
2021 디지털 산업혁신 빅데이터 플랫폼 활용
금융 아이디어 경진대회

재무제표를 이용한 종목 추천 서비스 개발

INDEX

01 제안배경

02 전처리

1) 상장기업

2) 신규상장기업

03 모델링

1) 상장기업

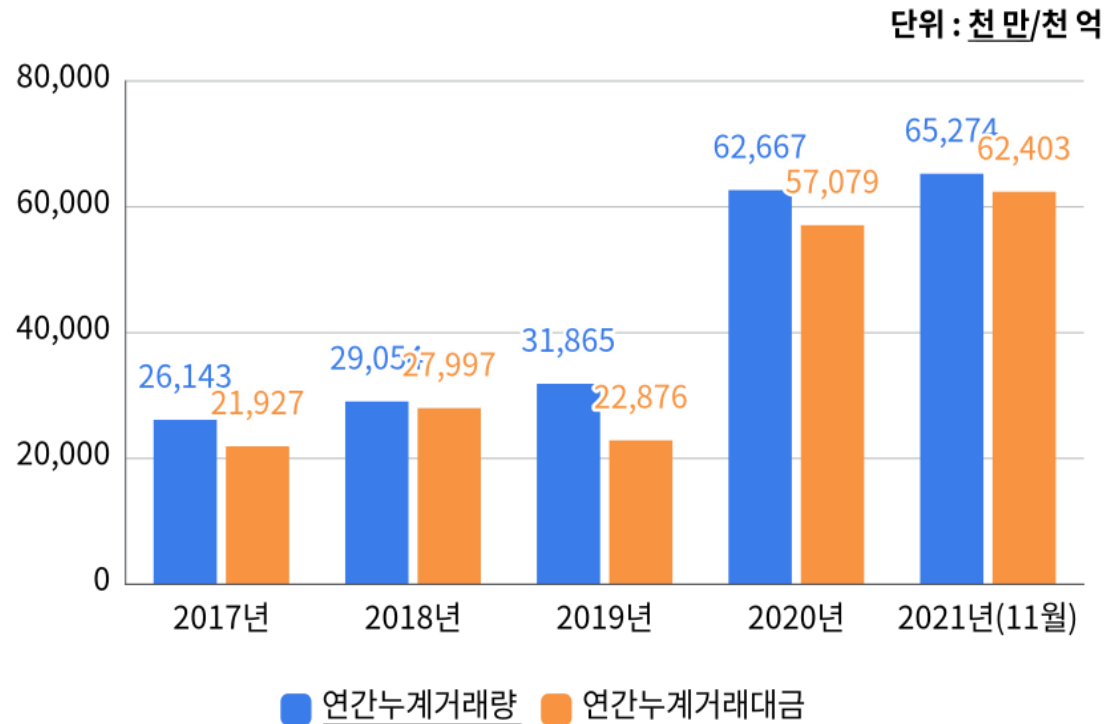
2) 신규상장기업

04 결론 및 한계

01 제안 배경

01 제안배경

COVID-19로 인한 주식열풍



출처: FNGUIDE

주식 투자자 43% "코로나 이후 시작"... 92% "계속할 것"

올해 4월 8일 한국은행이 공개한 '2020년 자금순환(잠정)' 통계에 따르면 지난해 가계 및 비영리단체의 국내외 주식 운용액은 83조 3천억 원으로 기존 최대 금액이었던 2018년 (23조 5천억 원)에 비해 3배 이상 증가한 사상 최대 금액을 기록한 것으로 나타났다.

- 한국일보 -



코로나로 인한 주식 투자 증가로 인해 주가예측 필요성 증가

기술적
분석

주식 그래프 분석을 통해
미래 주가를 예측

기본적
분석

기업의 가치를 분석하여
미래의 주가를 예측



기술적 분석은 주가를 활용하여 예측하는 것이다보니,
기업의 가치를 분석하기는 힘들다

따라서, 재무제표를 이용한 주가예측이 필요하다.

WICS기준 전체기업과 연간누계거래량 상위 300개 기업 비교 (2020년 기준)

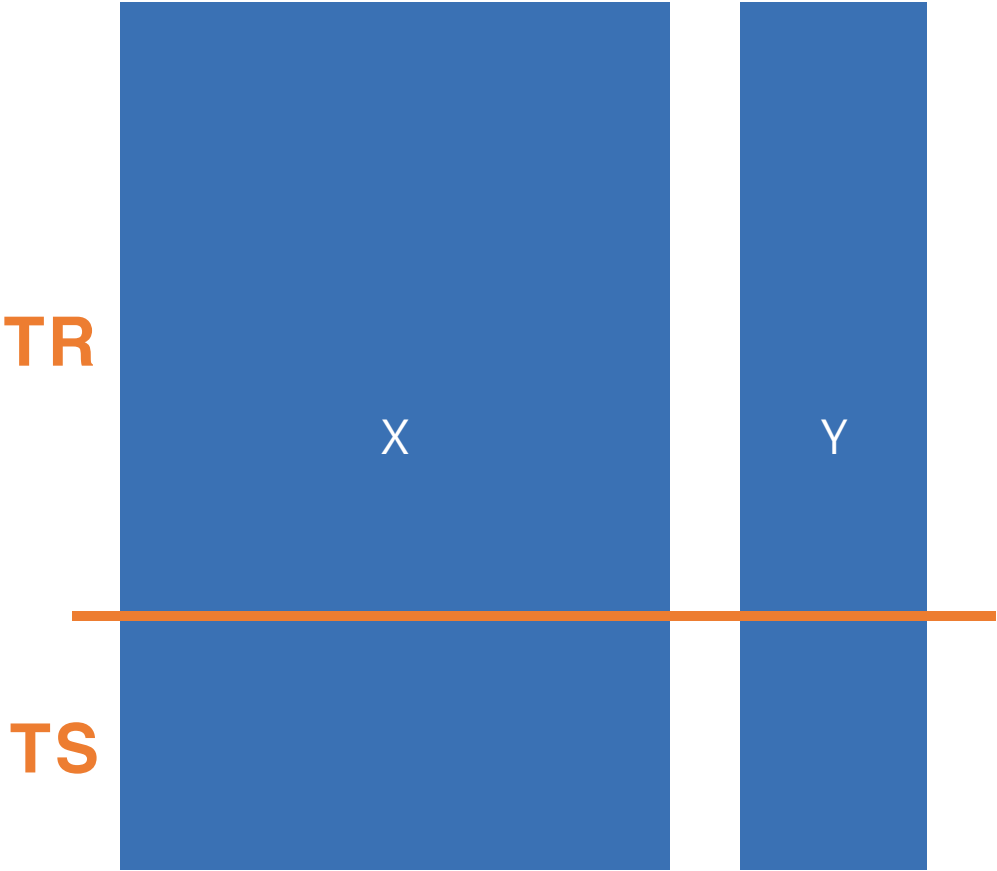
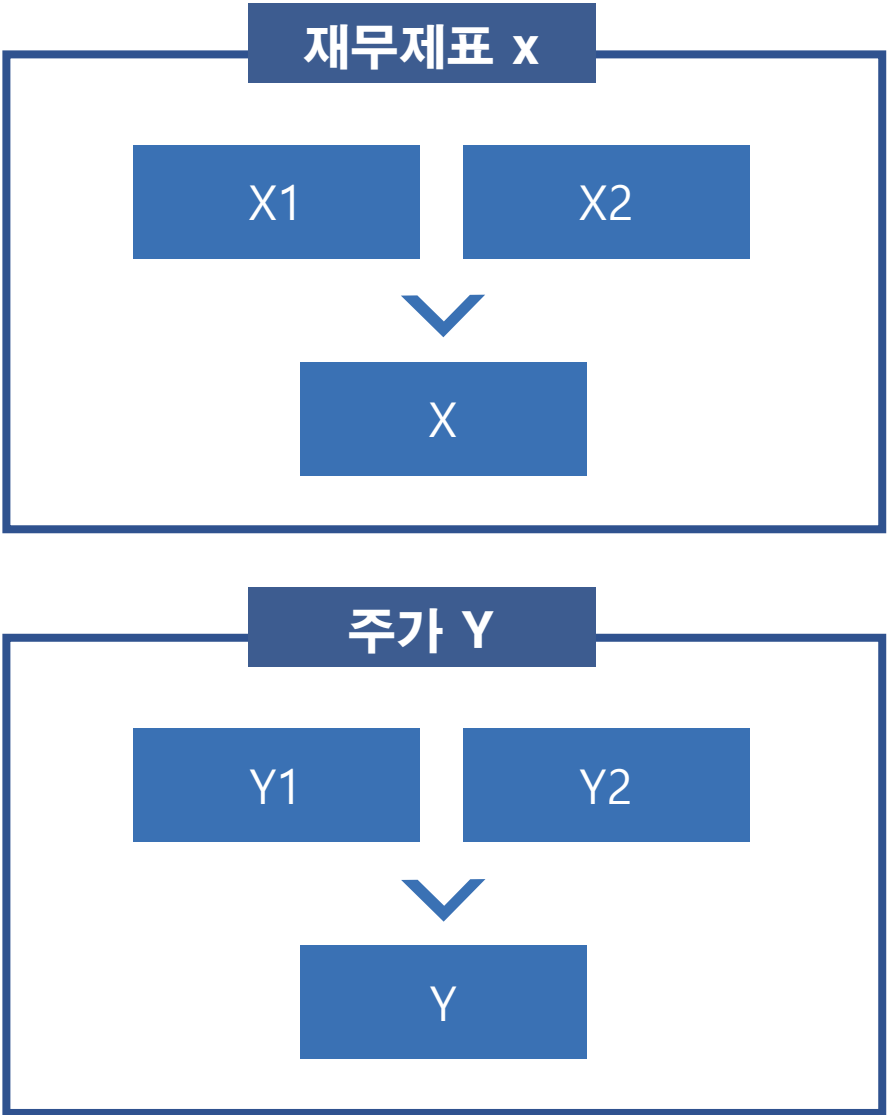
	전체기업 평균	거래량 상위 300개 기업 평균*	거래량 상위 100개 기업 평균	비교* (300과 전체비교)
연간누계거래량	274,301,268	1,315,942,795	2,364,679,229	약 4.8배
당기순이익(지배)	39,332,913	148,303,580	229,552,789	약 3.8배
매출액	1,202,403,013	2,823,731,094	3,085,456,165	약 2.3배
매출액증가율	85%	13%	11%	약 0.15배

자료출처:FNGUIDE

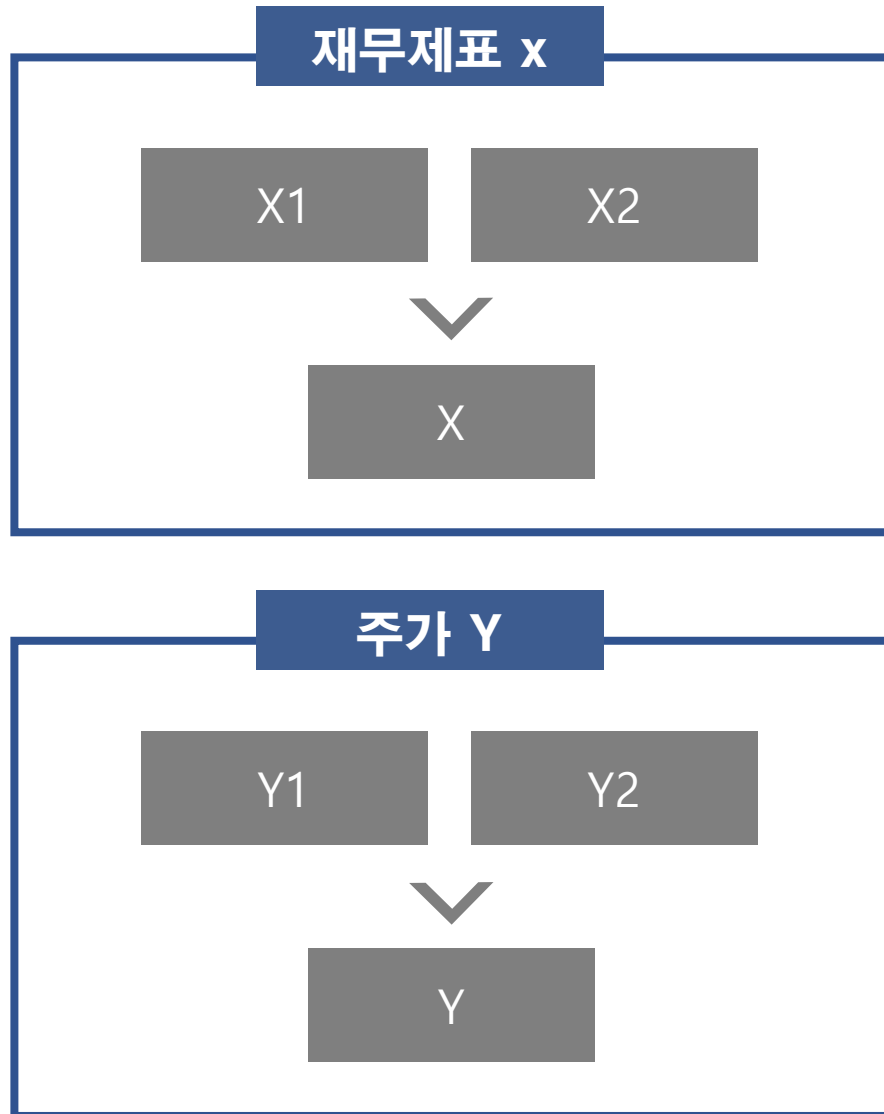
- 측정 지표만을 고려하여 주식을 하거나, 인기종목을 위주로 거래를 하는 성향이 있음을 파악할 수 있음 -> 투자의 안정성 ↓
- 다양한 재무제표 지표를 이용한 주가 예측을 통해 더욱더 안전한 주식 투자 유도 필요

02 전처리

상장기업



신규상장기업



TR



TS



02 전처리

1. 상장기업 - 지표선택

성장성	수익성	안전성	활동성	가치지표 및 재무제표
매출액증가율(QoQ)	매출총이익	부채비율	재고자산회전율	EBITDA
영업이익증가율(QoQ)	영업이익률	차입금의존도	총자산회전율	EV/EBITDAEPS(지배, 우선주제외, Adj.)
순이익증가율(지배, QoQ)	당기순이익(지배)	순부채비율	매출채권회전율	BPS(지배, 무형자산제외, Adj.)
자기자본증가율	매출총이익률	유동부채		P/B(Adj., FY End)
세전계속사업이익증가율(QoQ)	ROIC	유동비율		P/E(Adj., FY End)
EBITDA증가율(QoQ)	ROA	자본유보율		판매비와관리비
EPS증가율(QoQ)	ROE(지배)	부채총계		재무활동으로 인한 현금흐름
유동자산	EBITDA마진율	이자보상배율		영업활동으로인한현금흐름
자기자본비율	영업이익	이자비용		투자활동으로인한현금흐름
자산총계	자본잉여금			CAPEX
자본총계	이익잉여금			Cash Earnings
유형자산	매출액			
자본금				

02 전처리

1. 상장기업 - 재무제표 데이터 가공

처리 전

	Code	Name	자산총계	유동자산	유형자산	부채총계	유동부채	자본총계	자본금	자본잉여금	...
0	A000020	동화약품	447079558.61	227363314.51	129062487.05	95998588.02	66407251.65	351080970.60	27931470.00	27069234.98	...
1	A000040	KR모터스	175866324.91	70893051.99	81655140.86	123434535.82	100580167.73	52431789.09	48068684.00	7876577.00	...
2	A000050	경방	1284289875.52	116158139.40	127291631.76	532223783.90	300245057.04	752066091.63	13707635.00	NaN	...
3	A000060	메리츠화재	26272197931.47	NaN	723799970.82	23817140677.70	NaN	2455057253.78	60312500.00	601409747.88	...
4	A000070	삼양홀딩스	4209148472.68	1640567544.88	1510538057.76	1896599453.61	1013481474.23	2312549019.06	44341645.00	375649504.13	...
...
	Code	Name	자산총계	유동자산	유형자산	부채총계	유동부채	자본총계	자본금	자본잉여금	...
0	A000020	동화약품	437269331.80	226781447.99	124883148.45	96055587.67	68098817.86	341203744.12	27931470.00	27069234.98	...
1	A000040	KR모터스	167929038.19	64695271.01	81517508.73	114107898.77	90210486.67	53821139.42	48068684.00	9081506.00	...
2	A000050	경방	1276354292.04	105848583.15	132576721.80	531290231.17	295122365.91	745064060.87	13707635.00	NaN	...
3	A000060	메리츠화재	25516699467.37	NaN	724080383.30	23187092947.39	NaN	2329606519.98	60312500.00	601409747.88	...
4	A000070	삼양홀딩스	4082285820.95	1570941776.37	1508645250.04	1855692485.41	924196789.21	2226593335.54	44341645.00	376622934.28	...
...



처리 후

	Code	Name	자산총계	유동자산	유형자산	부채총계	유동부채	자본총계	자본금	자본잉여금	...
0	A000020	동화약품	2.24	0.26	3.35	-0.07	-2.48	2.89	0.00	0.00	...
1	A000040	KR모터스	4.73	9.58	0.17	8.17	11.49	-2.58	0.00	-13.27	...
2	A000050	경방	0.62	9.74	-3.99	0.18	1.74	0.94	0.00	NaN	...
3	A000060	메리츠화재	2.96	NaN	-0.04	2.72	NaN	5.39	0.00	0.00	...
4	A000070	삼양홀딩스	3.11	4.43	0.13	2.20	9.66	3.86	0.00	-0.26	...
...

```
df = df1.copy()
for col in df1.columns[2:51]:
    for j in range(len(df1)):
        try:
            if df1[col][j] == 0:
                df[col][j] = np.nan
                continue
            df[col][j] = (df2[col][j] - df1[col][j]) / df1[col][j] * 100
        except:
            pass
```

3월 대비 6월 증가율로 데이터 가공
선택이유: 최신데이터

02 전처리

1. 상장기업 - 주가 데이터 수집

9월 1일 주가 수집

```
for i in range(0, len(h)):
    c = h[i]
    n = df_n[i]

    data = stock.get_market_ohlcw_by_date('2021-09-01', '2021-09-01', c)
    data['회사명'] = n
    df8_start = pd.concat([df8_start, data], axis=0)
```

9월 30일 주가 수집

```
for i in range(0, len(h)):
    c = h[i]
    n = df_n[i]

    data = stock.get_market_ohlcw_by_date('2021-09-30', '2021-09-30', c)
    data['회사명'] = n
    df8_finish = pd.concat([df8_finish, data], axis=0)
```

11월 30일 주가 수집

```
for i in range(0, len(h)):
    c = h[i]
    n = df_n[i]

    data = stock.get_market_ohlcw_by_date('2021-11-30', '2021-11-30', c)
    data['회사명'] = n
    df10_finish = pd.concat([df10_finish, data], axis=0)
```

9월달 설정 이유

6월 재무제표 공시일 기간이 60일 이내로, 기업마다
다르기 때문에 그 이후인 9월 1일로 설정

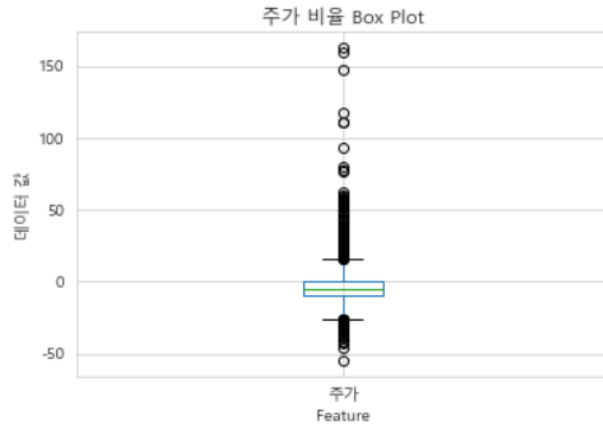
기간 설정 이유

9.1 - 9.30 한달과 다음 분기 재무제표 공시일을
고려하여 9.1-11.30 세 달 간격으로 설정

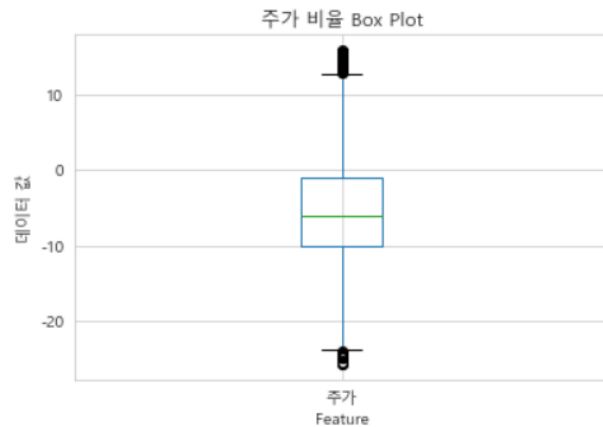
02 전처리

1. 상장기업 - 이상치 제거

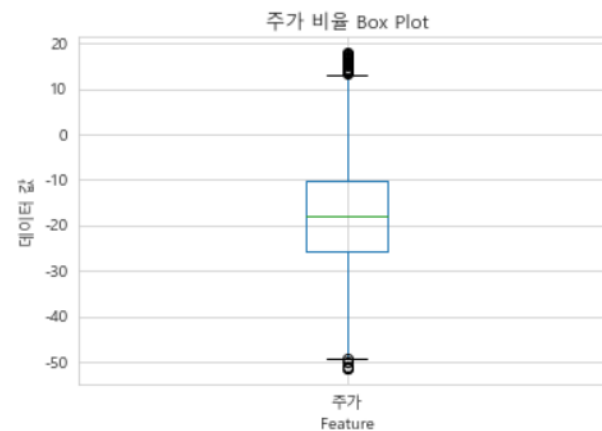
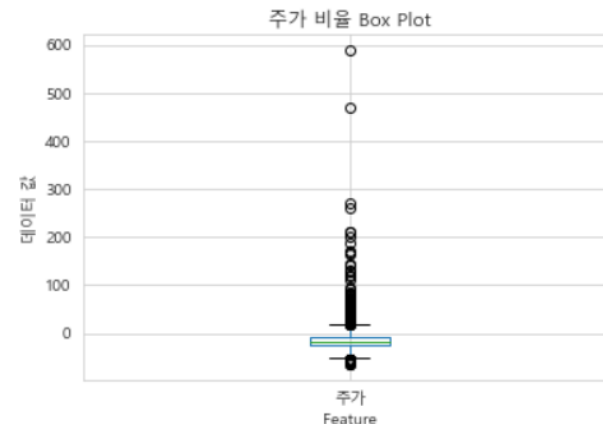
<한달 후 주가 상승률>



편차 감소



<세달 후 주가 상승률>



WHY?

극단값의 경우 재무제표보다는
화제성으로 인한 주식 증가 확률이
더 높을 가능성을 고려하여 IQR의
1.5배 밖에 있는 이상점을 제외하고
분석하였다.

02 전처리

1. 신규상장기업 - 지표선택

성장성	수익성	안전성	활동성	가치지표 및 재무제표
유동자산	매출총이익	부채비율	재고자산회전율	EBITDA
유형자산	영업이익률	차입금의존도	총자산회전율	EPS(지배, 우선주제외, Adj.)
자본총계	당기순이익(지배)	순부채비율	매출채권회전율	EV/EBITDA
자산총계	매출총이익률	유동부채		CAPEX
자기자본비율	ROIC	유동비율		판매비와관리비
자본금	ROA	자본유보율		Cash Earnings
	ROE(지배)	부채총계		투자활동으로인한현금흐름
	EBITDA마진율	이자보상배율		영업활동으로인한현금흐름
	영업이익			재무활동으로 인한 현금흐름
	매출액			

* 상장기업과 비교하여 주가가 들어간 경우, 증가율(신규상장의 경우 재무제표의 영향이 적기 때문에, 다 기간의 재무제표의 영향력이 적을 것이라 가정)인 경우 등을 제외하고 재선택

02 전처리

2. 신규상장기업 - 재무제표 데이터 가공

Code	Name	상장일	자산 총계	유동 자산	유형 자산	부채 총계	유동 부채	자본 총계	자본 금	...	자본 유보 율	순부 채비 율	자입 금의 존도	총자 산회 전율	매출 채권 회전 율	재고 자산 회전 율	EV/EBITDA	EPS(지 배, 우선 주제외, Adj.)	WICS 업종 코드 (대)
A383310	에코프로에이치엔	20210528	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	G15
A383800	LX홀딩스	20210527	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	G20
A276730	제주맥주	20210526	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	G30
A363250	진시스텍	20210526	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	G35
A361670	삼영에스엔씨	20210521	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	G45

상장직전과 상장직후 재무제표 변수를 증가율로 가공



```
df_copy = df_copy[df_copy.isnull().sum(1) < 6]
```

결측치 값이 6개 이상인 행 제거



Name	상장일	자산 총계	유동자 산	유형자 산	부채 총계	유동 부채	자본 총계	자본 금	...	자본 유보 율	순부 채비 율	자입 금의 존도	총자 산회 전율	매출 채권 회전 율	재고 자산 회전 율	EV/EBITDA	EPS(지 배, 우선 주제외, Adj.)	WICS 업종 코드 (대)
에이피오	20210514	71.40	147.09	10.39	-8.59	-10.58	91.17	17.65	...	86.72	104.91	-23.08	-30.79	-16.47	-0.53	NaN	-24.77	G30
쿠른	20210428	41.26	62.46	1.81	0.03	1.42	88.85	21.94	...	71.26	-0.30	-80.60	-12.70	-8.47	NaN	NaN	5.84	G45
엔시스	20210401	-7.16	-12.60	33.59	-52.13	-53.47	6.05	1.48	...	4.61	-14.51	-90.09	-23.96	-2.70	-25.22	NaN	430.77	G45
자이엔트스업	20210324	1.90	-12.36	65.96	13.38	14.06	-1.34	0.93	...	-2.81	-17.86	2.21	-25.54	-15.98	-19.46	-354.70	219.05	G50
재노코	20210324	7.04	-8.55	297.00	11.34	7.30	4.46	0.00	...	4.59	-2.33	-52.88	10.90	40.89	47.89	-69.02	106.67	G20

```
# 분기 마지막 날 +60일 추가 수집
```

```
df_d_ts = []
```

```
for i in range(len(df_d)):
```

```
    df_d2 = dt.datetime.strptime(df_d[i], '%Y-%m-%d')
```

```
    df_d2 = df_d2 + relativedelta(months=2)
```

```
    df_d_ts.append(df_d2)
```

```
# 토요일, 공휴일을 평일로 수정
```

```
XXRX = ecal.get_calendar("XXRX") # 한국 코드
```

```
for i in range(len(df_d_ts)):
```

```
    while XXRX.is_session(df_d_ts[i]) == False:
```

```
        df_d_ts[i] = df_d_ts[i] + relativedelta(days=1)
```

· 상장직후 재무제표 공시일이 다 다르기 때문에, 분기 마지막 날 + 60일 / 90일 추가 수집

· 수집주가 날짜가 공휴일인 경우, 임의로 일수를 추가하여 추가 수집

02 전처리 종속변수 세분화

Name	자산 총계	유동 자산	유형 자산	부채 총계	유동 부채	자본 총계	자본 금	자본 잉여금	...	매출 채권 회전율	재고 자산 회전율	EV/EBITDA	EPS(지 배, 우 선주제 외, Adj.)	BPS(지 배, 무 형자산 제외, Adj.)	EPS증 가율 (QoQ)	P/E(Adj., FY End)	P/B(Adj., FY End)	WICS 업종 코드 (대)	주 가
KR모터 스	4.73	9.58	0.17	8.17	11.49	-2.58	0.00	-13.27	...	46.86	47.22	-82.09	26.67	-9.56	-57.71	-29.60	-2.93	G25	1.00
경방	0.62	9.74	-3.99	0.18	1.74	0.94	0.00	0.00	...	0.11	-6.79	0.14	52.21	1.89	4405.22	-28.88	4.35	G25	2.00
메리츠 화재	2.96	0.00	-0.04	2.72	0.00	5.39	0.00	0.00	...	-2.68	0.00	-12.12	24.77	5.82	61.98	-15.49	0.00	G40	2.00
삼양홀 딩스	3.11	4.43	0.13	2.20	9.66	3.86	0.00	-0.26	...	1.78	1.17	-5.42	-6.05	6.88	-93.60	25.45	13.33	G30	1.00
하이트 진로	-11.55	-29.57	-0.41	-16.72	-22.17	2.25	0.00	0.00	...	8.94	0.31	39.48	-18.55	2.93	-94.01	25.63	0.00	G30	1.00
...

```

for i in range(len(model1)):
    if model1['주가'][i] < -15:
        model1['주가'][i] = 0
    elif model1['주가'][i] >= -15 and model1['주가'][i] < 0:
        model1['주가'][i] = 1
    elif model1['주가'][i] >= 0 and model1['주가'][i] < 15:
        model1['주가'][i] = 2
    else:
        model1['주가'][i] = 3

```



단순히 이진분류가 아닌, 좀 더 세부적으로
구분하기 위해 -15%와 0% 및 15% 기준
으로 다진분류하였다.

03 모델링

03 모델링

모델링 코드

< GBM >

```
boosting_gbm_ensemble = GradientBoostingClassifier()

boosting_gbm_ensemble.fit(X_train, y_train)
y_pred = boosting_gbm_ensemble.predict(X_test)

print("Training Score: {:.3f}".format(boosting_gbm_ensemble.score(X_train, y_train)))
print("Test Score: {:.3f}".format(boosting_gbm_ensemble.score(X_test, y_test)))
```

< XGBOOST >

```
xgb = XGBClassifier(num_class=7, learning_rate=0.1, max_depth=10, eval_metric='logloss')
xgb.fit(X_train, y_train)
xgb_pred = xgb.predict(X_test)

print("Training Score: {:.3f}".format(xgb.score(X_train, y_train)))
print("Test Score: {:.3f}".format(xgb.score(X_test, y_test)))
```

< SVM >

```
svm = SVC()
svm.fit(X_train, y_train)
svm_pred = svm.predict(X_test)

print("Training Score: {:.3f}".format(svm.score(X_train, y_train)))
print("Test Score: {:.3f}".format(svm.score(X_test, y_test)))
```

< RANDOMFOREST >

```
forest = RandomForestClassifier(random_state = 0)
model=forest.fit(X_train, y_train)

print("Training Score: {:.3f}".format(forest.score(X_train, y_train)))
print("Test Score: {:.3f}".format(forest.score(X_test, y_test)))
```

랜덤포레스트 파라미터 튜닝

```
params = {
    'n_estimators': [100], #100개의 sub Dataset, Decision Tree
    'max_depth': [6, 8, 10, 12],
    'min_samples_leaf': [8, 12, 18],
    'min_samples_split': [8, 16, 20]
}

model = RandomForestClassifier(random_state=0, n_jobs=-1)
grid_cv = GridSearchCV(model, param_grid=params, cv=2, n_jobs=-1)
grid_cv.fit(X_train, y_train)

print("Best hyperparameter: ", grid_cv.best_params_)
```

Best 예측 값

```
final_bagging_rf_ensemble = RandomForestClassifier(n_estimators=100, max_depth=6, min_samples_leaf=8,
                                                    min_samples_split=8, random_state=0)

final_bagging_rf_ensemble.fit(X_train, y_train)
y_pred = final_bagging_rf_ensemble.predict(X_test)

print(f'Best accuracy score: {accuracy_score(y_test, y_pred):.3f}')
```

03 모델링

1. 상장기업 - 모델링 결과

< 한 달 >

	GBM	XGBOOST	SVM	RANDOMFOREST
TRAINING SCORE	0.863	0.996	0.697	0.996
TEST SCORE	0.680	0.698	0.700	0.700

< 세 달 >

	GBM	XGBOOST	SVM	RANDOMFOREST
TRAINING SCORE	0.862	0.995	0.619	0.995
TEST SCORE	0.558	0.570	0.582	0.594

재무제표로 인한 주가변동 예측주기는 한달이 더 정확하다는 것을 알 수 있음

03 모델링

2. 신규상장기업 - 모델링 결과

*상장기업의 독립변수와 신규 상장기업의 독립변수가 서로 달라 독립변수 통일

< 한 달 >

	GBM	XGBOOST	SVM	RANDOMFOREST
TRAINING SCORE	0.813	0.996	0.700	0.996
TEST SCORE	0.385	0.355	0.391	0.379

상장직전과 상장직후의 재무제표는 주가에 미치는 영향이 적음을 알 수 있음

04 결론 및 한계

결론

1. 재무제표를 통한 주가 예측으로 사람들이 쉽게 추천종목을 파악할 수 있다.
2. 재무제표로 인한 주가변동예측 주기는 1달이 더 정확하다는 것을 알 수 있었다.
3. 기업이 변동성이 클 경우, 재무제표보다는 기업의 화제성으로 투자하는 경우가 많아 예측하기 어렵다.



한계

독립 변수

주가에 영향을 미치는 요소들이 많기 때문에, 재무제표 외에 다른 지표들을 사용할 시 더 좋은 결과가 나올 것으로 예상

재무제표 공시일

기업들의 공시기한이 60일로, 각 회사의 공시일이 다르기 때문에 주가날짜를 통합하여 예측하였다.



감사합니다.