코드 | 매출액 | 영업이익 | 순이익 | 자산 | 부채 | 자본 | 주식수 | 배당금 | 배당세공제 | 퀀트계산가 (증가) | 시가총액 |
| 2 | 1 | 두산 | A000150 | 16,969,331 | 274,997 | 963,892 | 30,117,623 | 22,408,094 | 123,738 | 16,523,835 | 2,000 | 1,692 | 40,500 | 6,692 |
| 3 | 2 | 한미반도체 | A042700 | 257,386 | 66,649 | 50,131 | 322,293 | 70,799 | 251,494 | 51,501,501 | 200 | 169 | | 0 |
| 4 | 3 | 대산 | A001680 | 3,113,204 | 174,392 | 127,011 | 2,528,335 | 1,426,785 | 1,101,550 | 34,648,025 | 700 | 592 | | 0 |
| 5 | 4 | SK가스 | A018670 | 4,412,319 | 190,187 | 265,350 | 4,498,920 | 2,600,205 | 1,898,716 | 9,230,244 | 4,000 | 3,384 | | 0 |
| 6 | 5 | 현대홈쇼핑 | A057050 | 2,298,676 | 154,892 | 112,714 | 2,688,595 | 891,006 | 1,797,589 | 12,000,000 | 2,200 | 1,861 | | 0 |
| 7 | 6 | 파미셀 | A005690 | 37,752 | 7,233 | 5,242 | 69,744 | 12,639 | 57,105 | 59,958,950 | | 0 | | 0 |
| 295 | 294 | 대한제당 | A001790 | 12144 | 463 | 226 | 11462 | 6676 | 4785 | 8969658 | 900 | 761 | | |
| 296 | 295 | GS글로벌 | A001250 | 28150 | 322 | -821 | 9568 | 6578 | 2990 | 82533764 | | 0 | | |
| 297 | 296 | 코오롱플라스트 | A138490 | 2951 | 38 | 29 | 3179 | 1196 | 1982 | 38000000 | | 0 | | |
| 298 | 297 | 동일산업 | A004890 | 3332 | 19 | 80 | 4175 | 570 | 3604 | 2425215 | 1,800 | 1,523 | | |
| 299 | 298 | 성신양회 | A004980 | 7223 | 211 | 101 | 10902 | 6828 | 4074 | 24516073 | 150 | 127 | | |
| 300 | 299 | 대영포장 | A014160 | 2607 | 47 | 33 | 2079 | 540 | 1538 | 108394549 | | 0 | | |
| 301 | 300 | 한즈코퍼레이션 | A143210 | 5726 | -52 | -81 | 8441 | 5293 | 3147 | 22263268 | | 0 | | |
| 302 | | | | | | | | | | | | | | |
| | | 2017실적 | 2018실적 | 2019실적 | 2020실적 | + | | | | | | | | |

사용할 종목들의 일자별 주가 조사 후 DB에 저장

[12003] 개별종목 시세 추이

통계 > 기본 통계 > 주식 > 종목시세 > 개별종목 시세 추이

종목명

000150/두산

조회

조회기간

20000424

~

20210423

1일

1개월

6개월

1년

Close x

2021.05.29 PM 03:30:47 (20분 지연 정보)

주

원

Y

| 일자 | 증가 | 대비 | 등락률 | 시가 | 고가 | 저가 | 거래량 | 거래대금 | 시가총액 | 상장주식수 |
|------------|--------|---------|-------|--------|--------|--------|---------|----------------|-----------------|------------|
| 2021/04/23 | 53,400 | ▼ 700 | -1.29 | 53,700 | 54,100 | 53,300 | 53,141 | 2,849,299,4... | 882,372,789,000 | 16,523,835 |
| 2021/04/22 | 54,100 | ▲ 1,000 | +1.88 | 53,500 | 54,300 | 52,800 | 112,744 | 6,065,952,7... | 893,939,473,500 | 16,523,835 |
| 2021/04/21 | 53,100 | ▼ 300 | -0.56 | 53,200 | 53,300 | 52,200 | 82,447 | 4,350,998,3... | 877,415,638,500 | 16,523,835 |
| 2021/04/20 | 53,400 | ▲ 300 | +0.56 | 52,500 | 54,000 | 52,400 | 107,642 | 5,729,616,6... | 882,372,789,000 | 16,523,835 |

제안 방법

주가 예측 모델에 사용할 종목의 일자별 주가 획득 및 가공

[12003] 개별종목 시세 추이

통계 > 기본 통계 > 주식 > 종목시세 > 개별종목 시세 추이

종목명

032940/일익

조회

조회기간

20000424 ~ 20210423

1일1개월6개월1년

Close

2021.05.29 PM 03:18:49 (20분 지연 정보)

주원

| 일자 | 종가 | 대비 | 등락률 | 시가 | 고가 | 저가 | 거래량 | 거래대금 | 시가총액 | 상장주식수 |
|------------|-------|-------|-------|-------|-------|-------|---------|----------------|-----------------|------------|
| 2021/04/23 | 5,450 | ▲ 60 | +1.11 | 5,340 | 5,470 | 5,300 | 106,109 | 572,336,300 | 99,153,103,500 | 18,193,230 |
| 2021/04/22 | 5,390 | ▼ 60 | -1.10 | 5,450 | 5,500 | 5,310 | 111,331 | 601,259,850 | 98,061,509,700 | 18,193,230 |
| 2021/04/21 | 5,450 | ▲ 70 | +1.30 | 5,390 | 5,610 | 5,390 | 175,252 | 963,379,380 | 99,153,103,500 | 18,193,230 |
| 2021/04/20 | 5,380 | ▲ 60 | +1.13 | 5,380 | 5,470 | 5,290 | 144,514 | 778,513,130 | 97,879,577,400 | 18,193,230 |
| 2021/04/19 | 5,320 | ▲ 10 | +0.19 | 5,390 | 5,430 | 5,240 | 136,198 | 728,372,540 | 96,787,983,600 | 18,193,230 |
| 2021/04/16 | 5,310 | ▲ 100 | +1.92 | 5,210 | 5,340 | 5,210 | 76,298 | 403,569,710 | 96,606,051,300 | 18,193,230 |
| 2021/04/15 | 5,210 | ▼ 40 | -0.76 | 5,290 | 5,290 | 5,200 | 90,565 | 473,356,570 | 94,786,728,300 | 18,193,230 |
| 2021/04/14 | 5,250 | ▼ 80 | -1.50 | 5,370 | 5,370 | 5,200 | 120,865 | 635,861,810 | 95,514,457,500 | 18,193,230 |
| 2021/04/13 | 5,330 | ▼ 30 | -0.56 | 5,380 | 5,500 | 5,290 | 174,098 | 934,242,050 | 96,969,915,900 | 18,193,230 |
| 2021/04/12 | 5,360 | ▲ 10 | +0.19 | 5,420 | 5,430 | 5,320 | 66,383 | 355,211,670 | 97,515,712,800 | 18,193,230 |
| 2021/04/09 | 5,350 | ▲ 40 | +0.75 | 5,300 | 5,440 | 5,290 | 100,878 | 538,560,270 | 97,333,780,500 | 18,193,230 |
| 2021/04/08 | 5,310 | ▼ 130 | -2.39 | 5,430 | 5,460 | 5,270 | 100,148 | 533,538,740 | 96,606,051,300 | 18,193,230 |
| 2021/04/07 | 5,440 | ▼ 40 | -0.73 | 5,480 | 5,580 | 5,380 | 108,952 | 596,708,250 | 98,971,171,200 | 18,193,230 |
| 2021/04/06 | 5,480 | ▼ 290 | -5.03 | 5,850 | 5,860 | 5,450 | 210,781 | 1,170,366,190 | 99,698,900,400 | 18,193,230 |
| 2021/04/05 | 5,770 | ▲ 340 | +6.26 | 5,490 | 5,790 | 5,440 | 281,561 | 1,592,029,9... | 104,974,937,100 | 18,193,230 |
| 2021/04/02 | 5,430 | 0 | 0.00 | 5,600 | 5,600 | 5,370 | 109,589 | 595,856,120 | 98,789,238,900 | 18,193,230 |
| 2021/04/01 | 5,430 | ▲ 30 | +0.56 | 5,470 | 5,470 | 5,330 | 79,905 | 430,466,110 | 98,789,238,900 | 18,193,230 |
| 2021/03/31 | 5,400 | ▼ 10 | -0.18 | 5,600 | 5,600 | 5,310 | 93,638 | 504,254,630 | 98,243,442,000 | 18,193,230 |
| 2021/03/30 | 5,410 | ▲ 130 | +2.46 | 5,300 | 5,450 | 5,300 | 163,021 | 877,438,060 | 98,425,374,300 | 18,193,230 |
| 2021/03/29 | 5,280 | ▲ 150 | +2.92 | 5,340 | 5,430 | 5,170 | 225,588 | 1,204,066,8... | 96,060,254,400 | 18,193,230 |



| | A | B | C | D | E | F |
|------|-----------|-------|------|------|------|--------|
| 1 | date | close | open | high | low | vol |
| 2 | 2000.4.24 | 5940 | 6070 | 6100 | 5800 | 409179 |
| 3 | 2000.4.25 | 5870 | 5700 | 6000 | 5610 | 249685 |
| 4 | 2000.4.26 | 5800 | 6150 | 6260 | 5650 | 391259 |
| 5 | 2000.4.27 | 5600 | 5650 | 5900 | 5400 | 205318 |
| 6 | 2000.4.28 | 5660 | 5700 | 5730 | 5570 | 224748 |
| 7 | 2000.5.2 | 5950 | 5660 | 6000 | 5660 | 320955 |
| 8 | 2000.5.3 | 6020 | 5850 | 6370 | 5700 | 743882 |
| 9 | 2000.5.4 | 6120 | 6100 | 6250 | 5960 | 430641 |
| 10 | 2000.5.8 | 6210 | 6230 | 6620 | 6110 | 487048 |
| 11 | 2000.5.9 | 6130 | 6090 | 6180 | 5960 | 257391 |
| 5178 | 2021.4.12 | 5360 | 5420 | 5430 | 5320 | 66383 |
| 5179 | 2021.4.13 | 5330 | 5380 | 5500 | 5290 | 174098 |
| 5180 | 2021.4.14 | 5250 | 5370 | 5370 | 5200 | 120865 |
| 5181 | 2021.4.15 | 5210 | 5290 | 5290 | 5200 | 90565 |
| 5182 | 2021.4.16 | 5310 | 5210 | 5340 | 5210 | 76298 |
| 5183 | 2021.4.19 | 5320 | 5390 | 5430 | 5240 | 136198 |
| 5184 | 2021.4.20 | 5380 | 5380 | 5470 | 5290 | 144514 |
| 5185 | 2021.4.21 | 5450 | 5390 | 5610 | 5390 | 175252 |
| 5186 | 2021.4.22 | 5390 | 5450 | 5500 | 5310 | 111331 |
| 5187 | 2021.4.23 | 5450 | 5340 | 5470 | 5300 | 106109 |

1. csv파일로 다운로드
2. 불필요한 칼럼 삭제
3. 칼럼명 영어로 변경
4. 날짜 오름차순으로 정렬
- (5. 액면 분할(병합)이 있는 경우 비율에 맞게 값 변환)

제안 방법

- 재무 정보를 바탕으로 설정한 조건에 맞는 종목을
선별해주는 프로그램
(Python 프로그램)
- 재무 정보를 바탕으로 선별한 종목들을
매수할지 판단해주는 모델
(LSTM 모델)

재무 정보를 바탕으로 설정한 조건에 맞는 종목을 선별해주는 프로그램

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|----|-----------|-----------|-----|-----|------------|-----|------------|-------------|-----------|--------------|---------------|--------------|---------|-----------|------------|
| | StockCode | StockName | EqP | PER | M | PBR | ROE | BuyingPrice | SellPrice | BuyingAmount | ProfitLossByS | InvestmentAr | Balance | Yield | |
| 0 | 84110 | 휴온스글로벌 | # | # | 0.48942597 | # | 0.56739481 | 115.93067 | 28050 | 76800 | 356 | 48750 | 9985800 | 27340800 | 173.796791 |
| 1 | 32940 | 원익 | # | # | 0.85708577 | # | 0.57081895 | 66.5999803 | 5570 | 4530 | 1795 | -1040 | 9998150 | 8131350 | -18.671454 |
| 2 | 7860 | 서연 | # | # | 1.78272054 | # | 0.22011737 | 12.347273 | 10150 | 6100 | 985 | -4050 | 9997750 | 6008500 | -39.901478 |
| 3 | 30530 | 원익홀딩스 | # | # | 2.35249679 | # | 0.64821874 | 27.5545004 | 6090 | 7390 | 1642 | 1300 | 9999780 | 12134380 | 21.3464696 |
| 4 | 23600 | 삼보판지 | # | # | 2.3803327 | # | 0.2906615 | 12.2109611 | 6860 | 9790 | 1457 | 2930 | 9995020 | 14264030 | 42.7113703 |
| 5 | 33160 | 엠케이전자 | # | # | 2.39128164 | # | 0.30002216 | 12.5465004 | 8360 | 10200 | 1196 | 1840 | 9998560 | 12199200 | 22.0095694 |
| 6 | 81660 | 휠라홀딩스 | # | # | 2.80111661 | # | 0.66927183 | 23.8930371 | 70400 | 126000 | 142 | 55600 | 9996800 | 17892000 | 78.9772727 |
| 7 | 37460 | 삼지전자 | # | # | 2.91807298 | # | 0.52772227 | 18.0846153 | 8130 | 13000 | 1230 | 4870 | 9999900 | 15990000 | 59.901599 |
| 8 | 5710 | 대원산업 | # | # | 2.97846631 | # | 0.57581982 | 19.3327624 | 7150 | 7300 | 1398 | 150 | 9995700 | 10205400 | 2.0979021 |
| 9 | 122450 | KMH | # | # | 2.98471715 | # | 0.38434863 | 12.8772213 | 7950 | 11700 | 1257 | 3750 | 9993150 | 14706900 | 47.1698113 |
| 10 | | | | | | | | | | | | | | 138872560 | 38.91195 |

실행 결과로 나오는 csv파일

백테스팅 검증

재무 정보 날짜: 2016년 4분기

종목들 매수 날짜: 2017.04.10

종목들 매도 날짜: 2018.04.16

조건: **PER > 0, PBR < 0.7, ROE > 11**, PER 오름차순 10개

총 수익률: 38.9 %

재무 정보를 바탕으로 선별한 종목들의 주가를 예측하는 모델

```
Model: "sequential"
```

| Layer (type) | Output Shape | Param # |
|---------------|--------------|---------|
| lstm (LSTM) | (None, 120) | 58560 |
| dense (Dense) | (None, 60) | 7260 |

```
Total params: 65,820  
Trainable params: 65,820  
Non-trainable params: 0
```

입력: 120개의 셀을 가진 LSTM layer

출력: 60개 값을 가진 Dense layer

120일의 주가를 바탕으로 이후 60일의 주가 예측

60일을 예측한 이유

PER, PBR, ROE 등의 값은 분기마다 발표하는 보고서에 따라 변한다.

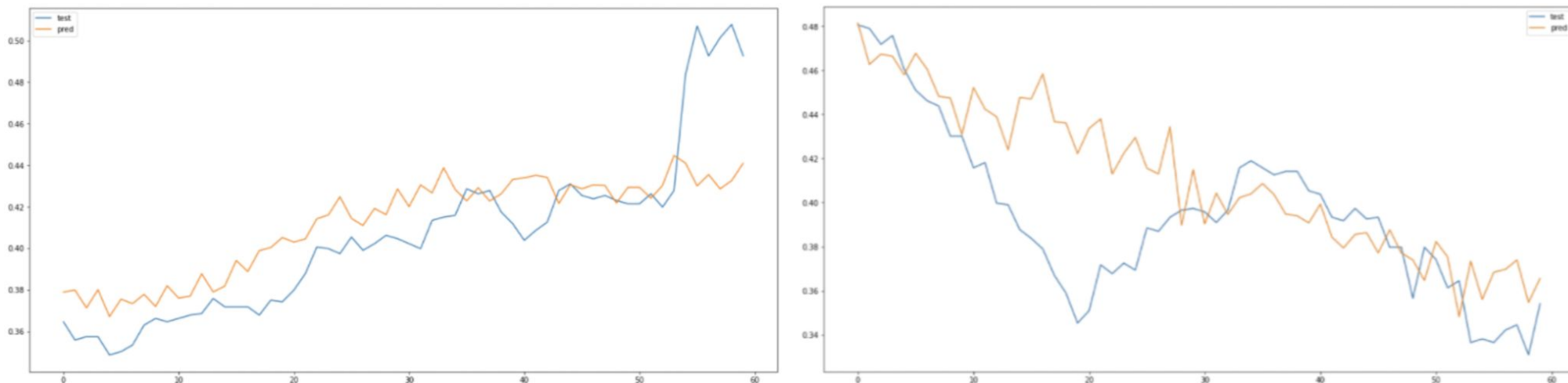
실제로 어떤 분기의 지표를 기준으로 분기, 반기, 연간 투자를 했을 경우 분기 투자가 가장 수익률이 높다.

보통 한 분기는 60일을 조금 넘는다.

따라서 한 분기인 60일을 예측하여, 해당 종목의 주가가 한 분기동안 상승할지 하락할지 예측한다.

예측 결과가 상승이면 매수하고, 하락이면 매수하지 않는다.

재무 정보를 바탕으로 선별한 종목들의 주가를 예측하는 모델



모델의 출력인 60일간 예측 주가(orange)와 같은 기간의 실제 주가 (blue)

```
total: 59.32 %  
up:    51.26 %   265 / 517  
down:  65.06 %   473 / 727
```

```
total: 63.2 %  
up:    83.1 %   354 / 426  
down:  52.77 %  429 / 813
```

```
total: 64.63 %  
up:    75.73 %  440 / 581  
down:  54.9 %   364 / 663
```

여러 종목으로 모델 평가

정확도: 60% 내외

정확도 판단 기준

상승(또는 하락)할지 예측한 것이 실제로 상승(또는 하락)했을 경우 True
전체 Test set에 대하여 True의 비율

기대효과

- 투자 종목을 선별할 때 PER, PBR, ROE 등의 지표를 원하는 조건으로 검색 가능
- 선별한 종목을 매수할지 판단
- 켈리 공식에 근거하여 60% 내외의 정확도면 분산 투자를 하였을 경우, 장기적으로 수익을 거둘 수 있다.

향후 연구

- 한 분기의 재무 정보가 아닌 여러 기간의 재무 정보로 종목을 선별하면 수익률 상승
- 주가 및 재무 정보 뿐만 아니라 뉴스의 키워드도 예측에 사용하면 정확도가 상승할 것이라 판단

감사합니다.