



Predicting stock prices  
based on people's reactions.

201600779 김영민

# Context



**01** Motivation

**02** Introduce My Project

**03** Interim findings

**04** Method of Data Collection

**05** Trouble Shooting

**06** Next Process

# Motivation

어론 속의 여론

주식 투자자 43% "코로나 이  
후 시작"... 92% "계속할 것"

♡ 1 ○ 0

주식 직접투자 행태 및 투자자산 인식 조사

업력 2021.05.06 04:30



〈그림 1〉 개인투자자 거래규모



Then



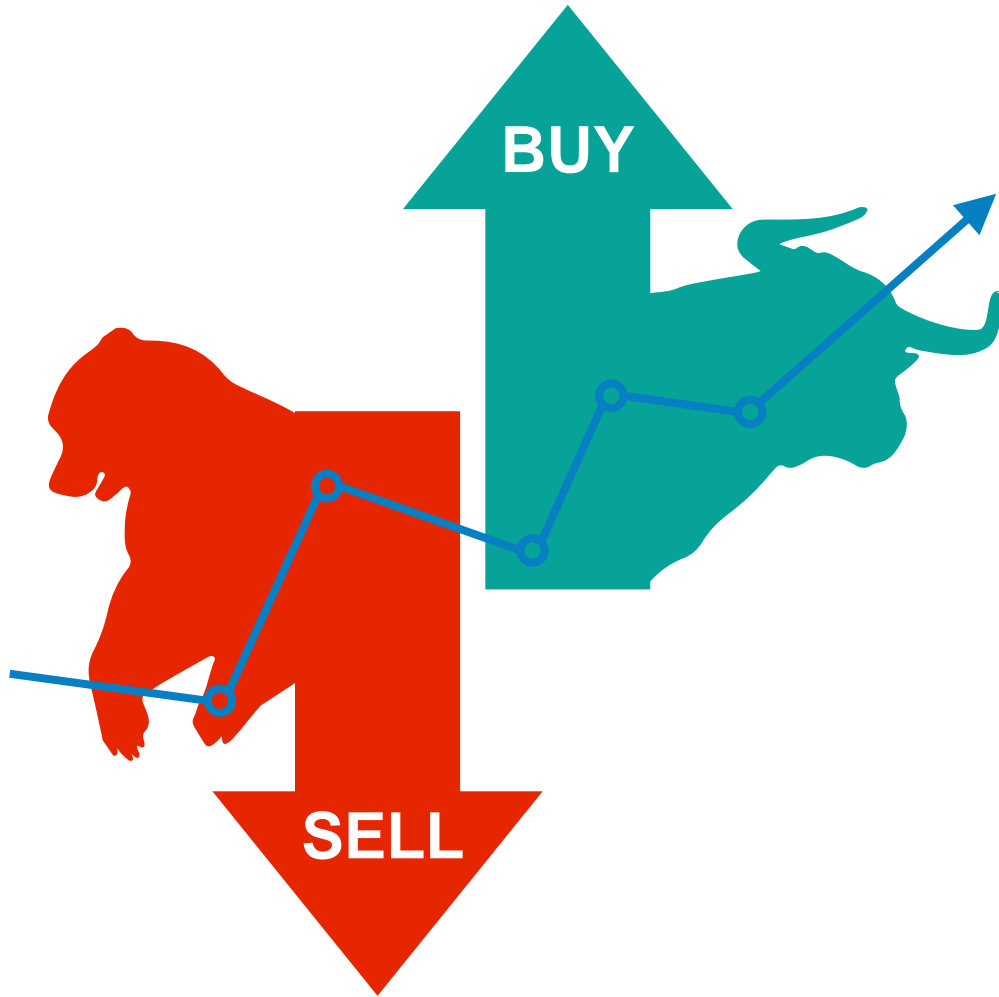
사람들의 기대 심리가  
주가에 영향을 미칠까???

개인들이 가장 많이 접하게 되는  
회사 소식인 기사들에 의해  
주가 변동에 영향이 있을까?

코로나 이후 개인 투자자 비율



# Introduce My Project



Is **Individual reaction** related to **stock prices**?

## 종목

- ❖ 삼성전자
- ❖ 네이버
- ❖ 셀트리온
- ❖ 기업은행

## 기간

- ❖ 2021.01~2021.10.15
- ❖ Test는 2021.11월 주가로 시행

## 수집 데이터

- ❖ Investing.com의 여론 데이터
- ❖ 네이버 기사 데이터
- ❖ 시작가와 종가 데이터

## Prediction Model

- ❖ RandomForest Classification
- ❖ LGBM Classification
- ❖ LSTM or GRU

## Semantic Analysis

- ❖ KNU 감정사전
- ❖ VADER
- ❖ Google Cloud API

## 발전 방향

- ❖ 코로나 전후 1년 비교
- ❖ 딥러닝을 활용한 감정 분석
- ❖ 여러 사이트를 활용한 감정 분석

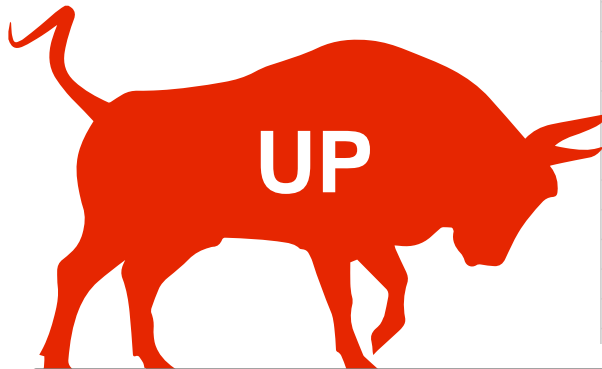
# Interim findings



# Investing.com Opinion



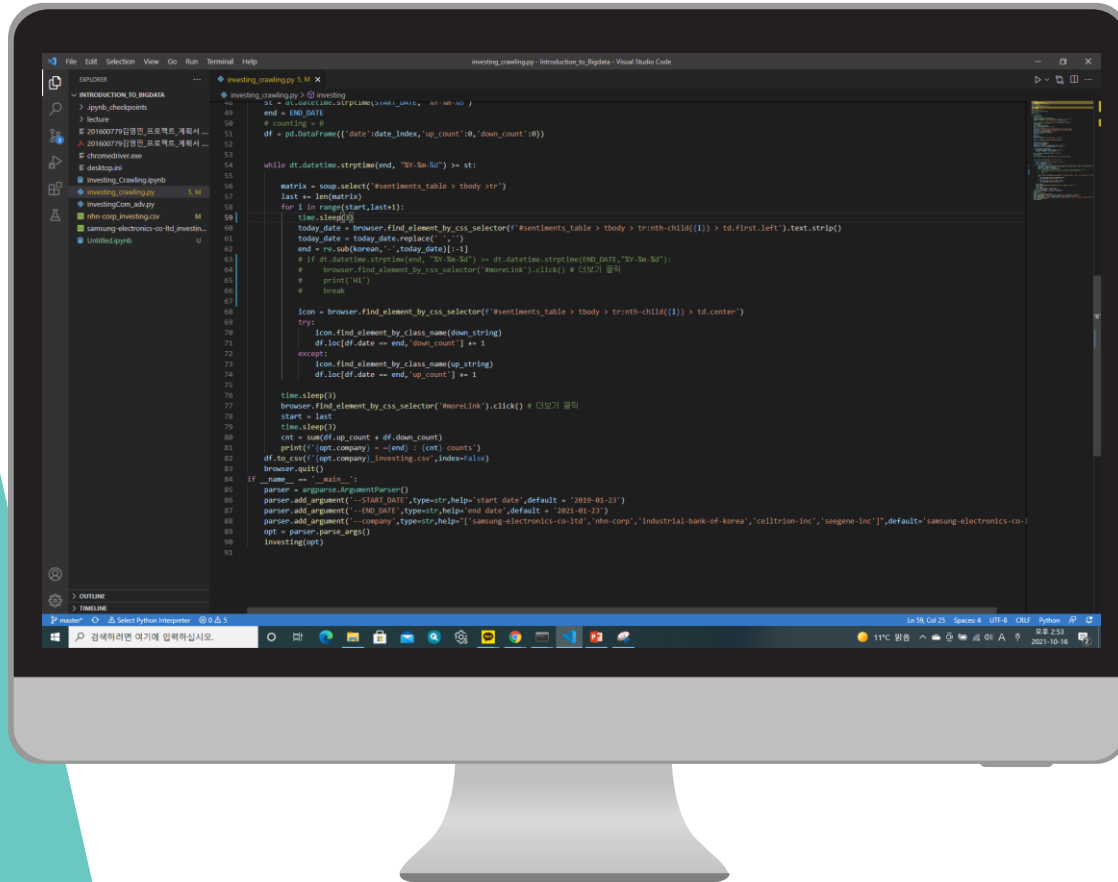
전망에 **빨간 소**가  
있으면 **상승** 예측



전망에 **파란 곰**이  
있으면 **하락** 예측



# Crawling



일반 BeautifulSoup 로는 접근이 불가능 하기 때문에  
Selenium을 이용하여 사이트 접근

## BeautifulSoup

BeautifulSoup를 사용하여 원하는 정보 추출

[https://github.com/winston1214/INU/blob/master/Introduction to Bigdata/investing\\_crawling.py](https://github.com/winston1214/INU/blob/master/Introduction%20to%20Bigdata/investing_crawling.py)

# Stock Price

## 주가 데이터 다운로드

005930 역사적 데이터

기간:

일간

데이터 다운로드

2021/09/16 - 2021/10/16

i

날짜	종가	오픈	고가	저가	거래량	변동 %
2021년 10월 15일	70,100	70,200	71,000	70,000	17.53M	1.01%
2021년 10월 14일	69,400	69,100	69,800	68,800	0.23K	0.87%
2021년 10월 13일	68,800	68,700	69,600	68,400	0.57K	-0.29%
2021년 10월 12일	69,000	70,800	70,900	68,700	18.42K	-3.50%
2021년 10월 11일	71,500	71,500	71,500	71,500	-	0.00%
2021년 10월 08일	71,500	72,300	72,400	71,500	13.97M	-0.14%
2021년 10월 07일	71,600	71,800	72,100	71,300	0.17K	0.42%
2021년 10월 06일	71,300	72,700	72,800	71,200	8.34K	-1.25%
2021년 10월 05일	72,200	72,900	73,000	71,400	1.12K	-1.37%
2021년 10월 04일	73,200	73,200	73,200	73,200	-	0.00%
2021년 10월 01일	73,200	73,900	74,000	72,900	15.70M	-1.21%
2021년 09월 30일	74,100	74,300	74,800	73,800	6.44K	0.00%
2021년 09월 29일	74,100	74,800	75,300	73,800	15.27K	-2.88%
2021년 09월 28일	76,300	77,600	77,600	76,200	0.94K	-1.80%
2021년 09월 27일	77,700	77,300	77,700	77,000	11.63M	0.52%
2021년 09월 26일	77,300	77,300	77,300	77,300	-	0.00%
2021년 09월 24일	77,300	77,600	77,700	77,100	11.85M	-0.13%
2021년 09월 23일	77,400	77,500	77,600	76,900	0.49K	0.26%
2021년 09월 22일	77,200	77,200	77,200	77,200	-	0.00%
2021년 09월 17일	77,200	76,300	77,200	75,900	15.66M	1.45%
2021년 09월 16일	76,100	77,300	77,400	76,100	12.88M	-1.17%
최고: 77,700	최저: 68,400	차이: 9,300	평균: 73,643	변동 %: -9		

데이터 다운로드가 한 달  
단위로 제공되기 때문에  
데이터 결합 필요



```
import pandas as pd
import os
ls = os.listdir('samsung')
df_list = []
for i in ls:
    data = pd.read_csv('samsung/'+i)
    df_list.append(data)
df = pd.concat(df_list).reset_index(drop=True)
```

	날짜	종가	오픈	고가	저가	거래량	변동 %
0	2021년 02월 26일	82,500	82,800	83,400	82,000	36.72M	-3.28%
1	2021년 02월 25일	85,300	84,000	85,400	83,000	6.35K	4.02%
2	2021년 02월 24일	82,000	81,900	83,600	81,400	0.98K	0.00%
3	2021년 02월 23일	82,000	81,300	82,900	81,100	7.03K	-0.24%
4	2021년 02월 22일	82,200	83,800	84,200	82,200	17.48K	-0.48%
...	...	...	...	...	...	...	...
230	2021년 09월 06일	77,300	76,800	77,600	76,600	3.42K	0.91%
231	2021년 09월 05일	76,600	76,600	76,600	76,600	-	0.00%
232	2021년 09월 03일	76,600	76,400	76,700	76,000	11.89M	0.79%
233	2021년 09월 02일	76,000	76,800	76,800	75,700	1.76K	-1.04%
234	2021년 09월 01일	76,800	76,600	77,100	75,900	0.84K	0.13%

235 rows × 7 columns

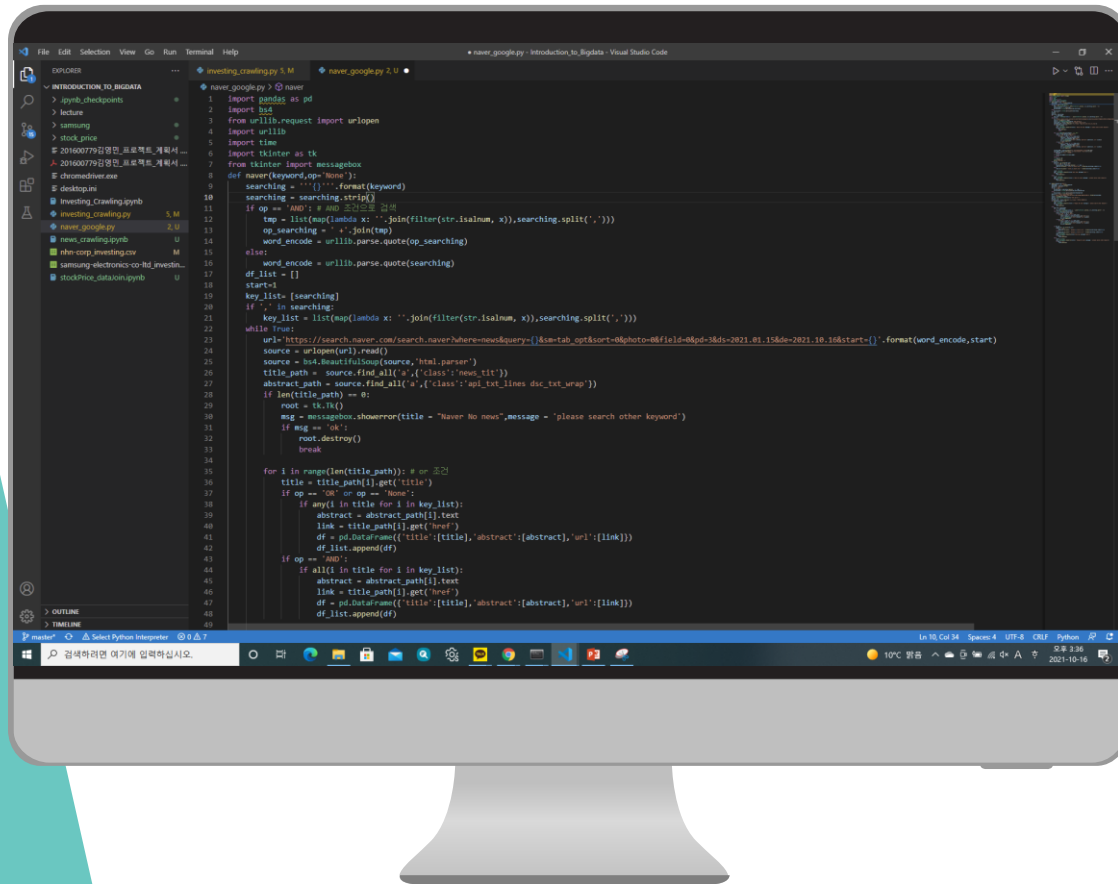
By investing.com



# Naver News Crawling

## BeautifulSoup

BeautifulSoup를 사용하여 원하는 정보 추출



[https://github.com/winston1214/project/tree/master/Keyword\\_Search\\_News\\_Crawling\\_Machine](https://github.com/winston1214/project/tree/master/Keyword_Search_News_Crawling_Machine)

title	abstract	url	date	engine
"20일 삼성	삼성전자의	<a href="http://mon">http://mon</a>	2021.10.15	naver
"삼성전자	개미(개인	<a href="https://www">https://www</a>	2021.10.16	naver
떨어지는	추가, 연고	<a href="http://yna">http://yna</a>	2021.10.14	naver
삼성전자	가조 바이든	<a href="http://new">http://new</a>	2021.10.16	naver
삼성전자,	삼성전자는	<a href="http://www">http://www</a>	2021.10.15	naver
삼성전자,	기사내용	<a href="http://www">http://www</a>	2021.10.15	naver
삼성전자,	삼성전자는	<a href="http://www">http://www</a>	2021.10.15	naver
[특징주]	삼성전자	<a href="http://yna">http://yna</a>	2021.10.15	naver
삼성전자,	반도체	<a href="https://biz">https://biz</a>	2021.10.15	naver
삼성전자	국내 코스	<a href="http://new">http://new</a>	2021.10.15	naver
[특징주]	삼성전자	<a href="http://yna">http://yna</a>	2021.10.15	naver
[속보]	삼성전자	<a href="http://www">http://www</a>	2021.10.15	naver
7만전자	삼성전자	<a href="http://www">http://www</a>	2021.10.15	naver
위기의 삼	최근 6만	<a href="https://pre">https://pre</a>	2021.10.16	naver
반동장 개	최근 코스	<a href="https://vie">https://vie</a>	2021.10.15	naver
삼성전자	삼성전자	<a href="http://yna">http://yna</a>	2021.10.14	naver
삼성전자,	기사내용	<a href="http://www">http://www</a>	2021.10.12	naver
삼성전자	삼성전자의	<a href="http://www">http://www</a>	2021.10.14	naver
출매만 5	분삼성전자	<a href="https://www">https://www</a>	2021.10.16	naver
대장주 삼	코스피 시	<a href="https://www">https://www</a>	2021.10.15	naver
삼성전자,	목표가 연	<a href="http://www">http://www</a>	2021.10.14	naver
"국민주	여삼성전자의	<a href="http://new">http://new</a>	2021.10.13	naver
삼성전자	삼성전자의	<a href="http://new">http://new</a>	2021.10.15	naver
삼성전자,	삼성전자	<a href="https://vie">https://vie</a>	2021.10.13	naver
삼성전자	!기사내용	<a href="http://www">http://www</a>	2021.10.12	naver
삼성전자,	삼성전자의	<a href="http://new">http://new</a>	2021.10.15	naver
"동학개미	15일 오후	<a href="http://www">http://www</a>	2021.10.15	naver
삼성전자	삼성전자	<a href="http://www">http://www</a>	2021.10.12	naver
미국발	훈: 6만전자	<a href="https://www">https://www</a>	2021.10.15	naver
코스피 30	국내 증시	<a href="https://www">https://www</a>	2021.10.15	naver
"6만전자	대장주 삼	<a href="http://www">http://www</a>	2021.10.13	naver
코스피 3,0	2,900선으	<a href="https://lvr">https://lvr</a>	2021.10.15	naver
미워도 삼	코스피 시	<a href="https://www">https://www</a>	2021.10.15	naver
코스피 0.8	구스피 대	<a href="https://www">https://www</a>	2021.10.15	naver

crawling machine

Engine

☐ Naver ☐ Google

Operator

☐ AND ☐ OR ☐ None

If you choice AND or OR

Please separate the two words by ','(comma)

Search

# Trouble Shooting

## Trouble

각 종목에 대한 기사 일자가 한정됨  
최대 3개월 남짓

Urlopen 모듈 사용 제한  
Header 값을 넣어도 사이트에 접속 불가

Selenium 사용 시 중간에  
끊김 현상 발생  
일정 개수가 넘어갈 때 데이터를  
못받아오는 경우 발생

Naver 증권의 기사 일자 제한

Investing.com의 urlopen 접속 불가

Investing.com 크롤링 제한

## Solution

네이버 검색으로 원하는  
일자 기사 검색

Urlopen 접근이 아니라  
Selenium으로 접근

time.sleep 이용과  
불필요한 정보는 최대한  
수집하지 않기

# Next Process

## 감정분석

### Unsupervised-Learning

VADER  
sentiwordnet

해당 회사에 대한 기사의  
긍정부정 확률 반환

비지도 학습 기반 긍부정  
예측

## 감정분석

### Supervised-Learning

KoBERT

해당 회사에 대한 기사의  
긍정부정 확률 반환

AI-HUB 데이터셋을 학습

지도 학습 기반 긍부정  
예측

## Prediction

### ML or DL

LGBM, LSTM

해당 주가 상승 하락 예측

시계열 순서를 고려하지  
않은 Classification

시계열 순서를 고려한  
Classification

