

3팀

머신러닝 / 딥러닝을 통한 주가 예측

1차 목표 :

KODEX 200 -> 코스피200 선물지수를 추종하는 ETF 상품 ( 개인이 투자 가능 / 접근성 좋음)

투자 관련 여러 지표들을 머신러닝 / 딥러닝을 학습시켜 다음날의 KODEX 200 방향성을 예측 ->  
0 : 하락 / 1 : 상승 -> 투자 수단으로 활용하거나, 적어도 마켓 타이밍에 도움이 되도록

개괄 : 지표 학습 모델 -> 스택킹으로 더 나은 모델 개발 -> 활용(웹 등)

1. 차트 지표 (Boosting)
2. 봉 차트 이미지 (CNN)
3. 거래주체(SVC, LSTM)
4. 시계열(LSTM)
5. 자연어(LSTM)

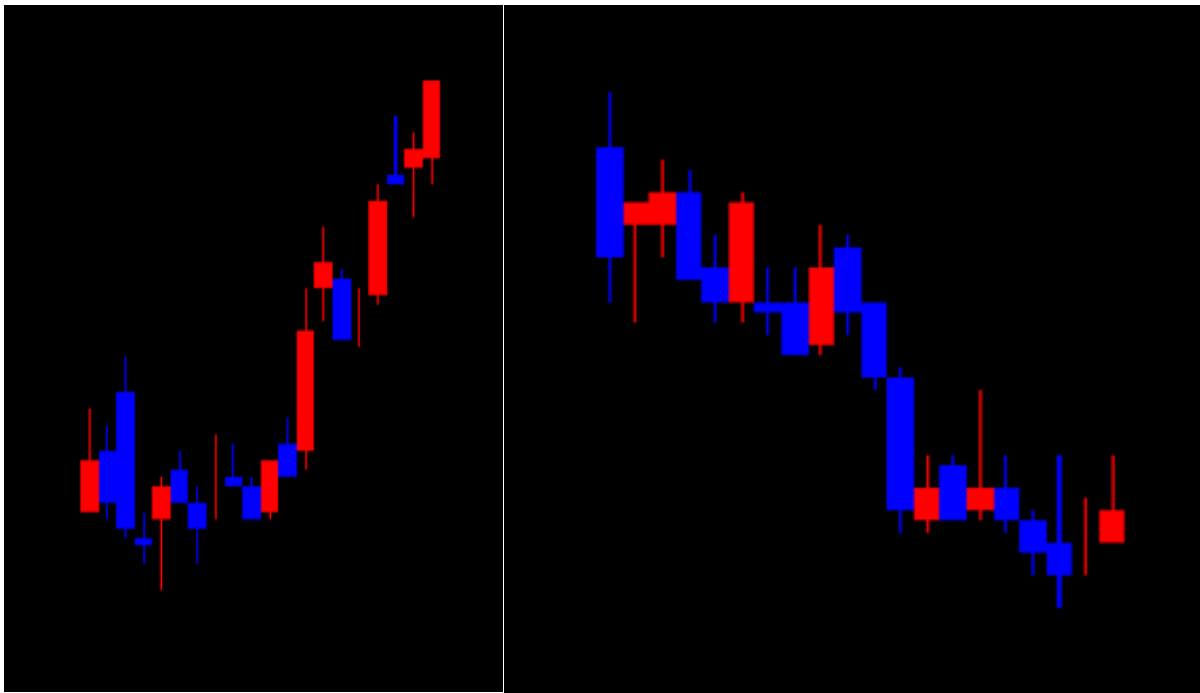
## 1. 차트 지표

	Open	High	Low	Close	Volume	Change	RSI	STOCASTIC_K	STOCASTIC_D	Bollinger ...	RSI_delta	D_delta	sto_diff	B_delta	MACD_delta	MA5	MA20	MA5_adj	MA20_adj	MA_diff	
Date																					
2020-12-16	46040.0	46120.0	46040.0	46120.0	58.0	0.001303	0.669643	0.955556	0.788730	0.879540	-	0.002976	0.083413	0.166826	0.103103	3.230736	46061.0	45962.75	-0.001279	-0.003410	46060.003410
2020-12-17	46180.0	46300.0	46180.0	46295.0	16.0	0.003794	0.753425	0.985915	0.854458	1.136869	-	0.083782	0.065729	0.131457	0.257330	15.510671	46113.0	45986.50	-0.003931	-0.006664	46112.006664
2020-12-18	46275.0	46280.0	46195.0	46195.0	42.0	-0.002160	0.611111	0.704225	0.804381	0.883646	-	-0.142314	-0.050078	-0.100155	-0.253224	3.017621	46144.0	46004.25	-0.001104	-0.004129	46143.004129
2020-12-21	46190.0	46235.0	46070.0	46070.0	236.0	-0.002706	0.571429	0.352113	0.653625	0.613213	-	-0.039683	-0.150756	-0.301512	-0.270433	-8.799072	46148.0	46021.50	0.001693	-0.001053	46147.001053
2020-12-22	46075.0	46080.0	46000.0	46000.0	102.0	-0.001519	0.455782	0.154930	0.487393	0.414248	-	-0.115646	-0.166232	-0.332463	-0.198964	-13.569425	46136.0	46032.00	0.002957	0.000696	46134.999304
2020-12-23	45970.0	45990.0	45935.0	45980.0	230.0	-0.000435	0.436242	0.123288	0.366025	0.335594	-	-0.019541	-0.121368	-0.242737	-0.078655	-13.163641	46108.0	46037.25	0.002784	0.001245	46106.998755
2020-12-24	46010.0	46115.0	46010.0	46080.0	308.0	0.002175	0.574324	0.397260	0.376436	0.605040	-	0.138083	0.010412	0.020824	0.269446	-3.123056	46065.0	46044.75	-0.000326	-0.000765	46064.000765
2020-12-28	46140.0	46145.0	46095.0	46145.0	8.0	0.001411	0.608696	0.575342	0.442738	0.772526	-	0.034371	0.066302	0.132604	0.167486	1.987067	46055.0	46055.75	-0.001950	-0.001934	46054.001934
2020-12-29	46085.0	46180.0	46085.0	46135.0	369.0	-0.000217	0.598765	0.547945	0.477807	0.722330	-	-0.009930	0.035069	0.070138	-0.050195	-0.014826	46068.0	46060.75	-0.001452	-0.001609	46067.001609
2020-12-30	46310.0	46310.0	46250.0	46270.0	251.0	0.002926	0.648649	0.893333	0.616316	1.028113	-	0.049883	0.138509	0.277017	0.305783	9.983859	46122.0	46076.25	-0.003199	-0.004187	46121.004187



Catboost -> 최근 100일 -> ACC : 57%

2. 봉차트 이미지



일봉 기준 5일, 10일, 20일 이미지 데이터 사용 ; 다음날의 주가 등락을 라벨

CNN 이미지 학습

하드웨어 이슈

Table 2: Our proposed CNN architecture.

Input	Conv2D-32 ReLU	max-pooling	Conv2D-48 ReLU	max-pooling	Dropout	Conv2D-64 ReLU	max-pooling	Conv2D-96 ReLU	max-pooling	Dropout	Flatten	Dense-256	Dropout	Dense-2
-------	----------------	-------------	----------------	-------------	---------	----------------	-------------	----------------	-------------	---------	---------	-----------	---------	---------

	Classifier	Period	Dimension	Sensitivity	Specitivity	Accuracy	MCC
with volume	CNN	5	50	83.2	83.8	83.5	0.67
	CNN	10	50	88.6	87.3	88.0	0.758
	CNN	20	50	<b>91.6</b>	<b>91.3</b>	<b>91.5</b>	<b>0.827</b>
	CNN	5	20	83.9	82.7	83.3	0.666
	Random Forest	10	20	87.0	88.3	87.6	0.751
	CNN	20	20	90.8	90.2	90.6	0.808
without volume	CNN	5	50	83.6	85.1	84.4	0.687
	CNN	10	50	89.2	88.1	88.7	0.773
	CNN	20	50	<b>93.3</b>	90.7	<b>92.2</b>	<b>0.84</b>
	CNN	5	20	84.8	83.0	83.9	0.678
	CNN	10	20	88.0	88.2	88.1	0.761
	CNN	20	20	81.7	91.4	91.0	0.817

	Classifier	Period	Dimension	Sensitivity	Specitivity	Accuracy	MCC
with volume	ResNet50	5	50	80.7	85.4	83.1	0.661
	ResNet50	10	50	88.6	88.4	88.5	0.77
	CNN	20	50	<b>90.0</b>	<b>90.1</b>	<b>90.0</b>	<b>0.798</b>
	ResNet50	5	20	78.8	82.3	80.6	0.612
	CNN	10	20	83.3	85.4	84.3	0.686
	CNN	20	20	89.1	84.6	87.1	0.738
without volume	ResNet50	5	50	79.1	87.9	83.3	0.671
	CNN	10	50	87.5	86.6	87.1	0.74
	CNN	20	50	<b>92.1</b>	92.1	<b>92.1</b>	<b>0.837</b>
	CNN	5	20	83.4	82.4	82.9	0.658
	CNN	10	20	85.4	85.6	85.5	0.708
	VGG16	20	20	91.5	89.7	90.7	0.808

Stock Data	Training Data		Testing Data		Independent Data	
	Start	End	Start	End	Start	End
TW50	2000/01/01	2016/12/31	2017/01/01	2018/06/14	2017/01/01	2018/06/14
ID10	2000/01/01	2016/12/31	2017/01/01	2018/06/14	2017/01/01	2018/06/14

### 3. 거래주체

흔히 수급이라고 불리는 '거래주체'

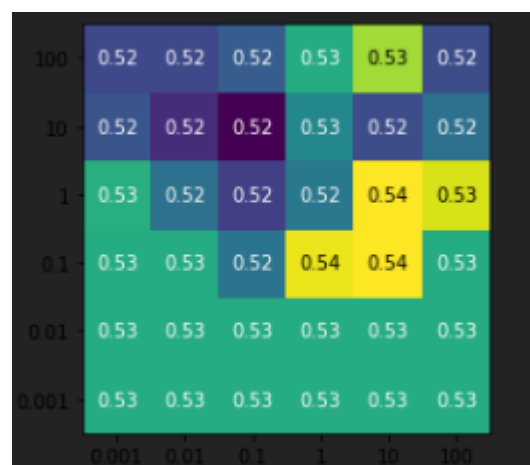
거래주체와 지수의 등락에는 명백한 상관관계가 있음

이는 , 다음날 지수의 흐름에도 영향을 줄 것.

일자	지수	등락폭	개인	외국인	기관
2022/05/24	2,621.47	▼ 25.91	3,937	-2,345	-1,707
2022/05/23	2,647.38	▲ 8.09	-1,652	-303	1,697
2022/05/20	2,639.29	▲ 46.95	-10,412	1,966	8,364
2022/05/19	2,592.34	▼ 33.64	5,167	-909	-4,828
2022/05/18	2,625.98	▲ 5.54	259	2,972	-3,495
2022/05/17	2,620.44	▲ 23.86	-3,082	-524	5,408
2022/05/16	2,596.58	▼ 7.66	650	17	-1,158
2022/05/13	2,604.24	▲ 54.16	-9,081	684	8,617
2022/05/12	2,550.08	▼ 42.19	3,908	-2,820	-1,575
2022/05/11	2,592.27	▼ 4.29	3,452	-2,899	-892
2022/05/10	2,596.56	▼ 14.25	2,856	-3,177	75
2022/05/09	2,610.81	▼ 33.70	3,380	-1,977	-1,767
2022/05/06	2,644.51	▼ 33.06	7,644	-4,863	-3,001
2022/05/04	2,677.57	▼ 2.89	1,009	306	-1,501
2022/05/03	2,680.46	▼ 6.99	1,041	1,150	-2,140

모델 프레임 짜는중

하이퍼파라미팅에 따른 테스트 결과 ; 외국인 콜옵션 순매수 만을 고려 , 정확도 상승 여력



#### 4. 자연어

Title	Date
0 [Asia마감] 中 '나홀로 돈풀기'에 환호...상하이 1.6% ↑	2022년05월20일(금)
1 [마켓뷰] 코스피, 국내외 기관 매수에 1.8% 상승 마감...개인 1조 '팔자'	2022년05월20일(금)
2 [외환마감]강달러 진정+위안화 강세...환율, 10원 가량 내려 1260원대	2022년05월20일(금)
3 [마감시황]코스피, 하루 만에 2600선 회복...'바이든 효과'	2022년05월20일(금)
4 [시황종합] 코스피, 한미정상회담 수혜 기대감에 상승...2639.29	2022년05월20일(금)
...	...
0 (마켓서핑)큰 고비 넘었다	2005년12월08일(목)
0 (마켓서핑)삼성전자 날다	2005년12월07일(수)
1 [내일의 전략]5년전 악몽이 떠오른다	2005년12월07일(수)
0 (마켓서핑)이유있는 불안감	2005년12월06일(화)
0 (마켓서핑)외국인 돌아오다	2005년12월05일(월)

인터넷의 정보(특히 뉴스기사) -> 오늘의 시장분위기 반영 -> 내일의 투심 / 방향을 예측할 수 있다

Date	new	encode	label	percent
2022년05월20일(금)	中 '나홀로 돈풀기'에 환호...상하이 1.6% ↑	([[[2, 1370, 7932, 6079, 1730, 7747, 5579, 5136...	부정	71.32
2022년05월20일(금)	코스피, 국내외 기관 매수에 1.8% 상승 마감...개인 1조 '팔자'	([[[2, 4672, 1140, 1261, 1994, 6896, 547, 2670,...	부정	82.65
2022년05월20일(금)	강달러 진정+위안화 강세...환율, 10원 가량 내려 1260원대	([[[2, 807, 5795, 4360, 7227, 7050, 7941, 814, ...	부정	60.08
2022년05월20일(금)	코스피, 하루 만에 2600선 회복...'바이든 효과'	([[[2, 4672, 4937, 1946, 553, 191, 6559, 5155, ...	긍정	70.98
2022년05월20일(금)	코스피, 한미정상회담 수혜 기대감에 상승...2639.29	([[[2, 4672, 4955, 6255, 7227, 6527, 7954, 2910...	긍정	75.46
2022년05월20일(금)	바이든 방한 훈풍에 1.8%대 상승...880선 눈앞	([[[2, 2186, 7096, 5928, 2267, 7828, 517, 7970,...	긍정	79.92
2022년05월20일(금)	기관 매수세에 코스피 2600선 회복...개인은 '팔자'	([[[2, 1261, 1994, 6579, 6896, 4672, 553, 191, ...	부정	52.53
2022년05월20일(금)	美 증시 하락 마감...코스피 기술 리오프닝株 중심 반등할까	([[[2, 4294, 4933, 1908, 7537, 1289, 6122, 6964...	부정	95.60
2022년05월20일(금)	102선으로 떨어진 달러인덱스...환율, 1260원대 하락 시도	([[[2, 533, 119, 6559, 7078, 1860, 1599, 7119, ...	부정	96.06
2022년05월20일(금)	뉴욕증시, 물가경기 우려에 하락...바이든 오늘 방한	([[[2, 1543, 7316, 6705, 2136, 5425, 3499, 6896...	부정	94.81
2022년05월20일(금)	해외 주식형 펀드, 3거래일간 1900억원 순유입	([[[2, 5012, 4230, 7921, 4828, 589, 5381, 5337,...	부정	72.38
2022년05월20일(금)	'약세장 문턱' 美증시 '흔들'...'실적실망' 시스코 13% ↓	([[[2, 3212, 7178, 2120, 7611, 4294, 5199, 6742...	부정	99.45
2022년05월20일(금)	계속되는 인플레이션 우려로 소크...나스닥 -0.26%	([[[2, 984, 5889, 3758, 7761, 6050, 6639, 5341,...	부정	99.16
2022년05월20일(금)	가상화폐시장 뒤흔든 '테라·루나 사태' 후폭풍 어디까지	([[[2, 517, 5330, 6527, 7941, 7727, 6716, 1783,...	부정	93.10
2022년05월20일(금)	인플레이션 경계심에 공포에 하락	([[[2, 3758, 7761, 6050, 6295, 961, 1054, 6896,...	부정	96.86
2022년05월19일(목)	美 따라 녹아내렸다...일본·홍콩 급락	([[[2, 1835, 1494, 6797, 5678, 6074, 7131, 7939...	부정	99.35
2022년05월19일(목)	美 급락 여파에 코스피 다시 2600선 밑...2차 전자株는 강세	([[[2, 1253, 3298, 7682, 6896, 4672, 1574, 553,...	부정	96.86
2022년05월19일(목)	코스피, 기관 매도에 1.3% 하락...2590선 마감	([[[2, 4672, 1261, 1990, 6896, 541, 4933, 137, ...	부정	99.16
2022년05월19일(목)	코스피 3일 만에 2600 무너져... 오후 낙폭 축소 2592 마감	([[[2, 4672, 589, 7126, 1946, 553, 191, 2095, 5...	부정	98.69
2022년05월19일(목)	강달러에 국내증시 하락...환율, 11원 이상 올라 1270원대	([[[2, 807, 5795, 6896, 1138, 7316, 6705, 4933,...	부정	98.86

라벨 / percent 데이터 -> LSTM을 통한 주가 방향 예측 모델(초안)

5. 시계열 – 오직 가격만을 변수로



6.

```
4/4 [=====] - 2s 11ms/step - loss: 0.6878 - accuracy: 0.5900  
Cross-entropy : 0.687831699848175  
Accuracy : 0.5899999737739563
```