삼성전자
주가 예측 모델 구현 PJT

조아진

CONTENTS

01 기획 배경

삼성전자 주가 "무리 없다" vs "예측 어렵다"

등록 2021-01-19 05:00:00 | 수정 2021-01-19 05:07:13

[서울=뉴시스]고승민 기자 = 이재용 삼성전자
역 2년6개월을 선고받아 법정구속된
kkssmm99@newsis.com

[서울=뉴시스] 김제이 기자 = 최근 상승세를 보이며 '9만전자'까지 올라갔던 삼성전자의 주가에 대한 관심이 높아지고 있다. 이재용 부회장 구속에 따라 변수가 생겼기 때문이다.

증권가에서는 단기적으로는 이번 일이 삼성전자 주가에 대한 조정의 빌미가 될 수 있다고 지적했다. 하지만 실적 호전이 이어지고 있기 때문에 장기 상승 추세는 이어질 것이란 전망이 더 우세한 상황이다.

삼성전자, '십만전자'는 언제쯤… "본격 상승은 하반기부터"

1분기 호실적에도 주가는 지지부진
반도체 실적 개선 효과…"3분기부터 나타날 것"

김건우 ohtdue@ekn.kr 최종 기사입력 2021-04-27 06:69:39

삼성전자 주식 투자자를 위한 주가 상승/하락 예측

3개월째 지지부진한 양상을 보이면서 언제쯤 주
고 있다. 증권가에서는 삼성전자 주가가 본격적으
소되는 3분기에나 가능할 것이라는 전망이 나온

26일 한국거래소에 따르면 삼성전자는 올해 1월 11일 종가 기준 9만1000원으로 최고가를 경신한 뒤 최근 들어서는 8만원대에서 횡보하고 있다. 현재 주가는 8만2900원대다.

개인투자자는 작년에 이어 올해도 삼성전자 주식을 대거 매수했지만, 주가는 여전히 지지부진한 흐름을 보이고 있다. 올해 들어 이달 26일까지 개인이 사들인 삼성전자 주식은 17조3171억원에 달한다. 이는 같은 기간 개인들의 코스피 전체 순매수 대금(41조2283억원)의 42%에 달하는 수치다. 불과 5개월도 안돼 지난해 1년간 삼성전자 순매수액(9조5952억원)을 뛰어넘은 셈이다.

02 데이터 소개

데이터 소개

- 삼성전자 주식 '일별 시세' data
- 2019.01.01부터 2021.06.30까지 '삼성전자' 종목의 시가, 고가, 저가, 종가, 거래량 데이터 수집

	Date	Open	High	Low	Close	Volume	Change
0	2019-01-02	39400	39400	38550	38750	7847664	0.001292
1	2019-01-03	38300	38550	37450	37600	12471493	-0.029677
2	2019-01-04	37450	37600	36850	37450	14108958	-0.003989
3	2019-01-07	38000	38900	37800	38750	12748997	0.034713
4	2019-01-08	38000	39200	37950	38100	12756554	-0.016774
...	...	...	...	...	...	...	...
612	2021-06-24	80400	81400	80100	81200	18771080	0.013733
613	2021-06-25	81500	81900	81200	81600	13481405	0.004926
614	2021-06-28	81700	82000	81600	81900	11578529	0.003676
615	2021-06-29	81900	82100	80800	81000	15744317	-0.010989
616	2021-06-30	81100	81400	80700	80700	13288643	-0.003704

617 rows × 7 columns

03 수집&전처리

데이터 불러오기

```
stock_data=pd.read_csv('C:₩study₩samsung5.csv', encoding='cp949')
stock_data
```

	Date	Open	High	Low	Close	Volume	Change
0	2019-01-02	39400	39400	38550	38750	7847664	0.001292
1	2019-01-03	38300	38550	37450	37600	12471493	-0.029677
2	2019-01-04	37450	37600	36850	37450	14108958	-0.003989
3	2019-01-07	38000	38900	37800	38750	12748997	0.034713
4	2019-01-08	38000	39200	37950	38100	12756554	-0.016774
...	...	...	...	...	...	...	...
612	2021-06-24	80400	81400	80100	81200	18771080	0.013733
613	2021-06-25	81500	81900	81200	81600	13481405	0.004926
614	2021-06-28	81700	82000	81600	81900	11578529	0.003676
615	2021-06-29	81900	82100	80800	81000	15744317	-0.010989
616	2021-06-30	81100	81400	80700	80700	13288643	-0.003704

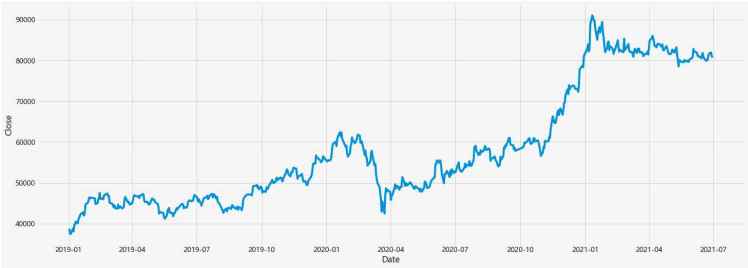
617 rows × 7 columns

```
stock_data.shape
```

(617, 1)

04 시각화

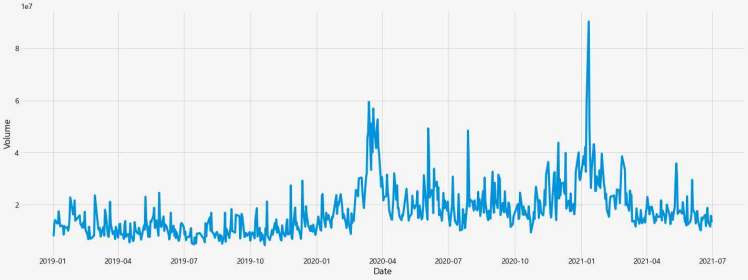
종가 데이터 시계열 추이



- 2019년 초부터 2019년 말까지 증가 추세.
- 2020년 초 '코로나19바이러스'의 영향으로 2020년 3월 20일날 역사적인 하락세를 보임.
- 이후 주가는 계속 증가하며 상승곡선을 보이다가 최근 들어 다소 하락세.

04 시각화

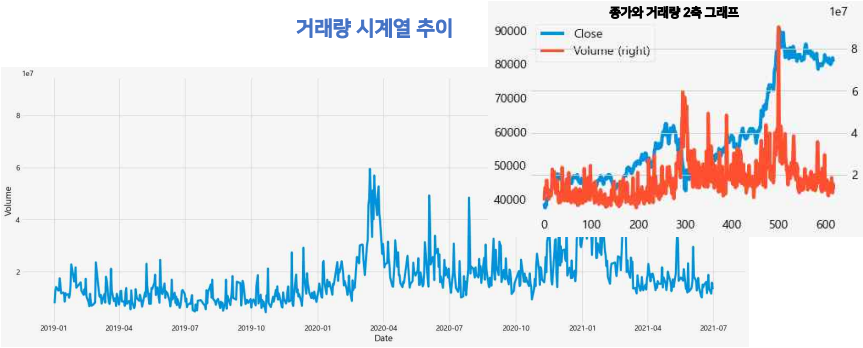
거래량 시계열 추이



- 2020년 초 '코로나19바이러스'의 영향으로 2020년 3월 경 역사적인 하락세를 기록했을 당시 거래량 크게 증가함.
- 2021년 1월 경 주가 상승 시기에 거래량 크게 증가.

04 시각화

거래량 시계열 추이



- 2020년 초 '코로나19바이러스'의 영향으로 2020년 3월 경 역사적인 하락세를 기록했을 당시 거래량 크게 증가함.
- 2021년 1월 경 주가 상승 시기에 거래량 크게 증가.

04 시각화

관련 기사 워드클라우드

최근 1년간 삼성전자 주식 관련 기사를 수집해서 워드클라우드 생성함.

05 ARIMA모형

ARIMA 모형

- AR(Autoregression) 모형과 MA(Moving Average) 모형을 합친 모형
- ARIMA모형은 시계열 데이터의 정상성(Stationary)를 가정함

'정상성'이란?

- 평균, 분산이 시간에 따라 일정한 성질
- 시계열 데이터의 특성이 시간의 흐름에 따라 변하지 않음을 의미함
- 추세나 계절성이 있는 시계열은 정상 시계열이 아님
- 정상성을 나타내지 않는 데이터는 정상성을 갖도록 로그변환, 차분 등 전처리 후 분석 시행

정상 시계열 변환 방법

- 변동폭이 일정하지 않은 경우 -> 로그 변환
- 추세, 계절성이 존재하는 경우 -> 차분 (1차 차분으로 정상성을 띄지 않으면 차분 반복)

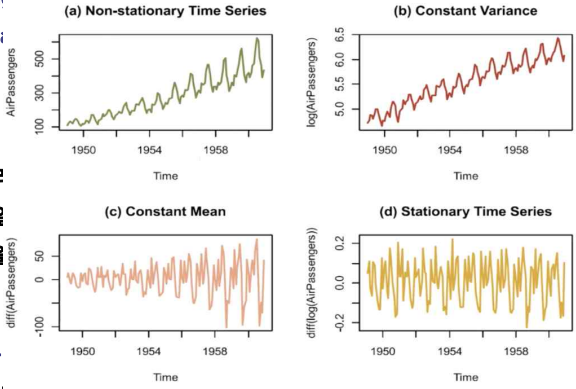
05 ARIMA모형

ARIMA 모형

- AR(Autoregression) 모형과 MA(Moving Average) 모형을 합한 모형
- ARIMA모형은 시계열 데이터의 정상성(Sta

'정상성'이란?

- 평균, 분산이 시간에 따라 일정한 성질
- 시계열 데이터의 특성이 시간의 흐름에 따라
- 추세나 계절성이 있는 시계열은 정상 시계열
- 정상성을 나타내지 않는 데이터는 정상성을



정상 시계열 변환 방법

- 변동폭이 일정하지 않은 경우 -> 로그 변환
- 추세, 계절성이 존재하는 경우 -> 차분 (1차 차분으로 정상성을 띄지 않으면 차분 반복)

05 ARIMA모형

AR모형이란?

자귀 회귀 모형으로, Auto Correlation의 약자이다.
자기상관성을 시계열 모형으로 구성하였으며, 예측하고자 하는 특정 변수의 과거 관측값의 선형결합으로 해당 변수의 미래값을 예측하는 모형이다.
이전 자신의 관측값이 이후 자신의 관측값에 영향을 준다는 아이디어에 기반하였다.

AR(p) 모형의 식은 다음과 같다.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

y_t는 t시점의 관측값, c는 상수, ϕ는 가중치, ε_t는 오차항을 의미한다.

MA모형이란?

Moving Average 모형으로, 예측 오차를 이용하여 미래를 예측하는 모형이다.

MA(q) 모형의 식은 다음과 같다.

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

ARIMA 모형이란?

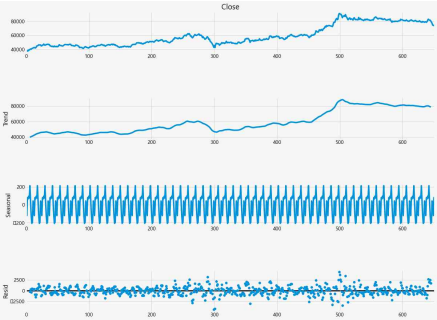
ARIMA(p,d,q) 모형은 d차 차분한 데이터에 위 AR(p) 모형과 MA(q) 모형을 합친 모형으로, 식은 다음과 같다.

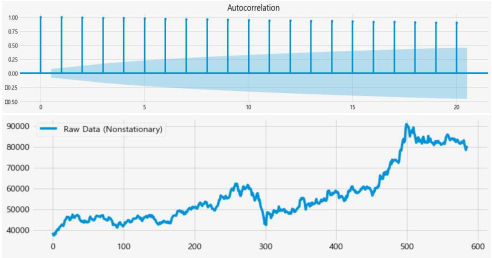
$$y_t' = c + \phi_1 y_{t-1}' + \phi_2 y_{t-2}' + \ldots + \phi_p y_{t-p}' + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

06 모델링

정상성 확인(Autocorrelation Function의 패턴 이용)

- 귀무가설(H0): 정상성을 만족하지 않는다.
- 대립가설(H1): 정상성을 만족한다.



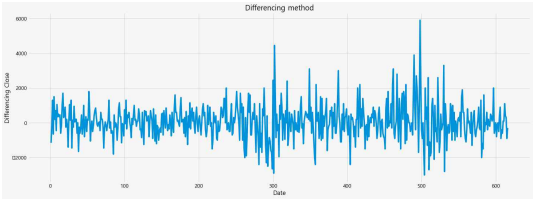


1. ADF : -0.6748918214310162
2. P-Value : 0.8531463786480 17 **P-value가 0.85이므로 유의확률 0.05에서 H0기각 불가**
3. Num Of Lags : 0
4. Num Of Observations Used For ADF Regression: 616
5. Critical Values :
   - 1% : -3.441010323593974 6
   - 5% : -2.866243743831338
   - 10% : -2.5692748053002195

➡ 해당 데이터는 정상성을 만족하지 못함 (->1차 차분 시행)

06 모델링

정상 시계열 변환 (1차 차분)




```
#ADF 검정 결과
ad_test(ts_diff[1:])
```

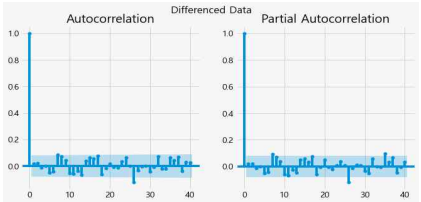
```
1. ADF :  -24.344905907582433
2. P-Value :  0.0
3. Num Of Lags :  0
4. Num Of Observations Used For ADF Regression: 615
5. Critical Values :
        1% :  -3.4410277306083668
        5% :  -2.8662510413264357
        10% :  -2.5692788902108853
```

P-value가 0.0으로 유의확률 0.05에서 귀무가설 H0기각

➡ 1차 차분한 데이터는 정성성을 만족함

ARIMA모델링(Autoregressvie Integrated MovingAverage)

- 현재값을 과거값 + 과거 예측 오차를 통해 설명함




```
ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=9766.265, Time=0.01 sec
ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=9768.110, Time=0.03 sec
ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=9770.031, Time=0.05 sec
ARIMA(0,1,3)(0,0,0)[0] intercept   : AIC=9771.770, Time=0.07 sec
ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=9768.107, Time=0.03 sec
ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=9770.111, Time=0.22 sec
ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=9771.333, Time=0.22 sec
ARIMA(1,1,3)(0,0,0)[0] intercept   : AIC=9772.749, Time=0.19 sec
ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=9770.030, Time=0.04 sec
ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=9772.016, Time=0.19 sec
ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=9773.022, Time=0.53 sec
ARIMA(2,1,3)(0,0,0)[0] intercept   : AIC=9774.181, Time=0.46 sec
ARIMA(3,1,0)(0,0,0)[0] intercept   : AIC=9771.744, Time=0.06 sec
ARIMA(3,1,1)(0,0,0)[0] intercept   : AIC=9773.716, Time=0.31 sec
ARIMA(3,1,2)(0,0,0)[0] intercept   : AIC=9775.726, Time=0.28 sec

Best model:  ARIMA(0,1,0)(0,0,0)[0] intercept
Total fit time: 2.706 seconds
```

07 모델 평가

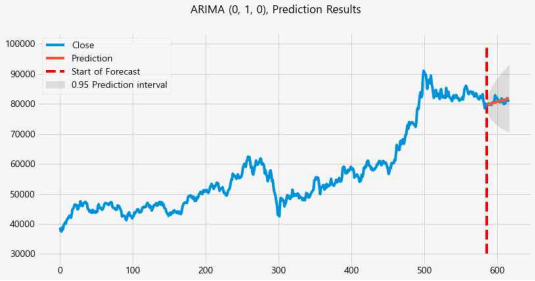
ARIMA모델링(Autoregressvie integrated MovingAverage)

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	586
Model:	SARIMAX(0, 1, 0)	**Log Likelihood**	-4881.132
Date:	Thu, 19 Aug 2021	**AIC**	9766.265
Time:	22:59:01	**BIC**	9775.008
Sample:	0	**HQIC**	9769.672
	- 586		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	69.8291	43.389	1.609	0.108	-15.211	154.869
sigma2	1.035e+06	4e+04	25.854	0.000	9.56e+05	1.11e+06

Ljung-Box (L1) (Q):	0.20	**Jarque-Bera (JB):**	228.56
Prob(Q):	0.65	**Prob(JB):**	0.00
Heteroskedasticity (H):	3.42	**Skew:**	0.54
Prob(H) (two-sided):	0.00	**Kurtosis:**	5.86



향후 주가가 다소 상승세를 보일 것으로 예측함

08 개발 후기

- 해당 프로젝트를 진행하기 전에 주가 예측은 불가능한 영역일 것이라고 생각했었다.
- 프로젝트를 진행하면서 정확한 주가는 아니더라도 주가 상승 혹은 하락 여부 정도는 예측이 가능할 수도 있겠다는 생각이 들었다.

- 본 프로젝트에서는 2019.01.01~2021.06.30까지의 데이터로 분석을 진행하여 삼성전자의 주가 상승을 예측하였으나, 현 2021.8월 기준 삼성전자 주가는 하락하였다.
- 지켜볼 필요가 있겠으나 단기적으로 보았을 때는 주가가 예상치 못한 수많은 변수가 관여할 수 있는 분야이기 때문에 모델링을 통한 예측이 쉽지 않은 영역인 것 같다.
- 해당 회사의 재정상황이나 실적, 행보뿐만 아니라 관련 종목의 기업 및 타 종목 흐름 해외의 움직임까지 고려해야 하기 때문에 다양한 환경적 요소를 더하여서 다각도의 분석이 필요할 것이라고 생각이 된다.

- A기업의 주가를 예측한다면, A기업 외에 관련 종목 데이터 (ex.B기업, C기업, D기업 등)도 함께 활용하여 예측하면 좀 더 예측력과 정확도가 높은 분석이 될 것 같다.
- 본 프로젝트는 삼성전자의 데이터만 가지고 분석을 진행하였기 때문에 다소 아쉬움이 있다.

Thank You