

머신러닝을 이용한 주가예측

모멘텀 투자에 다양한 기법 추가 적용

머신러닝을 이용한 주가예측

1. 프로젝트 개요 - 모멘텀 투자

2. 재무제표, 경제지표 반영 주가예측

3. 뉴스 NLP 반영 주가예측

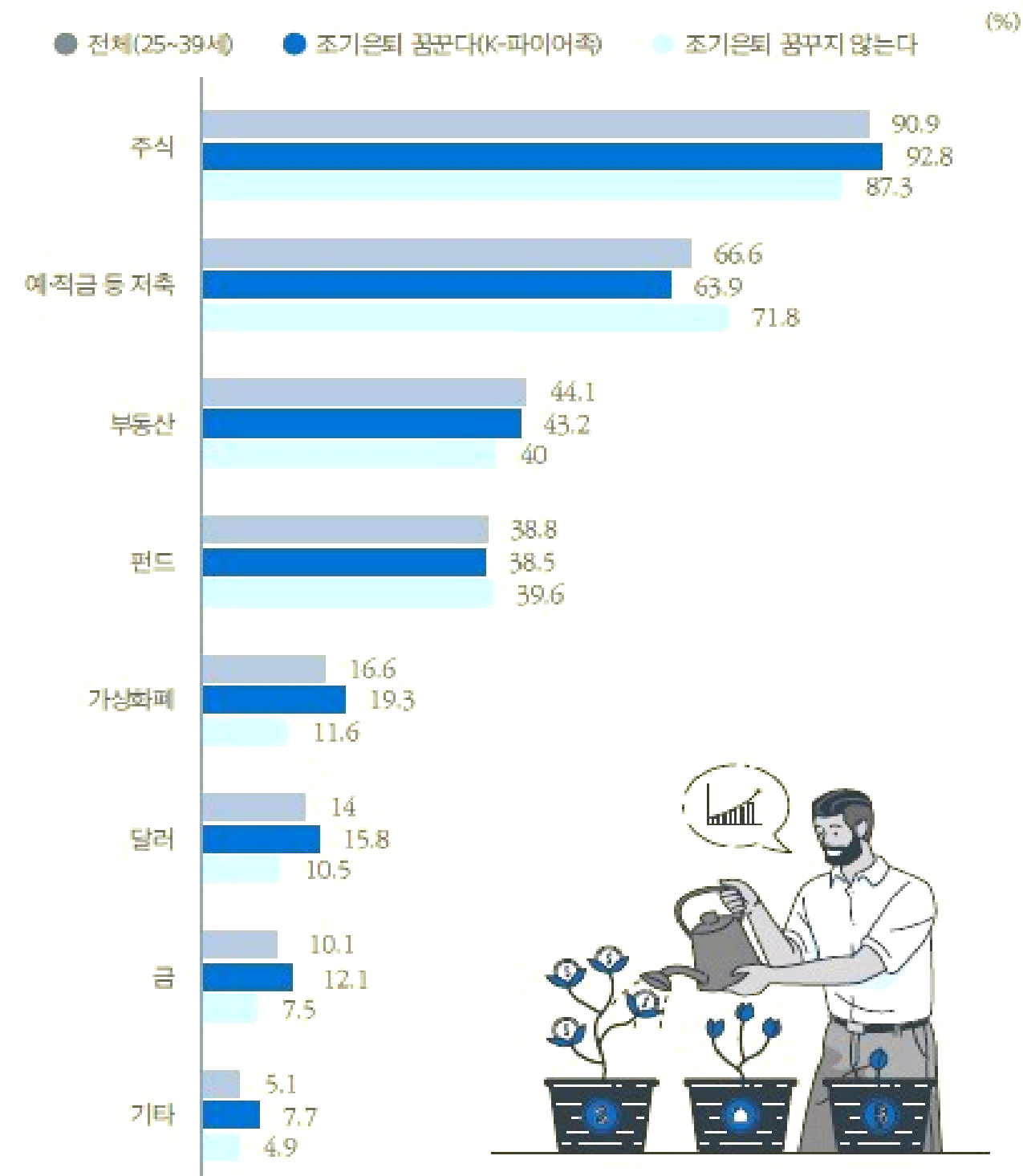
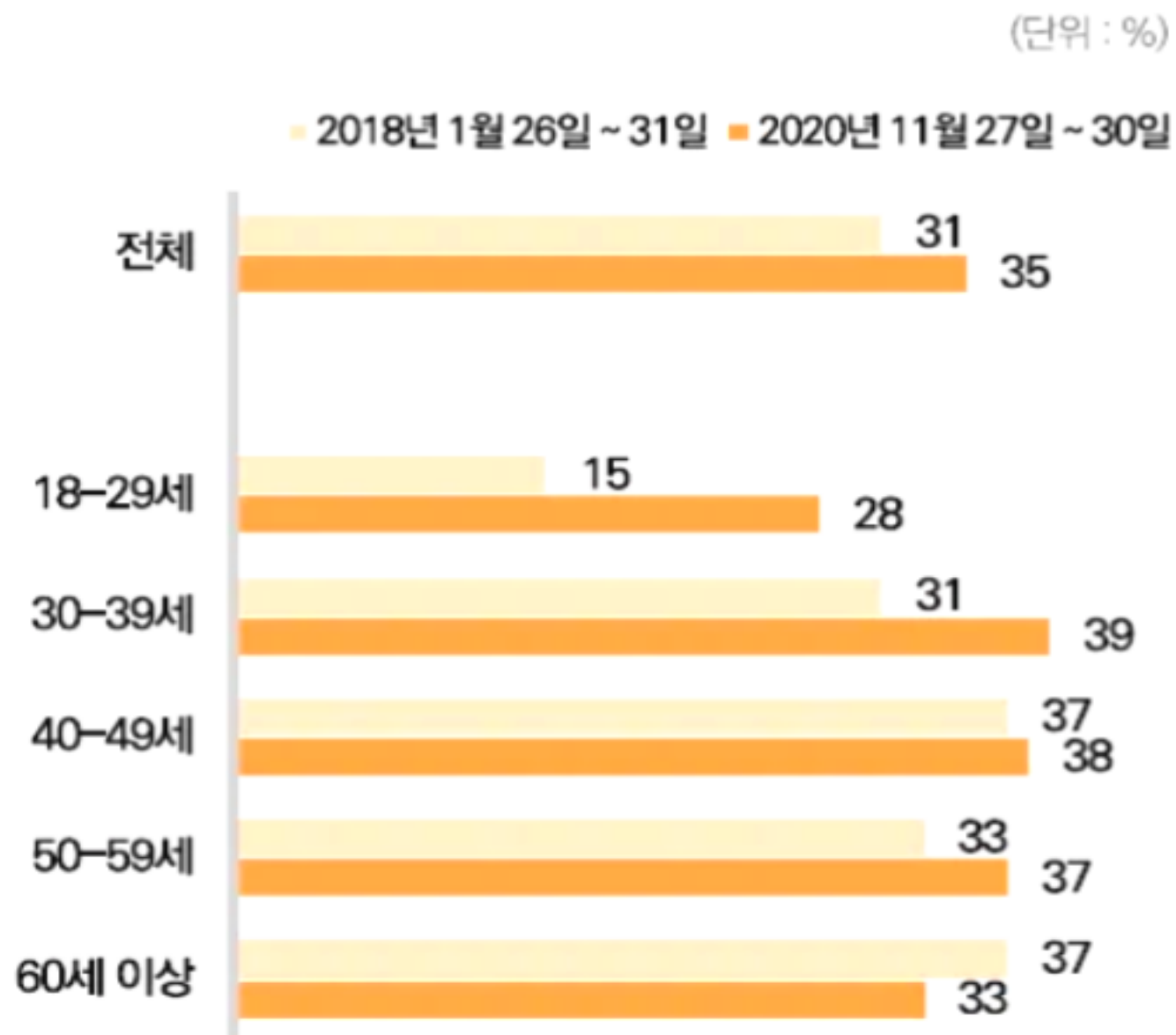
4. 모의투자 결론 및 시사점

주제 설정 배경

■ 코로나 이후 2030세대의 주식에 대한 관심 증가

연령별 주식 투자자 비율

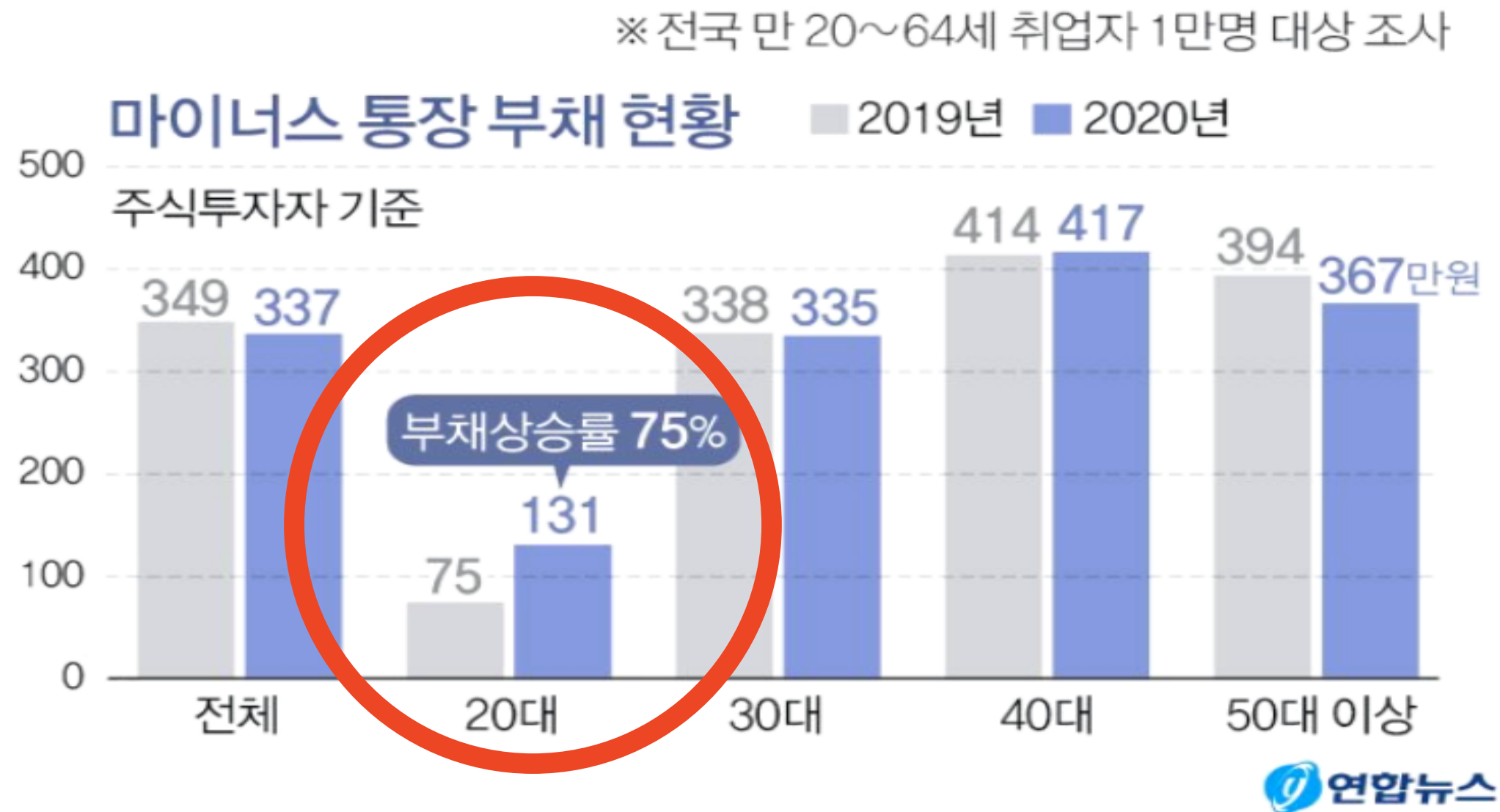
18-29세에서 13%p, 30대에서 8%p 상승



주제 설정 배경

■ 연령별 부채 현황

20대는 부채 상승률이 높아진것을 보아 공격적인 투자를 하고 있음



프로젝트 목표

재무제표, 경제지표, NLP를 활용하여 기존 모멘텀 투자보다 발전된 투자기법 제시



재무재표/경제지표에 머신러닝 기법을 활용하여 주가 예측



NLP로 뉴스가 회사 주가에 어떠한 영향을 미치는지 분석



빅데이터를 활용한 투자방식과 기존투자방식을 비교



■ 주가 추세의 속도가 증가하고 있는지, 감소하고 있는지를 추세 운동량으로 측정하여 나타내는 지표

■ 상승장에서 추세 가속도가 둔화된다면 매도하고, 하락장에서 추세 가속도가 증가하면 매수하는 전략을 취할 수 있음



재무제표

기업의 재무제표는 기업의 재무상태와 경영 성과 등을 담고 있기 때문에 주가와 긴밀한 상관관계를 가짐

+



거시경제지표

이자율, 국민소득, 환율 등의 경제지표는 주식시장에 직접적인 영향을 미치고 상관관계가 있기 때문에 주식의 매수, 매도를 판단하는 주요자료이다

+



NLP

NLP란 자연어(인간의 언어)를 컴퓨터가 처리하게 할 수 있는 기술
NLP 분석을 통해 언어의 패턴과 특성을 파악해 분류, 회귀등의 머신러닝에 사용 가능

재무제표 및 거시경제지표

재무·경제지표를 활용한 주가예측 및 모멘텀 투자방법



재무제표

기업의 재무제표는 기업의 재무상태와 경영 성과 등을 담고 있기 때문에 주가와 긴밀한 상관관계를 가짐

+



거시경제지표

이자율, 국민소득, 환율 등의 경제지표는 주식시장에 직접적인 영향을 미치고 상관관계가 있기 때문에 주식의 매수,매도를 판단하는 주요자료이다

+



NLP

NLP란 자연어(인간의 언어)를 컴퓨터가 처리하게 할 수 있는 기술
NLP 분석을 통해 언어의 패턴과 특성을 파악해 분류, 회귀등의 머신러닝에 사용 가능

STEP 01

재무·경제지표 산출

- 재무지표
기업의 수익성, 성장성, 건전성, 유동성 등을 나타내는 변수 21개 정의
- 경제지표
환율, CD금리, GDP상승률 등 거시경제를 대표하는 변수 5개 정의

STEP 02

주가 예측

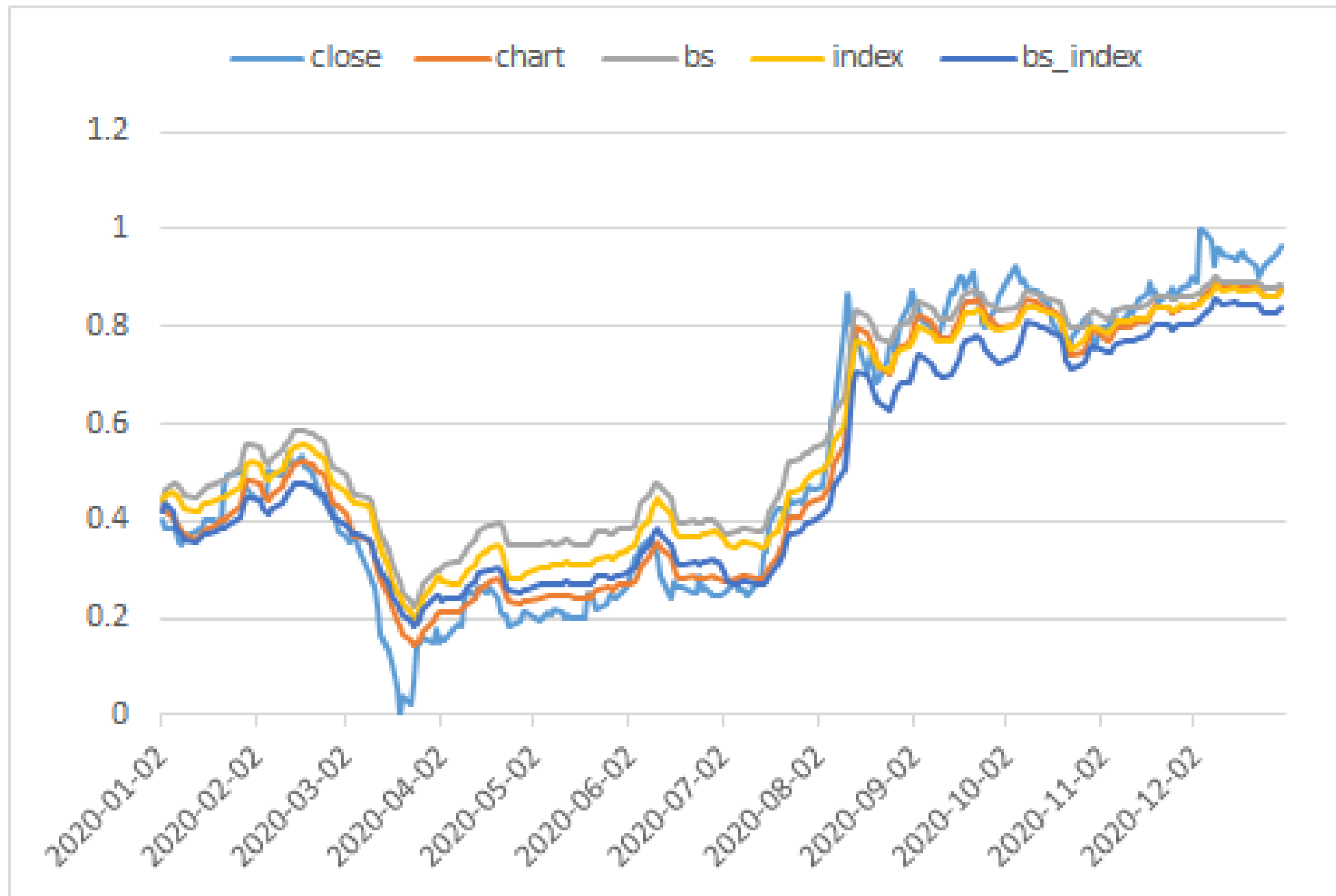
- 전처리한 재무·경제지표와 주가 데이터를 일별로 통합
- 로지스틱 회귀분석, DecisionTree, RandomForest, DNN, LSTM 등의 기법을 활용하여 주가 예측

STEP 03

결과 도출 및 투자 분석

- 개별 기업 및 분석대상 전체 기업의 예측 정확도와 시나리오별 특성 파악
- 모멘텀 투자와 결합해 주가변동에 따른 투자수익률 분석

딥러닝(LSTM) 기반 주가 예측 결과



- 주가정보만 활용한 경우와 재무·경제지표를 추가로 활용한 경우 모두 실제 주가를 잘 추정함
- bs_index의 경우 변동폭이 적은 보다 스무스한 움직임을 보임

딥러닝(DNN) 모델의 예측 정확도 분석

- 예측 정확도가 높은 회사들을 살펴본 결과, 한솔케미칼, 오뚜기, bgf리테일 등 화학 및 소비재 업종의 정확도가 높음
- 화학, 소비재 업종은 매출 발생시 바로 수익으로 인식하기 때문에 재무경제지표들과 연관성이 높은 것으로 추정됨

	code	company	DNN
0	14680	한솔케미칼	0.819328
1	7310	오뚜기	0.785714
2	282330	BGF리테일	0.781513
3	11780	금호석유	0.764706
4	2790	아모레G	0.764706
5	6650	대한유화	0.760504
6	90430	아모레퍼시픽	0.752101
7	670	영풍	0.714286
8	11170	롯데케미칼	0.710084
9	192820	코스맥스	0.684874
10	6260	LS	0.672269
11	139480	이마트	0.663866
12	4370	농심	0.655462
13	3240	태광산업	0.655462
14	5250	녹십자홀딩스	0.65126
15	52690	한전기술	0.62605

시가총액 기준 TOP10 기업 예측력 분석

	company	chart_RMSE	bs_RMSE	index_RMSE	nlp_RMSE	bs_index_RMSE	bs_nlp_RMSE	index_nlp_RMSE	bs_index_nlp_RMSE
0	삼성전자	0.0062	0.0050	0.0163	0.0068	0.0131	0.0049	0.0048	0.0047
1	SK하이닉스	0.2046	0.2862	0.1976	0.2195	0.2258	0.2498	0.1960	0.2205
2	LG화학	0.3943	0.3639	0.3864	0.3745	0.3696	0.3863	0.3583	0.3660
3	현대차	0.3635	0.3521	0.3448	0.3044	0.3417	0.2887	0.3104	0.2916
4	삼성바이오	0.3040	0.3106	0.3271	0.3163	0.3345	0.2990	0.3221	0.3244
5	카카오	0.3724	0.3857	0.3718	0.4996	0.3782	0.3984	0.3730	0.3901
6	삼성SDI	0.3295	0.2866	0.3009	0.3037	0.2784	0.2788	0.2690	0.2648
7	기아차	0.3085	0.2827	0.2893	0.2716	0.2886	0.2512	0.2681	0.2510
8	POSCO	0.1476	0.1656	0.1637	0.1759	0.1610	0.1437	0.1961	0.1460
9	현대모비스	0.2768	0.2976	0.2826	0.2751	0.2908	0.2848	0.2716	0.2803

- 평균적으로 차트만 활용한 기술적 분석보다 지표들을 활용한 분석이 예측력이 좋음

- 지표에 NLP까지 추가적으로 반영한 경우 평균 예측력이 가장 좋음



재무제표

기업의 재무제표는 기업의 재무상태와 경영 성과 등을 담고 있기 때문에 주가와 긴밀한 상관관계를 가짐



거시경제지표

이자율, 국민소득, 환율 등의 경제지표는 주식시장에 직접적인 영향을 미치고 상관관계가 있기 때문에 주식의 매수,매도를 판단하는 주요자료이다



NLP

NLP란 자연어(인간의 언어)를 컴퓨터가 처리하게 할 수 있는 기술
NLP 분석을 통해 언어의 패턴과 특성을 파악해 분류, 회귀등의 머신러닝에 사용 가능

STEP 01

뉴스와 주가데이터
크롤링 및 전처리

■ 뉴스 데이터 |

- 3개 경제 언론사의 2018.1.1~ 2021.3.31 크롤링
- 형태소 단위의 TDM으로 벡터화 및 불용어 제거

■ 주가 데이터 |

- 극성부여 : 어제 대비 오늘의 주가 상승시 1, 아닌 경우 0
- 전처리된 뉴스 데이터와 주가 데이터의 극성을 날짜기준으로 통합

STEP 02

머신러닝 분류 모델
주가예측

- 날짜 기준 TDM과 처리된 주가 데이터를 Logistic Regression와 Random Forest, Cosine Similarity를 이용해 내일의 주가 예측
- 그리드 서치를 이용한 교차검증과 최적 파라미터 적용
- 각 모델이 뉴스의 내용을 보고 예측한 극성과 실제 극성을 비교
- 혼동 행렬을 이용한 정확도를 도출

STEP 03

주가변동과 상관관계가
있는 형태소 추출

- TDM을 통해 주가 변동과 상관관계가 있는 형태소 파악
- 양의 상관관계가 있는 형태소, 음의 상관관계가 있는 형태소 리스트화
- 강한 상관관계가 있는 단어 실제상황과 비교 해석

분석 결과 | 머신러닝 분류 모델 주가예측

로지스틱 회귀의 정확도	랜덤포레스트의 정확도	코사인 유사도의 정확도
평균 : 0.5647	평균 : 0.5504	평균 : 0.5356

- 로지스틱 회귀의 정확도가 가장 좋고, 코사인유사도가 가장 낮은 정확도를 가짐
- 뉴스만으로 주가를 보편적으로 예측하는데 한계가 존재

분석 결과 | 머신러닝 분류 모델 주가예측

- 정확도가 최상위, 최하위 기업은 뉴스가 적은 기업이 대부분.
※ 충분한 기사량이 있어야 NLP분석의 의의가 있다.
- 기업별 각 모델의 정확도는 약간의 차이 존재.
- 코사인 유사도 정확도 0.6이상 상위 기업 30개중_
15개 소비재(경기 소비재, 생활 소비재) 비중 50%

	company	code	ras_cos
140	세방전지	004490	1.000000
125	오리온홀딩스	001800	1.000000
100	코오롱인더	120110	0.791667
105	한국콜마	161890	0.756494
141	삼양홀딩스	000070	0.750000
64	DB손해보험	005830	0.741830
110	영진약품	003520	0.708333
73	OCI	010060	0.700549
69	일진머티리얼즈	020150	0.700000
75	한화생명	088350	0.700000
70	BGF리테일	282330	0.689394
26	한온시스템	018880	0.675000
84	현대위아	011210	0.665441
121	한성	020000	0.662879
21	삼성에스디에스	018260	0.634921
92	영원무역	111770	0.633333
18	신한지주	055550	0.629870
126	대상	001680	0.625000
55	롯데쇼핑	023530	0.625000
129	두산	000150	0.625000
4	현대차	005380	0.624239
35	강원랜드	035250	0.623656
61	호텔신라	008770	0.623355
133	LG하우시스	108670	0.614583
47	SKC	011790	0.610248
34	현대제철	004020	0.610000
33	CJ제일제당	097950	0.607328
57	들서	026960	0.607097
108	한올바이오파마	009420	0.604167
62	마도	204320	0.600000

목 표 | 주가변동과 상관관계가 있는 형태소 리스트 추출

- 주가와 TDM 형태소의 상관관계를 추출

상관계수	상관관계 정도
± .9 이상	상관관계가 아주 높다
± .7~.9미만	상관관계가 높다
± .4~.7미만	상관관계가 다소 높다
± .2~.4미만	상관관계가 있느냐 낮다
± .2미만	상관관계가 거의 없다

- 주가와 상관관계가 강한 형태소 추출

이를 통해 사후적으로 NLP 분석의 설명력을 파악가능

농심

양의 상관관계	음의 상관관계
기생충	꽃새우
실시간	군산
전문	어민
풀이	품질
알고리즘	대도시

목 표 | 주가변동과 상관관계가 있는 형태소 리스트 추출

"계획이 통했구나"... 짜파게티, 기생충 특수에 해외 매출 120% '깡충'

이선목 기자

입력 2020.03.18 14:19

농심이 영화 ‘기생충’ 열풍의 특수를 누리고 있다. **농심** (279,500원 ▲ 500 0.18%)은 지난달 짜파게티의 해외 매출이 지난해 동기 대비 2배(120%) 이상 증가한 150만달러(약 18억6000만원)를 기록했다고 18일 밝혔다. 월간 기준 최대 실적이다.

"서해 오염탕 군산 꽃새우 안써" 농심 설명에 군산 민심 '부글'

송고시간 | 2019-07-29 15:32

김동철 기자

기자 페이지

군산 어민들 "서해 모든 생선류가 환경오염됐다는 인식 심어줘"

(군산=연합뉴스) 김동철 기자 = 국민 과자 새우깡의 원료를 군산 꽃새우에서 수입산으로 바꾼 농심의 결정으로 직격탄을 맞은 전북 군산 어민들과 지역 정치권이 뿔났다.

다른 기업의 경우에도 상관관계 강한 형태소는
그 기업의 이슈를 잘 설명함

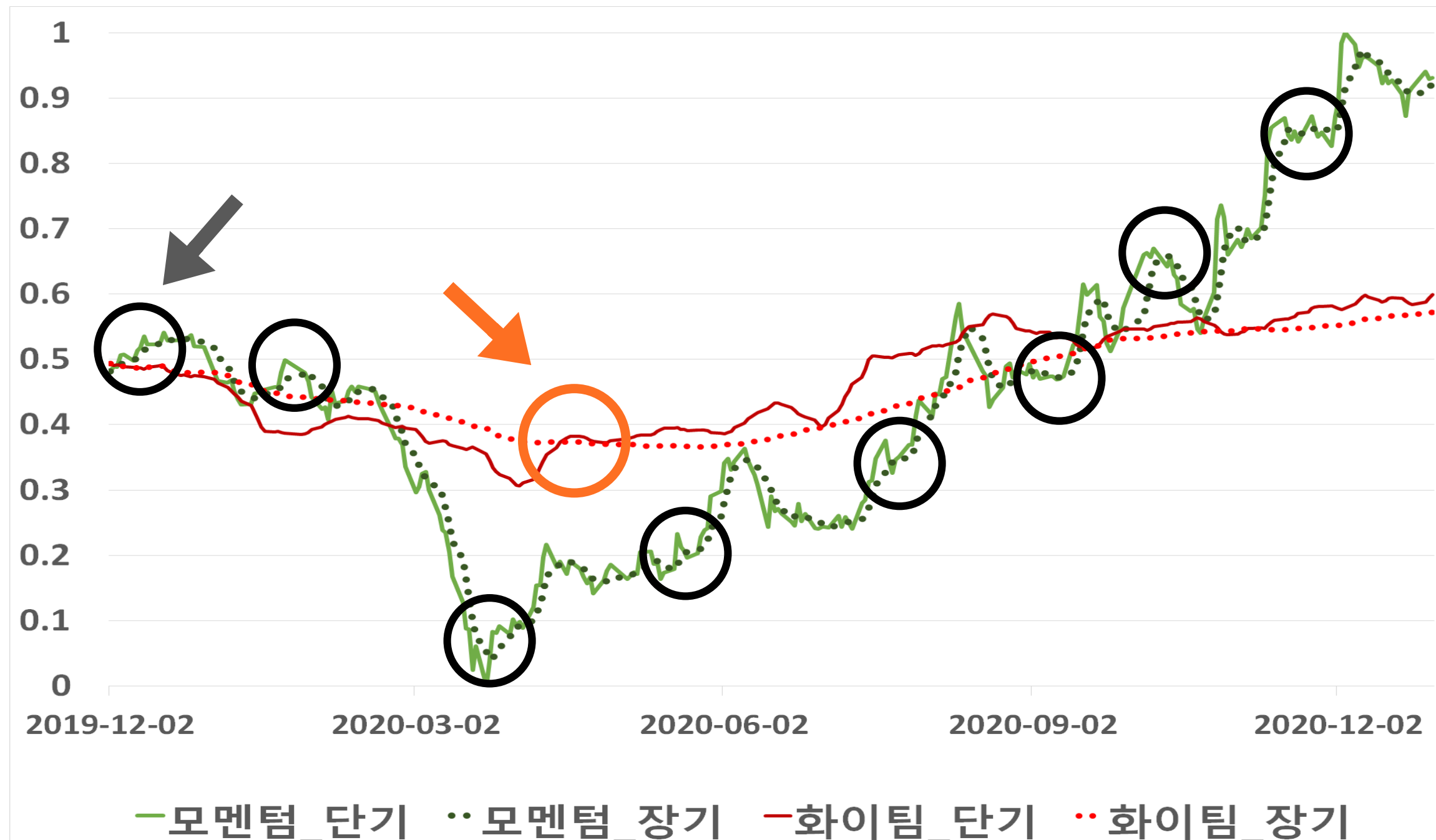
농심	
양의 상관관계	음의 상관관계
기생충	꽃새우
실시간	군산
전문	어민
풀이	품질
알고리즘	대도시

모의투자 모멘텀투자자와 비교



화이팀 투자 전략 |

재무제표, 경제지표, NLP 를 활용한 예측 주가의 이동평균선의 교차점을 매수, 매도 타이밍으로 하여 투자



[기아차 모의투자]

기존 모멘텀투자자 VS 화이팀 모멘텀투자자 수익률 비교 |



모멘텀 투자 : 121%의 수익률
화이팀 투자 : 138%의 수익률

모의 투자 결과 시사점 |



경제지표, 재무지표, NLP는 주가를 예측할 때 정확도를 더 높이는 요소



시가총액 상위 회사 모의투자는 화이팀 투자전략의 수익률 높음



기존 모멘텀 투자에 추가적인 매매 시점을 반영하여 투자방법을 개선 가능

화이팅



감사합니다
