

해커톤 프로젝트

: 머신러닝에 대한 이해 및 주식 가격 예측을 통한 실습

7조 칠성사이다

[CONTENTS]

1. 중간발표 Review

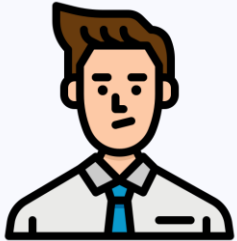
팀원 소개 | 분석 주제

2. 분석 내용 및 결과

분석 내용 | ARIMA모형 분석 결과

3. 한계점 및 느낀 점

한계점 및 느낀 점 | Q&A



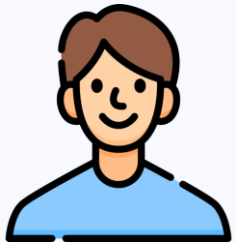
정해랑©

- 조장
- 모델결과 분석 및 해석
- 웹 구현
- 해커톤 일지 작성



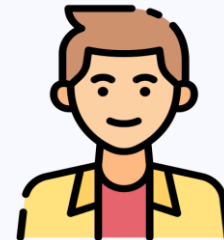
김영주

- 머신러닝 모델 조사 및 정리
- 모델결과 분석 및 해석
- 모델 설계 보조
- 중간 발표



이정환

- 머신러닝 모델 조사
- 분석 모델 설계
- 모델 결과 분석 및 해석
- 데이터 수집



김준호

- 머신러닝 모델 조사
- 모델결과 분석 및 해석
- 모델 설계 보조
- 최종 발표 및 PPT제작

머신러닝에 대한 이해 & 주가 예측을 통한 실습



머신러닝 모델 이해

: 수업시간에 배운 머신러닝 모델 이해



아리마모형을 통한 주가 예측

: 분석 툴, 아리마(ARIMA)모형 채택



웹 구현

: 분석내용 웹에다 구현



세부 데이터 설정

: 항공 테마주를 분석 데이터로 삼고 분석 진행



왜 항공주?	<div>팀원들의 관심분야</div> <div>· 팀원 4명 모두 여행을 좋아함</div>
주가 상승이 기대되는 산업	<div>거리두기 완화 & 입국방역절차 간소화 등 항공 업계 회복 기대감</div> <div>여행 수요 살아나자 미국 항공주 급등</div> <div>...</div> <div>미국 항공사들 높은 여행 수요 확인 & 호실적</div> <div>(출처: 아주경제 22.04)</div>
데이터 내용	<div>분석대상 기업의 분당 주가 & 거래량 데이터</div> <div>기간 : 21.04.01 ~ 21.04.26 (약 1년)</div>
분석대상기업	AK홀딩스 / 대한항공 / 아시아나항공 / 에어부산 / 제주항공 / 진에어 / 티웨이항공
Train Set	21.04.01 ~ 21.12.31 기간의 주가 & 거래량
Test Set	22.01.03 ~ 22.04.26 기간의 주가 & 거래량

분석 방향

: 항공 테마주의 공통된 최적의 p, d, q 도출 후 모델 성능 평가

STEP 01

같은 테마안에 있는
개별 주식은 공통적인
추세와 주기를 띤다고
생각

STEP 02

Auto Arima 이용
항공 테마주
최적의 p, d, q 도출

STEP 03

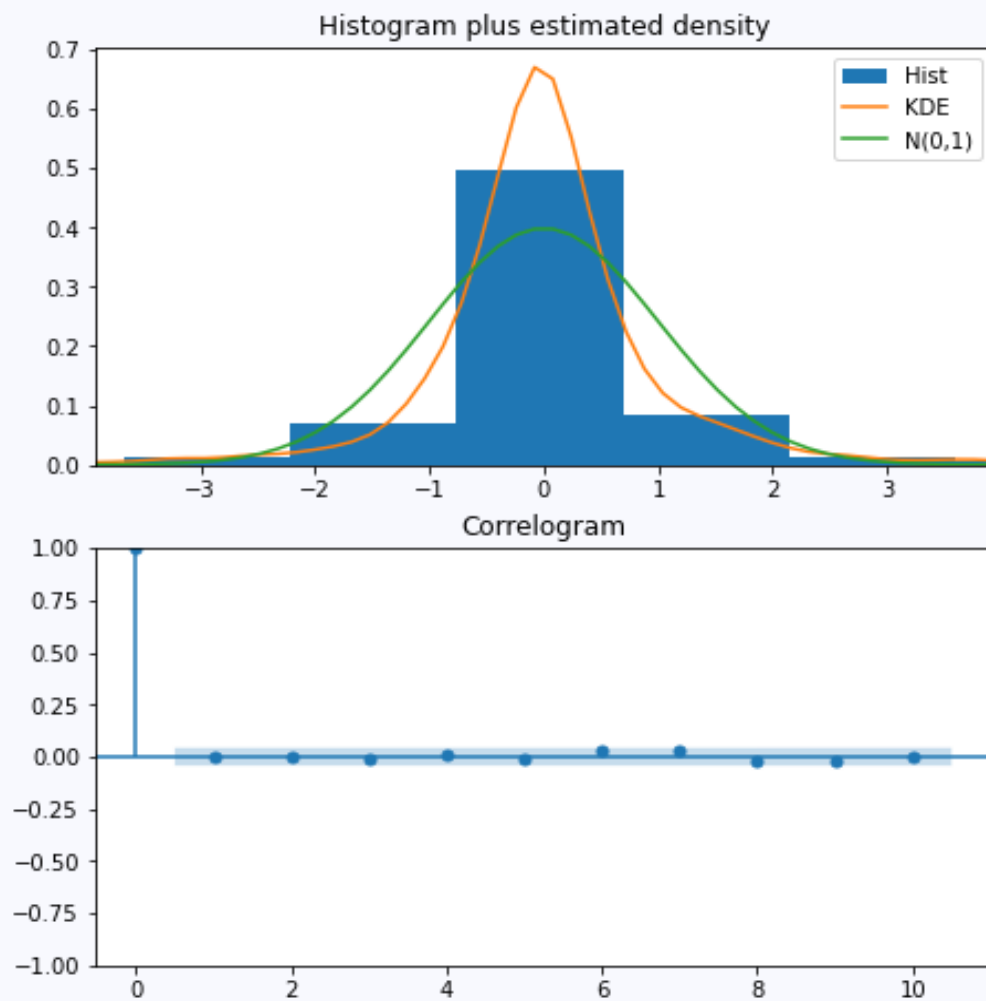
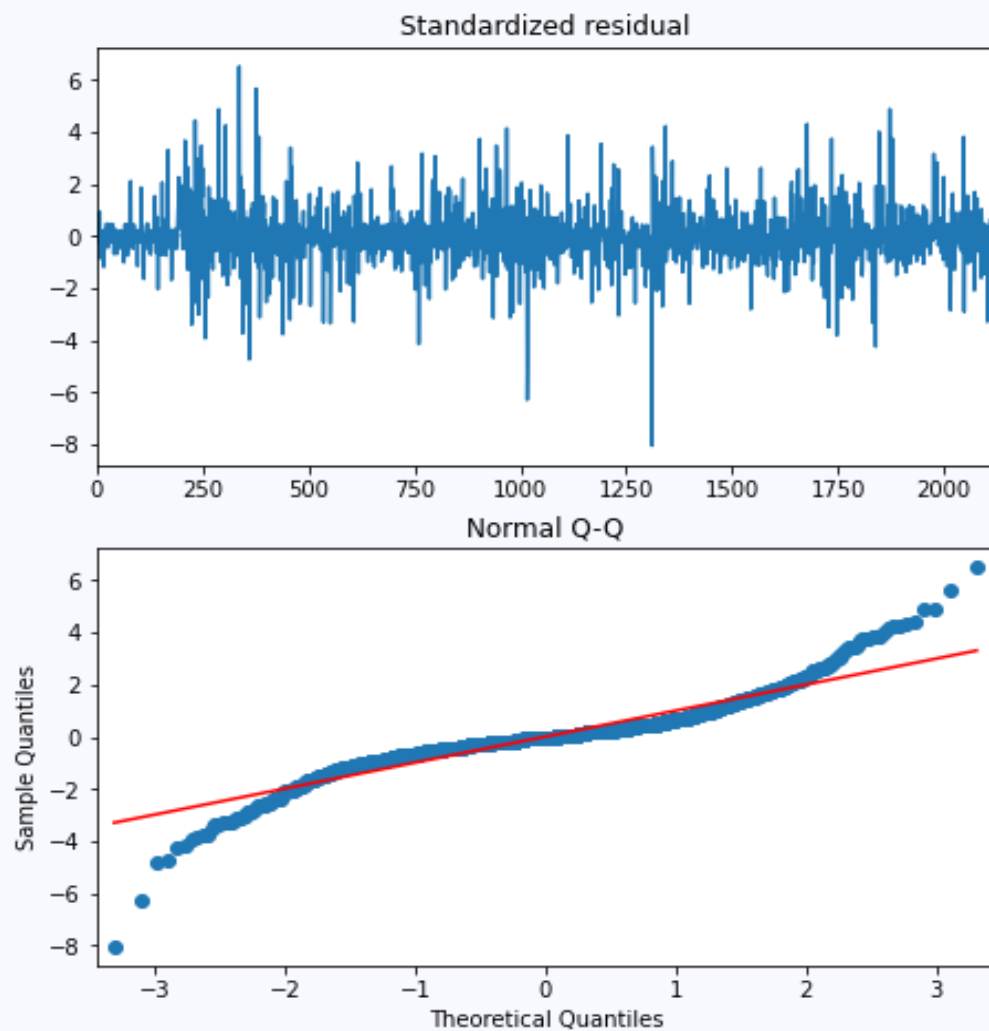
최적의 p, d, q 를
적용하여 예측.
모델 성능 평가.

[CONTENTS]

2. 분석 내용 및 결과

분석 내용 | ARIMA모형 분석 결과

각 기업 분석마다 데이터 확인과정 이후 분석진행



2. 분석내용 및 결과

분석 내용 | ARIMA모형 분석 결과

p,d,q 도출

: Auto-Arima tool을 이용하여 최적의 p,d,q 도출

```
columns = ["AK_close", "korea_air_close", "aisana_close",
           "airbusan_close", "jeju_close", "jin_close", "tway_close"]

title = ["AK 홀딩스(006840)", "대한항공(003490)", "아시아나항공(020560)",
         "에어부산(298690)", "제주항공(089090)", "진에어(272450)", "티웨이항공(091810)"]

for i in range(0,7,1):
    # print("="*60)
    # print(i)
    a = title[i]

    y_train = air_hour_1[columns[i]].iloc[:188] #기준행 188
    y_test = air_hour_1[columns[i]].iloc[188:]
    y_train.plot()
    y_test.plot()

    #차분 차수 찾기
    kpss_diffs = ndiffs(y_train, alpha=0.05, test='kpss', max_d = 6)
    adf_diffs = ndiffs(y_train, alpha=0.05, test='adf', max_d=6)
    n_diffs = max(adf_diffs, kpss_diffs)

    n_diffs_print = f'적정 차분 차수는 {n_diffs}'

    print(n_diffs_print)
    # print_data = []
    # print_data.append(n_diffs_print)

    model = pm.auto_arima(y=y_train,
                          d = 1,
                          start_p=0,
                          max_p=3,
                          start_q=0,
                          max_q=3,
                          seasonal=False,
                          stepwise=True,
                          trace = True)

    #append위한 함수
    def model2():
        model = pm.auto_arima(y=y_train,
                              d = 1,
                              start_p=0,
                              max_p=3,
                              start_q=0,
                              max_q=3,
                              seasonal=False,
                              stepwise=True,
                              trace = True)

    return model
```

시간별

- AK홀딩스 = (3,1,1)
- 대한항공 = (0,1,2)
- 아시아나 = (0,1,0)
- 에어부산 = (3,1,3)
- 제주항공 = (2,1,3)
- 진에어 = (0,1,0)
- 티웨이 = (0,1,0)

일별

- AK홀딩스 = (0,1,0)
- 대한항공 = (0,1,0)
- 아시아나 = (1,1,0)
- 에어부산 = (0,1,0)
- 제주항공 = (0,1,0)
- 진에어 = (0,1,0)
- 티웨이 = (0,1,0)

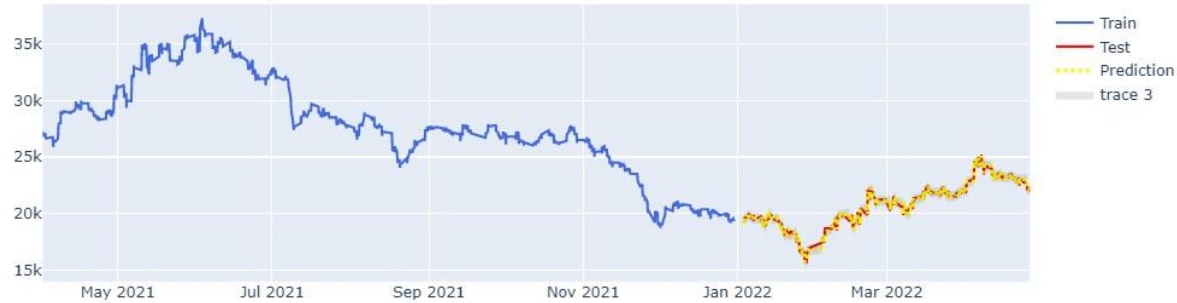
예측 평가 지표

: MAPE와 괴리율 지표 사용

<div>MAPE</div> <div>(평균 절대 비율 오차)</div> <div>(Mean Absolute Percentage Error)</div>	측정 지표	<div>. MAPE는 실제값과 예측값 사이의 차이를 실제값으로 나눠준 값들의 평균 x 100</div> <div>즉, 오차가 실제값에서 차지하는 상대적인 비율</div>
	수식	<div>$MAPE = \frac{100}{n} \sum_{i=1}^n \left \frac{y_i - \hat{f}(x_i)}{y_i} \right$</div>
<div>괴리율</div>	측정 지표	<div>. 애널리스트 리포트 내 괴리율 지표는 예측정확도 측정지표로 활용</div> <div>. 애널리스트가 설정한 목표주가와 실제 주가와의 차이를 확인하는 지표</div>
	수식	<div>$\frac{(\text{예측 주가} - \text{실제 주가})}{(\text{실제 주가})} \times 100$</div>

시간별 데이터로 ARIMA모형 재분석

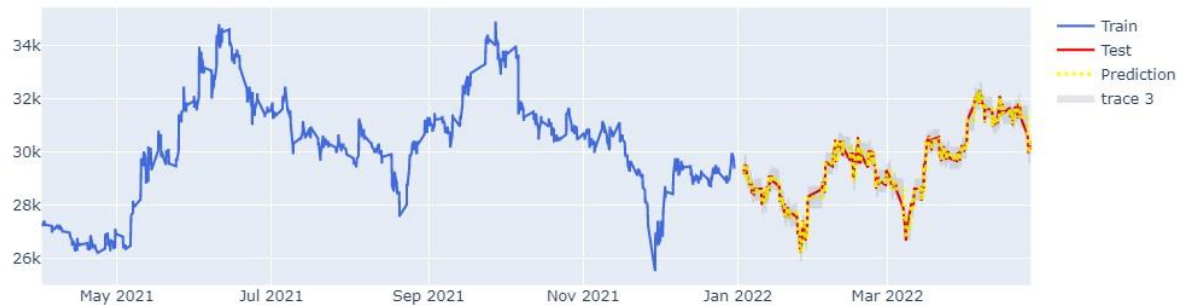
AK 홀딩스(006840)



[AK홀딩스 (006840)]

- MAPE = 0.73%
- 괴리율 = -0.014%

대한항공(003490)



[대한항공 (003490)]

- MAPE = 0.44%
- 괴리율 = -0.03%

시간별 데이터로 ARIMA모형 재분석

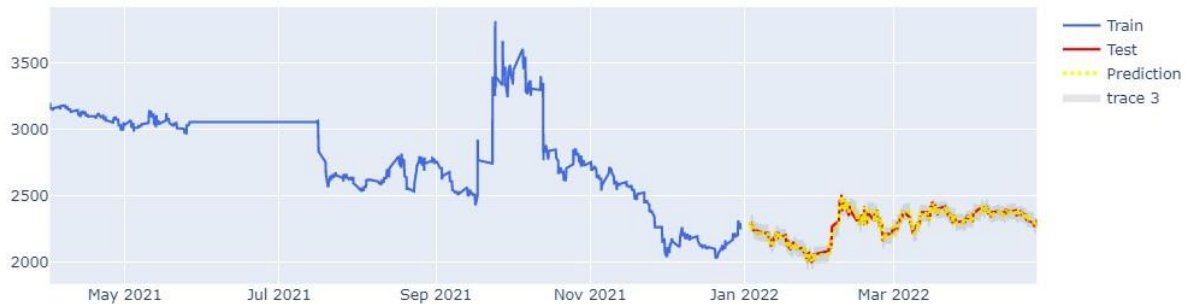
아시아나항공(020560)



[아시아나 항공 (020560)]

- MAPE = 0.69%
- 괴리율 = -0.000..%

에어부산(298690)

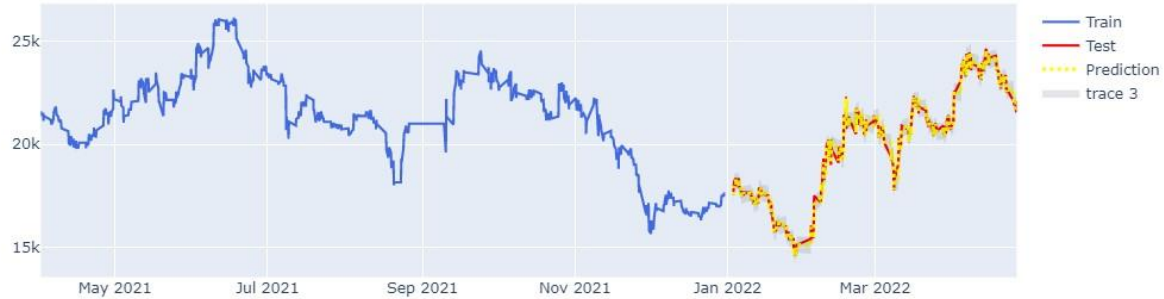


[에어부산 (298690)]

- MAPE = 0.60%
- 괴리율 = 0.001%

시간별 데이터로 ARIMA모형 재분석

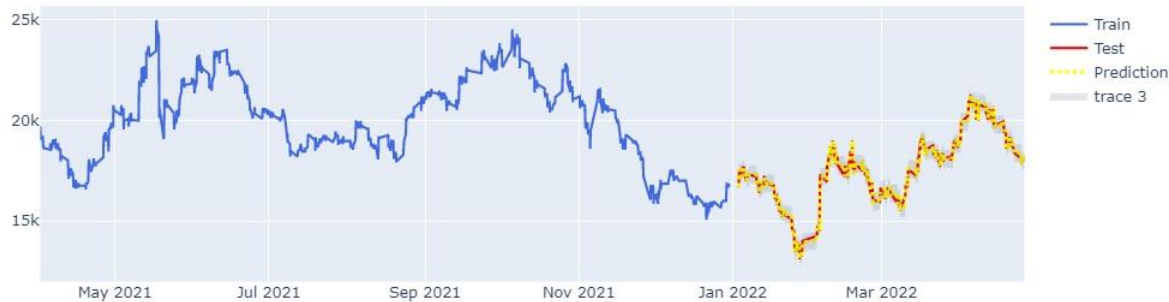
제주항공(089090)



[제주항공 (089090)]

- MAPE = 0.86%
- 괴리율 = -0.028%

진에어(272450)



[진에어 (272450)]

- MAPE = 0.84%
- 괴리율 = -0.005%

시간별 데이터로 ARIMA모형 재분석

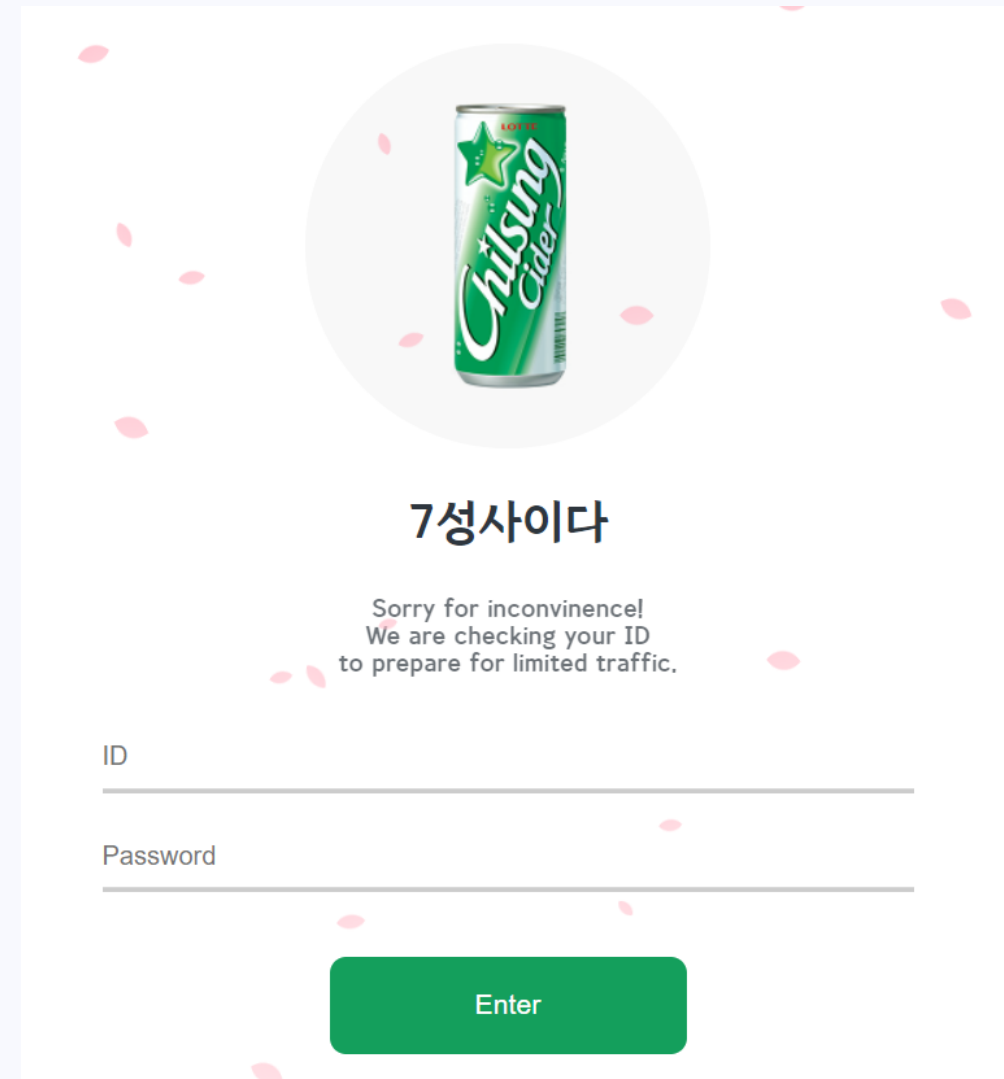


[티웨이항공 (089090)]

- MAPE = 0.68%
- 괴리율 = -0.011%

접속해보세요😊

ID & PW = asd



Chilsung Cider

7성사이다

Sorry for inconvenience!
We are checking your ID
to prepare for limited traffic.

ID

Password

Enter

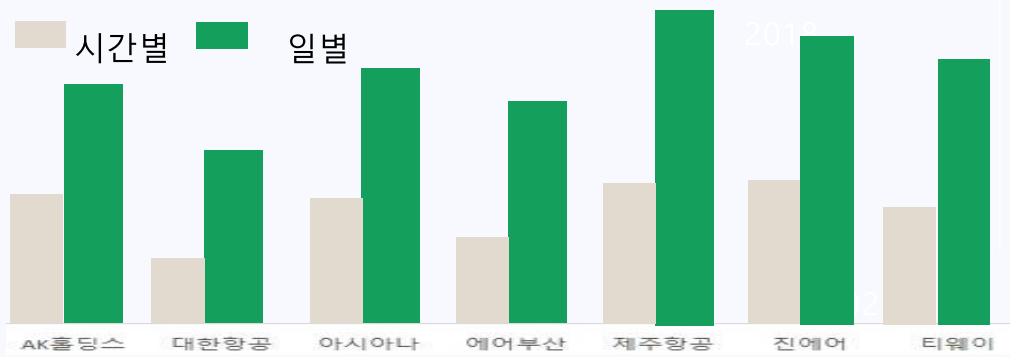
2. 분석내용 및 결과

분석 내용 | ARIMA모형 분석 결과

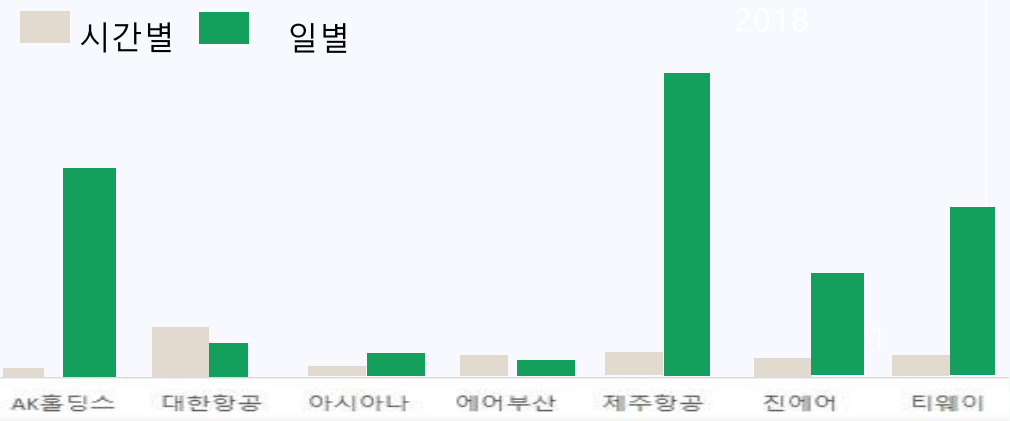
도출된 시간별, 일별 p,d,q 로 예측 후 평가

MAPE]

MAPE(%)



과리율(%)



▶ [MAPE]

일별 데이터보다 시간별 데이터를 사용한 ARIMA모형이 더 정확

▶ [과리율]

일별 데이터나 시간별 데이터를 사용한 ARIMA모형의 정확도가 개별 기업마다 다르게 나오는 모습을 확인.

	ARIMA(p,d,q)		MAPE(%)		과리율(%)	
	시간별	일별	시간별	일별	시간별	일별
AK홀딩스	(3,1,1)	(0,1,0)	0.730%	2.030%	-0.014%	-0.138%
대한항공	(0,1,2)	(0,1,0)	0.440%	1.320%	-0.030%	-0.025%
아시아나	(0,1,0)	(1,1,0)	0.690%	2.070%	0.000%	-0.004%
에어부산	(3,1,3)	(0,1,0)	0.600%	1.660%	0.001%	0.000%
제주항공	(2,1,3)	(0,1,0)	0.860%	2.450%	-0.028%	-0.216%
진에어	(0,1,0)	(0,1,0)	0.840%	2.350%	-0.005%	-0.050%
티웨이	(0,1,0)	(0,1,0)	0.680%	2.090%	-0.011%	-0.102%

2. 분석내용 및 결과

분석 내용 | ARIMA모형 분석 결과

Auto Arima로 도출한 p,d,q 와 다르게 주었을 때 평가지표를 비교

	ARIMA(p,d,q)				ARIMA(p,d,q)				ARIMA(p,d,q)				ARIMA(p,d,q)		
	시간별	시간별	시간별		시간별	시간별	시간별		시간별	시간별	시간별		시간별	시간별	시간별
AK홀딩스	(0,1,0)	0.730%	-0.034%	대한항공	(0,1,0)	0.450%	0.001%	아시아나	(0,1,0)	0.690%	0.009%	에어부산	(0,1,0)	0.580%	-0.031%
	(0,1,1)	0.730%	-0.037%		(0,1,1)	0.450%	0.001%		(0,1,1)	0.690%	0.010%		(0,1,1)	0.570%	-0.022%
	(0,1,2)	0.730%	-0.039%		(0,1,2)	0.450%	0.003%		(0,1,2)	0.700%	0.011%		(0,1,2)	0.580%	-0.026%
	(0,1,3)	0.730%	-0.039%		(0,1,3)	0.450%	0.002%		(0,1,3)	0.700%	0.012%		(0,1,3)	0.580%	-0.026%
	(1,1,0)	0.730%	-0.037%		(1,1,0)	0.450%	0.001%		(1,1,0)	0.690%	0.009%		(1,1,0)	0.580%	-0.022%
	(1,1,1)	0.730%	-0.040%		(1,1,1)	0.450%	0.002%		(1,1,1)	0.690%	0.011%		(1,1,1)	0.570%	-0.025%
	(1,1,2)	0.730%	-0.042%		(1,1,2)	0.450%	0.001%		(1,1,2)	0.700%	0.009%		(1,1,2)	0.580%	-0.026%
	(1,1,3)	0.730%	-0.032%		(1,1,3)	0.450%	0.002%		(1,1,3)	0.700%	0.009%		(1,1,3)	0.580%	-0.027%
	(2,1,0)	0.730%	-0.042%		(2,1,0)	0.450%	0.001%		(2,1,0)	0.700%	0.010%		(2,1,0)	0.580%	-0.025%
	(2,1,1)	0.730%	-0.041%		(2,1,1)	0.450%	0.002%		(2,1,1)	0.700%	0.011%		(2,1,1)	0.580%	-0.026%
	(2,1,2)	0.730%	-0.033%		(2,1,2)	0.450%	0.003%		(2,1,2)	0.700%	0.011%		(2,1,2)	0.600%	-0.025%
	(2,1,3)	0.730%	-0.034%		(2,1,3)	0.450%	0.001%		(2,1,3)	0.710%	0.008%		(2,1,3)	0.600%	-0.028%
	(3,1,0)	0.730%	-0.040%		(3,1,0)	0.450%	0.002%		(3,1,0)	0.700%	0.012%		(3,1,0)	0.580%	-0.027%
최저값	(3,1,1)	0.730%	-0.032%	최저값	(3,1,1)	0.450%	0.001%	최저값	(3,1,1)	0.700%	0.010%	최저값	(3,1,1)	0.580%	-0.027%
	(3,1,2)	0.730%	-0.035%		(3,1,2)	0.450%	0.001%		(3,1,2)	0.710%	0.015%		(3,1,2)	0.600%	-0.024%
	(3,1,3)	0.730%	-0.034%		(3,1,3)	0.450%	0.001%		(3,1,3)	0.710%	0.016%		(3,1,3)	0.600%	-0.029%
		0.730%	-0.034%			0.450%	0.001%			0.690%	0.009%			0.570%	-0.022%
제주항공	(0,1,0)	0.840%	-0.012%	진에어	(0,1,0)	0.840%	-0.012%	티웨이	(0,1,0)	0.690%	-0.001%				
	(0,1,1)	0.850%	-0.034%		(0,1,1)	0.830%	-0.013%		(0,1,1)	0.690%	-0.001%				
	(0,1,2)	0.850%	-0.036%		(0,1,2)	0.830%	-0.014%		(0,1,2)	0.690%	-0.001%				
	(0,1,3)	0.850%	-0.042%		(0,1,3)	0.830%	-0.015%		(0,1,3)	0.690%	-0.001%				
	(1,1,0)	0.850%	-0.034%		(1,1,0)	0.830%	-0.013%		(1,1,0)	0.690%	-0.001%				
	(1,1,1)	0.850%	-0.035%		(1,1,1)	0.830%	-0.014%		(1,1,1)	0.690%	-0.001%				
	(1,1,2)	0.850%	-0.042%		(1,1,2)	0.830%	-0.013%		(1,1,2)	0.690%	0.002%				
	(1,1,3)	0.850%	-0.044%		(1,1,3)	0.840%	-0.013%		(1,1,3)	0.690%	-0.001%				
	(2,1,0)	0.850%	-0.038%		(2,1,0)	0.830%	-0.014%		(2,1,0)	0.690%	-0.001%				
	(2,1,1)	0.850%	-0.041%		(2,1,1)	0.830%	-0.014%		(2,1,1)	0.690%	-0.001%				
	(2,1,2)	0.860%	-0.036%		(2,1,2)	0.840%	-0.017%		(2,1,2)	0.690%	-0.004%				
	(2,1,3)	0.860%	-0.039%		(2,1,3)	0.840%	-0.017%		(2,1,3)	0.690%	-0.004%				
	(3,1,0)	0.850%	-0.042%		(3,1,0)	0.830%	-0.015%		(3,1,0)	0.690%	-0.001%				
최저값	(3,1,1)	0.850%	-0.014%	최저값	(3,1,1)	0.840%	-0.014%	최저값	(3,1,1)	0.690%	-0.001%				
	(3,1,2)	0.860%	-0.040%		(3,1,2)	0.840%	-0.017%		(3,1,2)	0.690%	-0.002%				
	(3,1,3)	0.860%	-0.045%		(3,1,3)	0.840%	-0.012%		(3,1,3)	0.690%	0.000_				
		0.840%	-0.012%			0.830%	-0.012%			0.690%	0.000_				

Auto Arima로 도출한 최적의 p,d,q 적용한 값

Auto Arima로 도출한
최적의 p,d,q 를 적용했을 때,
평가 지표가
항상 가장 좋게 기록된 것은 아님을
비교 확인.

시간별 데이터로 도출된 p,d,q 로 일별 데이터 ARIMA 분석

	ARIMA(p,d,q)	MAPE(%)		과리율(%)	
	시간별	일별	시간→일	일별	시간→일
AK홀딩스	(3,1,1)	2.030%	2.050%	-0.138%	-0.274%
대한항공	(0,1,2)	1.320%	1.320%	-0.025%	0.015%
아시아나	(0,1,0)	2.070%	2.090%	-0.004%	0.091%
에어부산	(3,1,3)	1.660%	1.800%	0.000%	-0.018%
제주항공	(2,1,3)	2.450%	2.500%	-0.216%	-0.259%
진에어	(0,1,0)	2.350%	2.350%	-0.050%	-0.094%
티웨이	(0,1,0)	2.090%	2.100%	-0.102%	-0.020%



[시간별 데이터의 추출된 (p,d,q)]로
일별 데이터 ARIMA 분석할 시,
평가지표의 변화를 확인



같은 기업의 데이터라도 일별이냐 시간별이냐에
따라서 결과가 다르게 나타날 뿐더러,
 p,d,q 설정에 따라서도 평가지표 값이 달라짐.

[CONTENTS]

3. 한계점 및 느낀 점

한계점 및 느낀 점 Q&A

[한계점 및 느낀 점]

- 모델의 성능 지표로 우수 모델?
분석이 동일한 조건으로 시행되지 않았음.
- p, d, q 가 동일할 것이라는 아쉬운 가정
개별주식에만 적용되는 이슈가 작용
- 통계적 & 이론적으로 우수한 모델로도
실제 주가 예측은 어렵



모델 성능
의문점



아쉬운 가정
이해 미숙



주가 예측
어려움



Q&A

THANK YOU