

유튜브 주식 콘텐츠 기반 kospi 지수 분석 및 투자 보조 서비스 제안

채널별 선/후행성 분석을 바탕으로



금융A팀 | 주식전문채널 DARtv

Contents

- I. 목적 및 문제 정의
- II. 연구 진행 방향
- III. 연구 진행 과정
- IV. 인사이트 도출
- V. 한계점

목적 및 문제 정의

최근 유튜브는 주식 투자 정보를 얻는 핵심 채널로 급부상 중



BUT

주식 투자 정보, 많아도 너무 많아...'선택의 역설' 나타나

"100배 오른다"…개미 홀리는 유튜브 리딩방

"유튜브 주식채널이 개인투자자 의사결정 왜곡"



0. 목적 및 문제정의



오늘 시황 어쩌고,
저쩌고 그러합니다.
그래서 내일은 XX 방향
으로 양상이 나타날..

증권사

리스트

경제
유튜버

주식
유튜버



BUT
“유튜브 정보가 시장을 선행하는가?”

"100배 오른다"…개미 홀리는 유튜브 리딩방

"유튜브 주식채널이 개인투자자 의사결정 왜곡"



연구 진행 방향

1

본 탐구 목적에 맞게 대표성을 띠는 주식 관련 채널들을 선정

2

각 채널마다 코스피 관련 재생클립 선정 후 자막 데이터 추출

3

텍스트 전처리 및 감성분석 실행

4

감성지수-시장지수 매핑 후 시차 분석

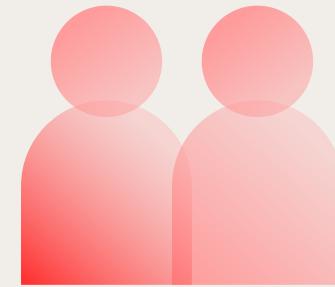
5

기존 매체와 유튜브 경향성 차이 분석

6

감정지수 기반 백테스팅 및 투자 보조 서비스 제안

1. 유튜브 채널 선정



경제 관련 카테고리 中



상위 채널부터 탐색
(구독자수, 조회수 기준)

1. 주식 관련 채널인지
2. 업로드 주기성이 짧은지

선정된 채널 리스트



<3PRO tv>



<한국경제 tv>



<전인구경제연구소>



<MTN 머니투데이방송>



<매일경제 tv>



<E트렌드>



<서울경제tv>



<불사조>



<위너스tv>

2. 스크립트 추출



선정된 채널들 中

다음을 충족하는
재생클립 수집

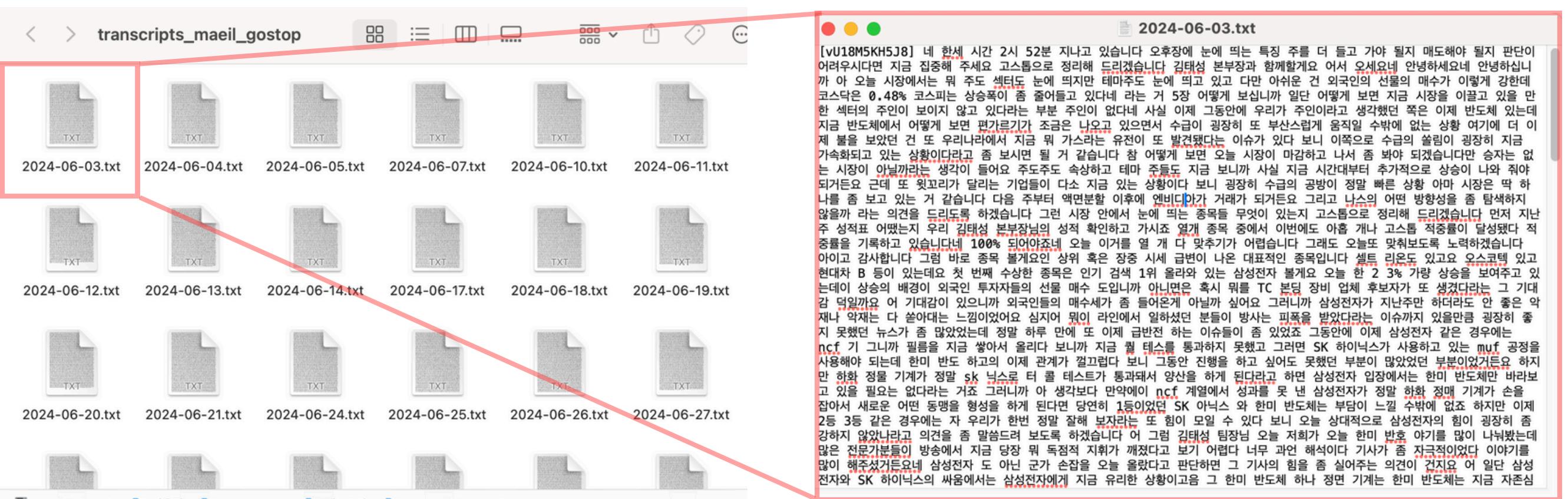
1. 국내 시장을 (주로) 다루는지
2. 수집기간 내 데이터가 충분한지

*수집기간: 24.06.01~24.12.31

스크립트 추출 예시



재생클립
[수상한 종목
GOSTOP]
[투자의 눈]



3. 텍스트 전처리 및 감성분석

1

스크립트

2

문장 단위로 분리

3

512토큰 이하 청크로 나누기

2024-06-03.txt
간 2시 52분 지나고 있습니다 오후장에 눈에 띠는 특징 주를 더 들고 가야
요 고스톱으로 정리해 드리겠습니다 김태성 본부장과 함께할게요 어서 오세요
센터도 눈에 띄지만 테마주도 눈에 띠고 있고 다만 아쉬운 건 외국인의 선물
폭이 좀 줄어들고 있다네 라는 거 5장 어떻게 보십니까 일단 어떻게 보면
있다라는 부분 주인이 없다면 사실 이제 그동안에 우리가 주인이라고 생각했던
평가하기가 조금은 나오고 있으면서 수급이 굉장히 또 부산스럽게 움직일 수밖에
에서 지금 뭐 가스라는 유전이 또 발견됐다는 이유가 있다 보니 이쪽으로 수
좀 보시면 될 거 같습니다 참 어떻게 보면 오늘 시장이 마감하고 나서 좀 봐
들어요 주도주도 속상하고 테마 주들도 지금 보니까 사실 지금 시간대부터 추가
되는 기업들이 다소 지금 있는 상황이다 보니 굉장히 수급의 공방이 정말 빠른
다음 주부터 액면분할 이후에 엔비디아가 거래가 되거든요 그리고 나스의 어
하겠습니다 그런 시장 안에서 눈에 띠는 종목들 무엇이 있는지 고스톱으로 정리
부분장님의 성적 확인하고 가시죠 열개 종목 중에서 이번에도 아홉 개나 고
10% 되어야죠 오늘 이거를 열 개 다 맞추기가 어렵습니다 그래도 오늘도 및
종목 볼게요인 상위 혹은 장중 시세 급변이 나온 대표적인 종목입니다 셀트리
째 상승한 종목은 인기 검색 1위 올라와 있는 삼성전자 불과요 오늘 한 2 3
투자들의 선물 매수 도입니까 아니면 혹시 뭐를 TC 본딩 장비업체 후보자
니까 외국인들의 매수세가 좀 들어온게 아닐까 싶어요 그러니까 삼성전자가 제
되었어요 심지어 뭐이 리인에서 일하셨던 분들이 방사는 피포를 벌았다는 이
게 정말 하루 만에 또 이제 급반전 하는 이슈들이 좀 있었죠 그동안에 이제 삼
아서 올리다 보니까 지금 웰 테스를 통과하지 못했고 그러면 SK 하이닉스가
그의 이제 관계가 껄끄럽다 보니 그동안 진행을 하고 싶어도 못했던 부분이 많
닉스로 터 클 테스트가 통과돼서 양산을 하게 된다라고 하면 삼성전자 입장에
그러니까 생각보다 만약에이 ncf 계열에서 성과를 못 낸 삼성전자가 정말
借此를 하게 된다면 당연히 1등이었던 SK 하이닉스 와 한미 반도체는 부담이 느낄
수가 한번 점말 잘해 보자라는 또 힘이 모일 수 있다 보니 오늘 상대적으로 삼
삼성드려 보도록 하겠습니다 어 그럼 김태성 팀장님 오늘 저희가 오늘 한미 반
금 당장 뭐 독점적 지휘가 깨졌다고 보기 어렵다 너무 과언 해석이다 기사가
도 아닌 군가 손잡을 오늘 올랐다고 판단하면 그 기사의 힘을 좀 실어주는 의
는 삼성전자에게 지금 유리한 상황이고 그 한미 반도체 하나 정면 기계는

✓ 원로: 2024-06-01.txt → 1594 문장
✓ 원로: 2024-06-03.txt → 630 문장
✓ 원로: 2024-06-04.txt → 659 문장
✓ 원로: 2024-06-05.txt → 1025 문장
✓ 원로: 2024-06-07.txt → 658 문장
✓ 원로: 2024-06-08.txt → 2041 문장
✓ 원로: 2024-06-10.txt → 598 문장
✓ 원로: 2024-06-11.txt → 633 문장
✓ 원로: 2024-06-12.txt → 1054 문장
✓ 원로: 2024-06-13.txt → 613 문장
✓ 원로: 2024-06-14.txt → 744 문장
✓ 원로: 2024-06-17.txt → 643 문장
✓ 원로: 2024-06-18.txt → 668 문장
✓ 원로: 2024-06-19.txt → 1012 문장
✓ 원로: 2024-06-20.txt → 632 문장
✓ 원로: 2024-06-21.txt → 595 문장
✓ 원로: 2024-06-22.txt → 2044 문장
✓ 원로: 2024-06-24.txt → 581 문장

청크 1: [_n-I2DHHkEU] [음악] 네 여러분 안녕하십니까 2024년 6월..
청크 2: 글쎄요 이게 끊는 말하긴 쉽니다 그런데 진짜 여러분들한테 도움되는
청크 3: 제가 뭐 뭐 소문은 여러 가지로 외국이 될 수도 있고 뭐 한국에 뭐
청크 4: 그러다가 이제 빛 보는 거고요 막 그래요 근데 어마어마하게 예 봤던
청크 5: 그러니까 제가 왜이 말씀을 드리냐면 hbm이라는이 부분은요 실험 반도
청크 6: 여러분들 이제 미국 시장에서는 AI 대해서 물론 앞으로도 그렇지만
청크 7: 여러분들 우리 신재생 얼마든지 할 수 있어요 공에 하나도 없는 에너
청크 8: 왜냐면 전환이 끝 때문에 전환을 다 이겨내면서 올린 거 보면은 운도
청크 9: 어 하나 오션에 어떤 부분을 반장한테 전화 연결하 물어봐야지 이런
청크 10: 그러면 제가 여기서 잠깐만 우리 수혁이 어머님이 어예 저한테 그
청크 11: 제가 왜이 얘기를 드리냐면 여러분들 이거 어떻게 보면 저도 누워서
청크 12: 지금도 마찬가지예요 대우주 하나 오 왜 샀을까요 하반기에 좋아진
청크 13: 예 수 그런데 제가 30년 가까이 하면서 엔터 터키 못 쳐서 돈 못
청크 14: 그래서 아이고 이거 물 많이 먹겠네 했는데 아니나 다를까 왜냐면
1% | 1/152 [00:01<04:33, 1.81s/it] 청크 15:
청크 16: 예예 매 얼마 4 6원 마 4 6원 4 6천원 3 비중은 비중은 20%
청크 17: 더 더 그리고 못 오르는 거예요 그리고 얘네는 지금 센터 같은 경
청크 18: 이것도 예 그래서 예 안 파시면 나중에 후회하시지 않을까라는 생각
청크 19: 아 삼성화재 저하고 상담했어요 아니 먼저번에 또 했었다고 이게 자
청크 20: KT 보여 드리고 하면서 예예 예 이걸 어떻게 할지이 감당을 어떻게
청크 21: 쉽게 말하면은 이렇게 경험을 통해서 뭔가 느껴지는 것들이 있어요

YouTube API**KSSDS****FinBert*****FinBERT**

금융 도메인 감성 분석을 위한 BERT 모델

금융 텍스트만으로 학습됨 -> 금융 문맥을 통해 문장 이해 가능

3. 텍스트 전처리 및 감성분석

4

각 청크별 감성 분류 진행

청크1 : ABCDEFGHIJK

부정

0.XXX

중립

0.XXX

긍정

0.XXX

청크2 : CCCCCCCFDDE

0.XXX

0.XXX

0.XXX

...

청크N : AVSDFERSEA

0.XXX

0.XXX

0.XXX

Avg.

= 1

2024-07-11 자막 분석 중...

21% | 26/125 [00:09<00:29, 3.35it/s] 청크 1: [LQTFUOChbco] 지금부터는 투자해 놓 시간으로 이어가 보겠습니다... → 확률: [0.0001238690165337175, 0.9998766282828283]
청크 2: 일단 삼성전자 실적 대만 보더라도 재고 재고자산에 대한 평가 환입이 또 ... → 확률: [0.31383416056632996, 0.6474233269691467, 0.038742545992136]
청크 3: 근데 테스트가 통과가 되고 이제 제품 계약을 맺게 된다라면 거기에 맞춰 ... → 확률: [0.00937306135892868, 0.06772923469543457, 0.9228977560997009]
청크 4: 또 사상 최초로 또 AI 탑재된 애플 폰이기 때문에 22억 유저가 또 가... → 확률: [0.0016006184741854668, 0.0008792884182184935, 0.9975200295448303]
청크 5: 애플은 이제 채체 피트가 탑재가 되는 거고 어 삼성은 이제 재미 아이가 ... → 확률: [0.016907157376408577, 0.031281162053346634, 0.9518117308616638]
청크 6: 성능이 더고 고성능화 될 수밖에 없는 부분이기 때문에 국내 반도체 기업에... → 확률: [0.05526184290647507, 0.24654065072536469, 0.6981975436210632]
청크 7: 해적 케블 시장 정복에 나서는 데 어제 LS 전선이 1조원을 투자해서 미... → 확률: [0.0035077708307653666, 0.015001005493104458, 0.981491208076477]
청크 8: 저희 투자 있는 시간에서도 이런 전력 전선 쪽은 조정시 계속해서 매수해 ... → 확률: [0.00023559249530080706, 0.9991533756256104, 0.000611043069511652]
청크 9: 지금 한 20% 대에서 급등세를 보이면서 지금 상승폭이 다소 둔화됐기 있... → 확률: [0.02258470468223095, 0.45184069871902466, 0.5255746245384216]
청크 10: 그래서 크래프톤의 탑픽 종목으로 좀 생각해 보시면 되겠습니다 네 게임주 ... → 확률: [5.326087921275757e-05, 0.9998546838760376, 9.197730832966045e-05]
청크 11: 오늘 김명진 매니저는 매수해 놓 탑픽 주로 해저 케블 모멘텀을 가진 LS... → 확률: [0.009859232231974602, 0.8610281348228455, 0.1291126310825348]
→ 평균 확률: 부정=0.039, 중립=0.484, 긍정=0.477
 최종 감정 (청크 기반 Soft Voting): 1

4. 감성지수-시장지수 매핑

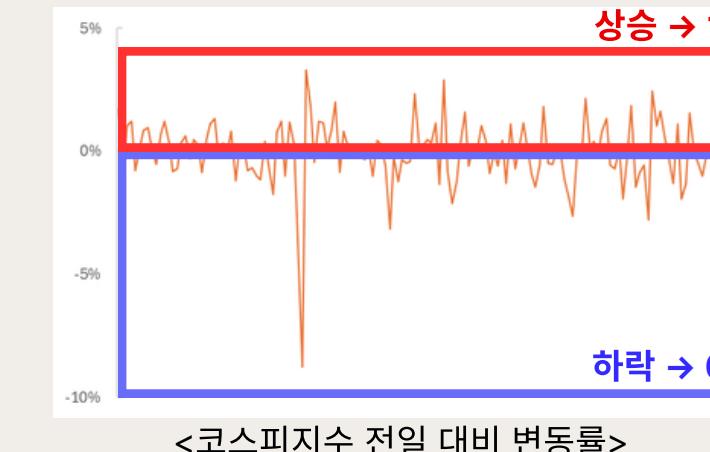
코스피 지수 전처리

1) 전일 기준 지수 상승 여부에 따라 이진 분류

*상승(1), 하락(0)

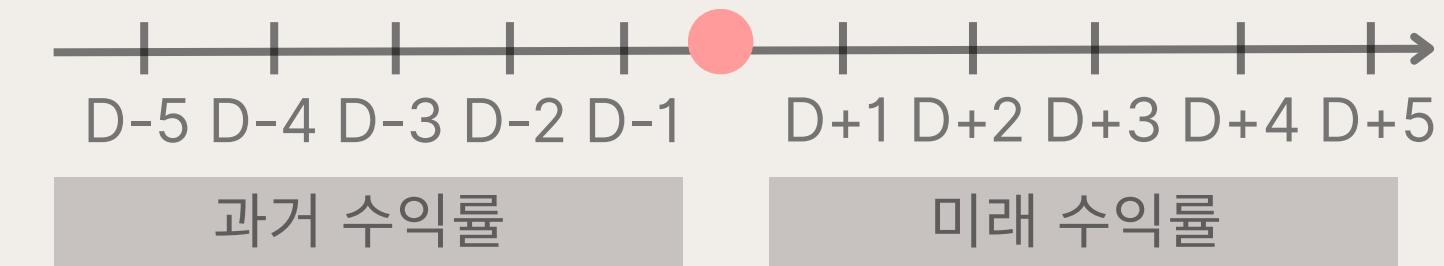


<코스피 지수>



<코스피지수 전일 대비 변동률>

2) 감정과 주가의 선후관계 분석을 위해 수익률 컬럼 생성



감성데이터와 주가 병합

*날짜별 감정과 그 날짜 혹은 ±N일의 주가 상승 여부를 연결

#	날짜	종가	시가	고가	저가	거래량	변동 %	변동 float	상승 여부	D+1_수익률	...	D-1_수익률	D-1_상승	D-2_수익률	D-2_상승	D-3_수익률	D-3_상승	D-4_수익률	D-4_상승	D-5_수익률	D-5_상승
19	2024-07-01	2804.31	2,792.96	2,805.95	2,783.98	459.22M	0.23%	0.23	1	-0.008362	...	0.002320	1	0.007274	1	0.004391	1	0.010784	1	0.014316	1
20	2024-07-02	2780.86	2,798.65	2,805.42	2,776.62	538.84M	-0.84%	-0.84	0	0.004729	...	-0.008362	0	-0.006062	0	-0.001149	0	-0.004008	0	0.002332	1
21	2024-07-03	2794.01	2,797.14	2,799.70	2,774.71	340.34M	0.47%	0.47	1	0.011070	...	0.004729	1	-0.003673	0	-0.001362	0	0.003574	1	0.000702	1
22	2024-07-04	2824.94	2,817.06	2,824.94	2,808.88	393.74M	1.11%	1.11	1	0.013200	...	0.011070	1	0.015851	1	0.007357	1	0.009693	1	0.014684	1
23	2024-07-05	2862.23	2,836.21	2,871.96	2,836.21	378.16M	1.32%	1.32	1	-0.001562	...	0.013200	1	0.024417	1	0.029261	1	0.020654	1	0.023021	1
24	2024-07-08	2857.76	2,862.58	2,868.99	2,855.53	425.06M	-0.16%	-0.16	0	0.003366	...	-0.001562	0	0.011618	1	0.022817	1	0.027653	1	0.019060	1

4. 감정지수 - 시장지수 시차 분석

4-1. 비율 검정

감정 분석 결과와 코스피 지수 상승 여부가 얼마나 일치하는지 시차별로 검정해보는 정확도 분석

H0 : 감정으로 주가 상승 여부를 맞출 확률은 0.5이다.(무작위 수준이다)

H1 : 감정으로 주가 상승 여부를 맞출 확률은 0.5보다 크다.

시차	맞춘 수	전체	정확도	z 통계량	p-value
0 D-5_상승	83	142	0.5845	2.043	0.0205
1 D-4_상승	83	142	0.5845	2.043	0.0205
2 D-3_상승	80	142	0.5634	1.523	0.0639
3 D-2_상승	81	142	0.5704	1.695	0.0450
4 D-1_상승	88	142	0.6197	2.939	0.0016
5 D+1_상승	82	142	0.5775	1.869	0.0308
6 D+2_상승	79	142	0.5563	1.351	0.0883
7 D+3_상승	82	142	0.5775	1.869	0.0308
8 D+4_상승	81	142	0.5704	1.695	0.0450
9 D+5_상승	83	142	0.5845	2.043	0.0205

<한경tv>

시차	맞춘 수	전체	정확도	z 통계량	p-value
0 D-5_상승	42	65	0.6462	2.464	0.0069
1 D-4_상승	41	65	0.6308	2.185	0.0145
2 D-3_상승	37	65	0.5692	1.127	0.1298
3 D-2_상승	39	65	0.6000	1.646	0.0499
4 D-1_상승	35	65	0.5385	0.622	0.2670
5 D+1_상승	36	65	0.5538	0.873	0.1912
6 D+2_상승	30	65	0.4615	-0.622	0.7330
7 D+3_상승	29	65	0.4462	-0.873	0.8088
8 D+4_상승	34	65	0.5231	0.373	0.3548
9 D+5_상승	31	65	0.4769	-0.373	0.6452

<전인구tv>

4. 감정지수 - 시장지수 시차 분석

4-2. 감성-주가 연관성이 통계적으로 확인된 채널들

유튜브채널명	D-5	D-4	D-3	D-2	D-1	D-0	D+1	D+2	D+3	D+4	D+5
매일경제tv	0.5659	0.5891	0.6047	0.5814	0.6202	0.6280	0.5039	0.4341	0.4961	0.4806	0.4574
머니투데이	0.5659	0.5891	0.6047	0.5814	0.6202	0.4600	0.5039	0.4341	0.4961	0.4806	0.4574
서울경제tv	0.3750	0.4583	0.5000	0.4167	0.4167	0.4170	0.4375	0.5208	0.5000	0.4792	0.3750
한국경제tv	0.5845	0.5845	0.5634	0.5704	0.6197	0.6270	0.5775	0.5563	0.5775	0.5704	0.5845
이트렌드	0.6234	0.5584	0.5455	0.5325	0.4935	0.4930	0.5455	0.4805	0.4545	0.5325	0.4935
위너스tv	0.4909	0.5545	0.5727	0.5818	0.5364	0.5450	0.5636	0.5727	0.5273	0.5545	0.5364
전인구tv	0.6462	0.6308	0.5692	0.6000	0.5385	0.5380	0.5538	0.4615	0.4462	0.5231	0.4769
불사조	0.5000	0.5500	0.6000	0.5250	0.4750	0.4750	0.5250	0.6000	0.5250	0.5500	0.5750
삼프로tv	0.6124	0.6279	0.6047	0.5581	0.5039	0.4960	0.5349	0.5194	0.4729	0.4729	0.4341

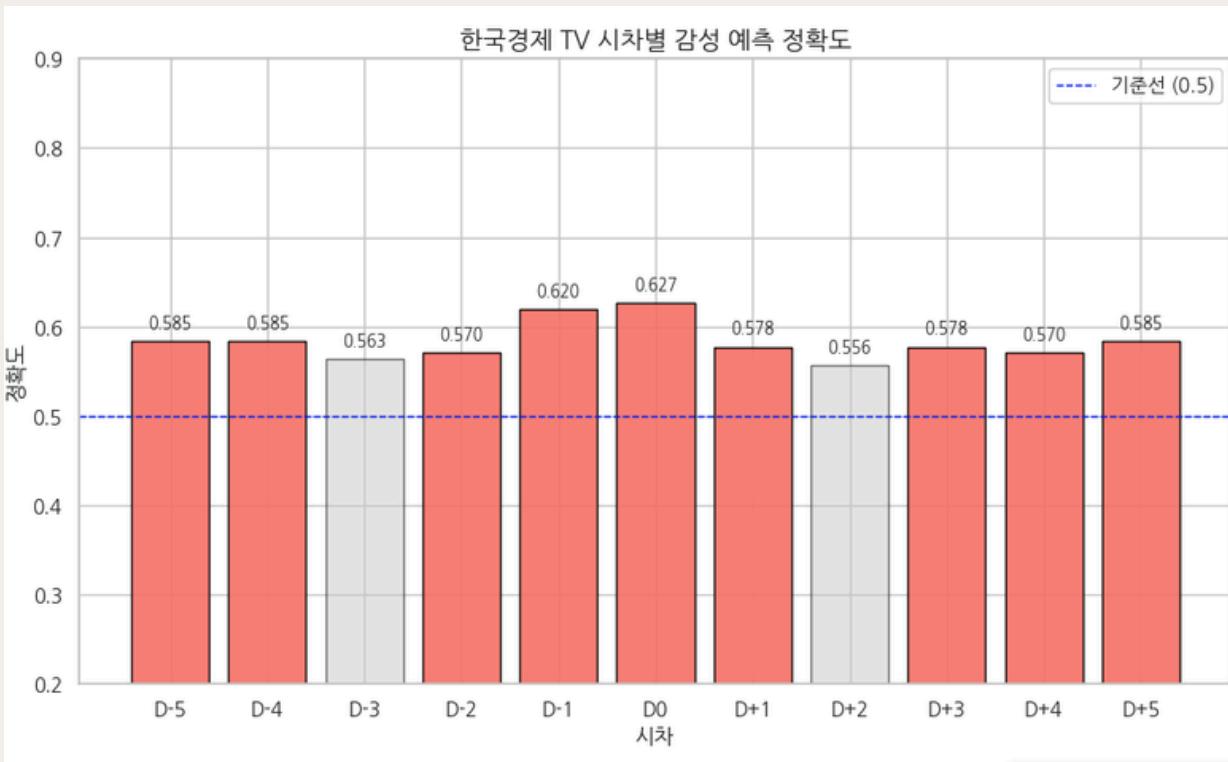
p<0.05 p<0.01

4. 감정지수 - 시장지수 시차 분석

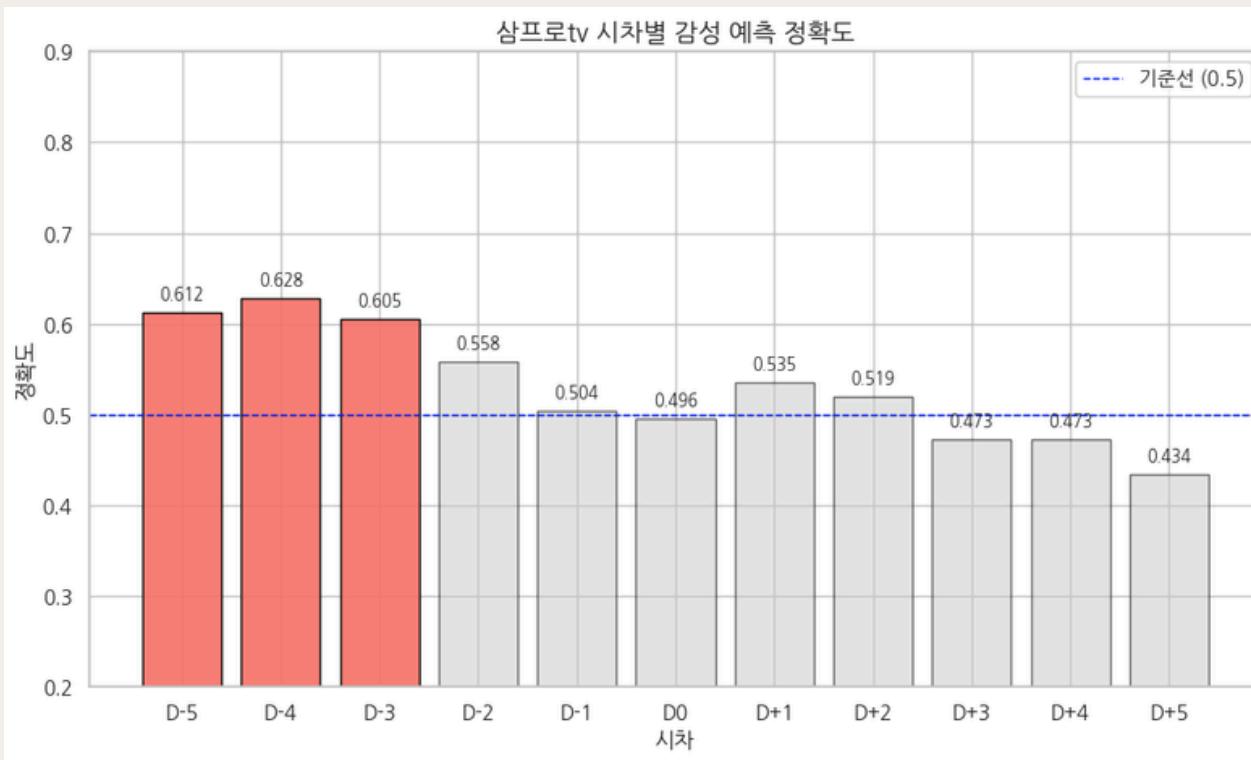
4-2. 감성-주가 연관성이 통계적으로 확인된 채널들



<한국경제 tv>



<3PRO tv>



한국경제 tv : 선행적 정보전달 채널

전체 시점에서 시장의 방향성과 감성분석의
결과가 유사한 양상을 띤다.
특히 미래 방향성에 대해 통계적으로 유의했던
'유일한' 채널

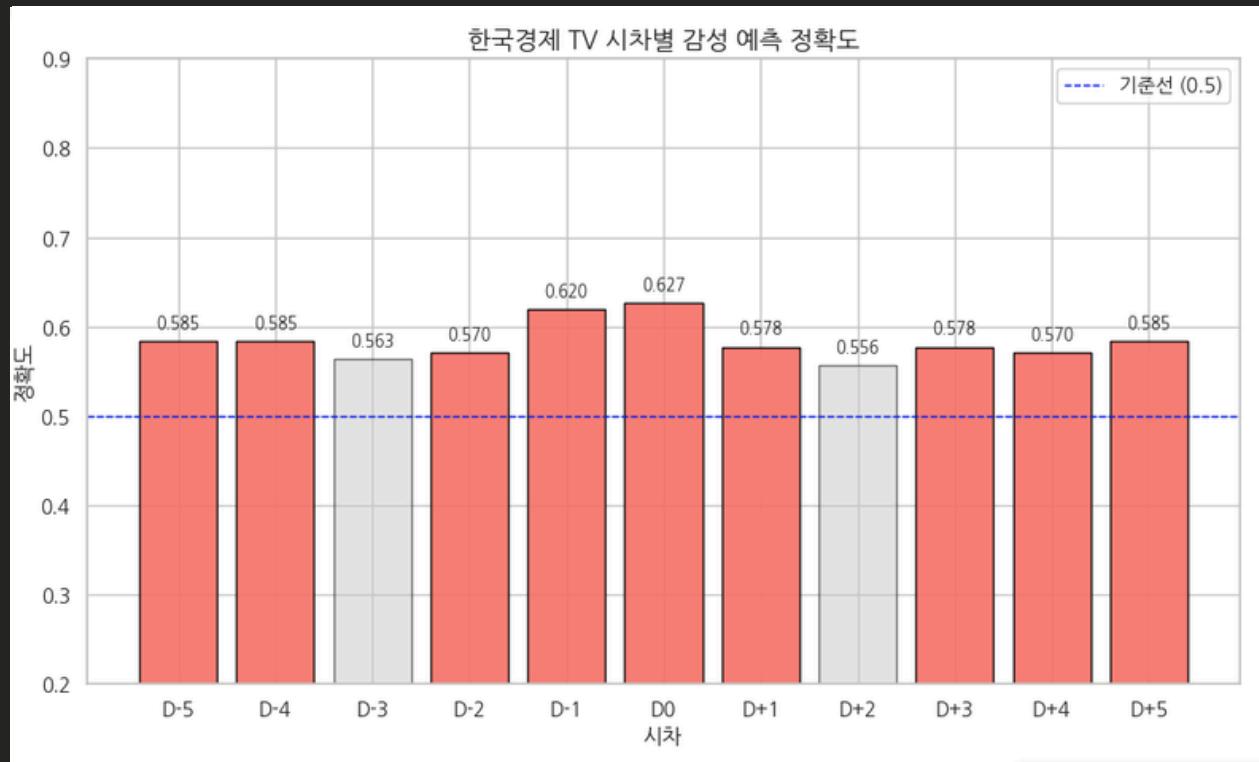
3PRO tv : 후행적 정보전달 채널

5일~3일 전 시장의 방향성과
감성분석 결과가 유사한 양상을 띤다.
D-5,4,3 : 통계적 유의한 시점

4. 감정지수 - 시장지수 시차 분석



<한국경제 tv>



<3PRO tv>



!!! 한국경제 tv 선행성 확인 !!!

한국경제 tv : **선행적** 정보전달 채널

전체 시점에서 시장의 방향성과 감성분석의
결과가 유사한 양상을 띤.
특히 미래 방향성에 대해 통계적으로 유의했던
'유일한' 채널

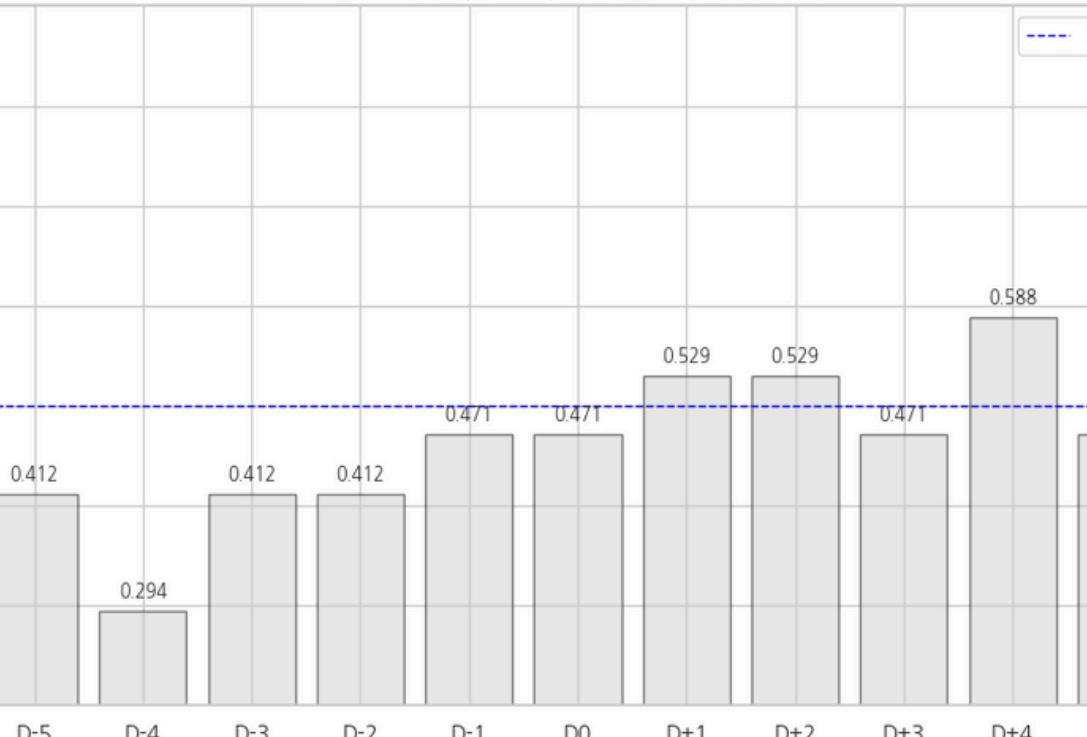
5일~3일 전 시장의 방향성과
결과가 유사함

3PRO tv : **후행적** 정보전달 채널

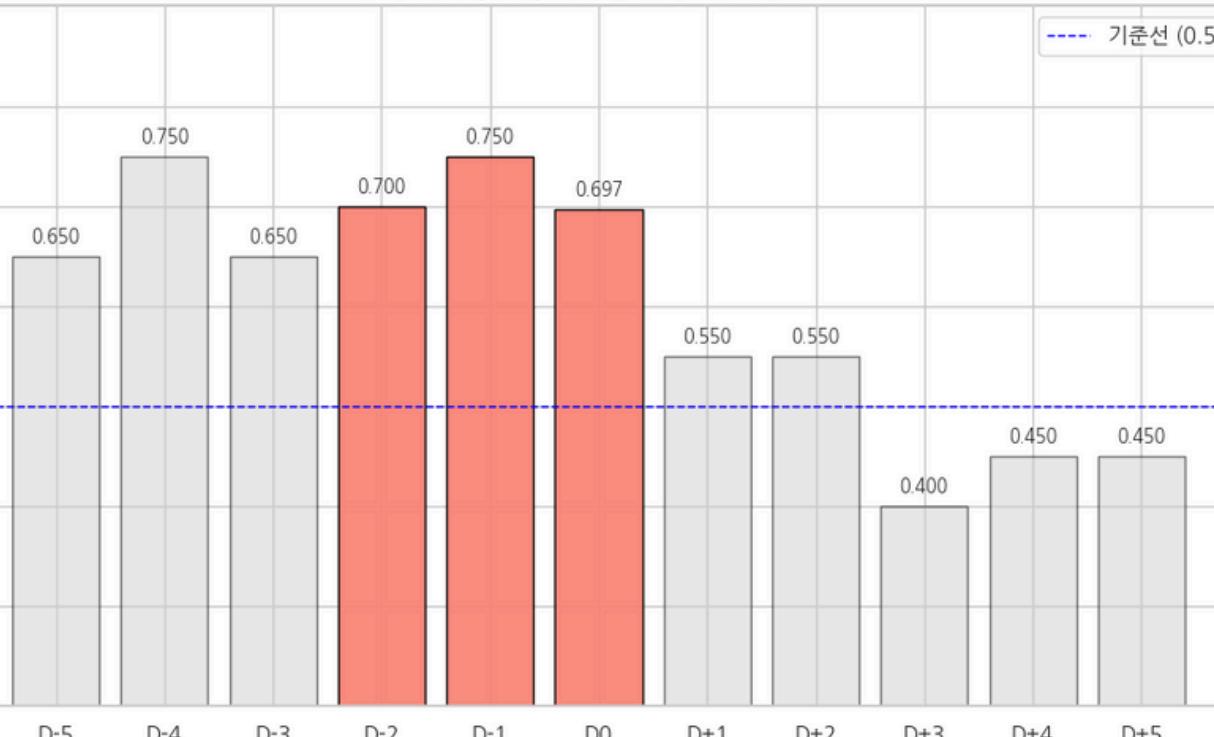
5. 전통매체 vs 유튜브 비교

한경tv 유튜브-금융리포트, 뉴스기사의 비교

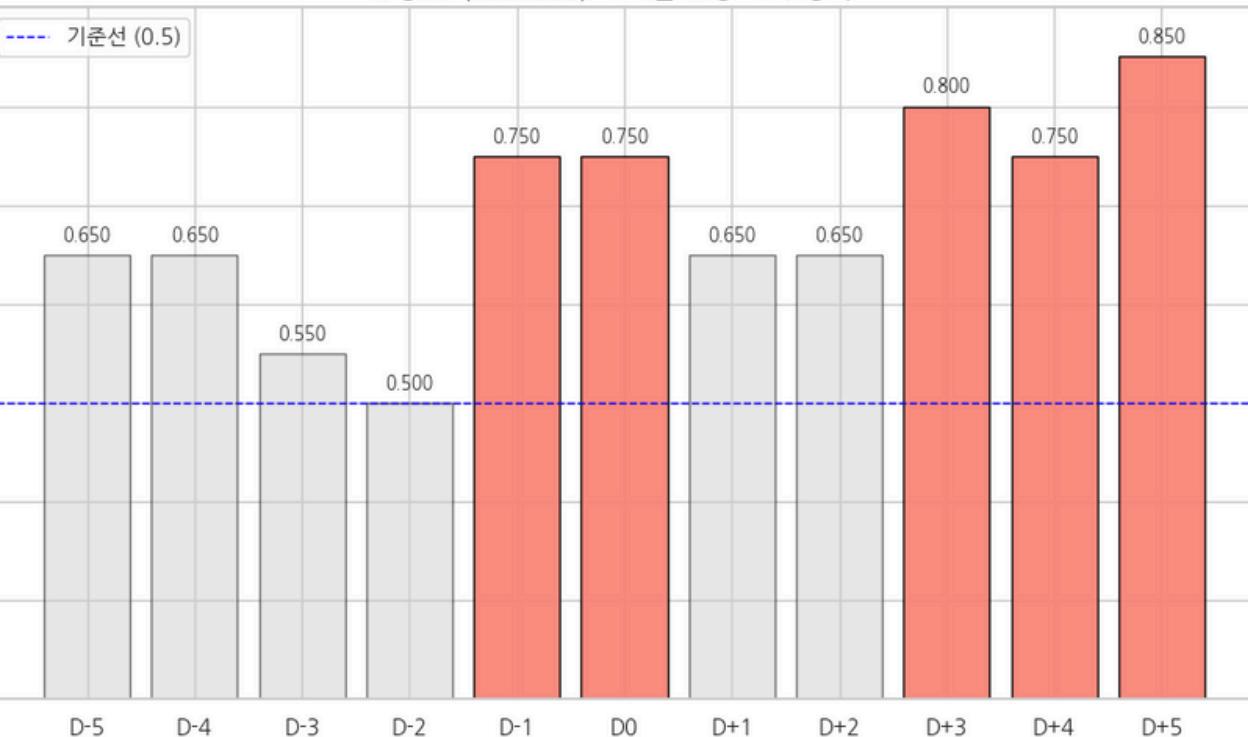
한경컨서스(리포트) 시차별 감성 예측 정확도



한경기사 (국내) 시차별 감성 예측 정확도



한경TV (YouTube) 시차별 감성 예측 정확도



<한경 컨서스 리포트>

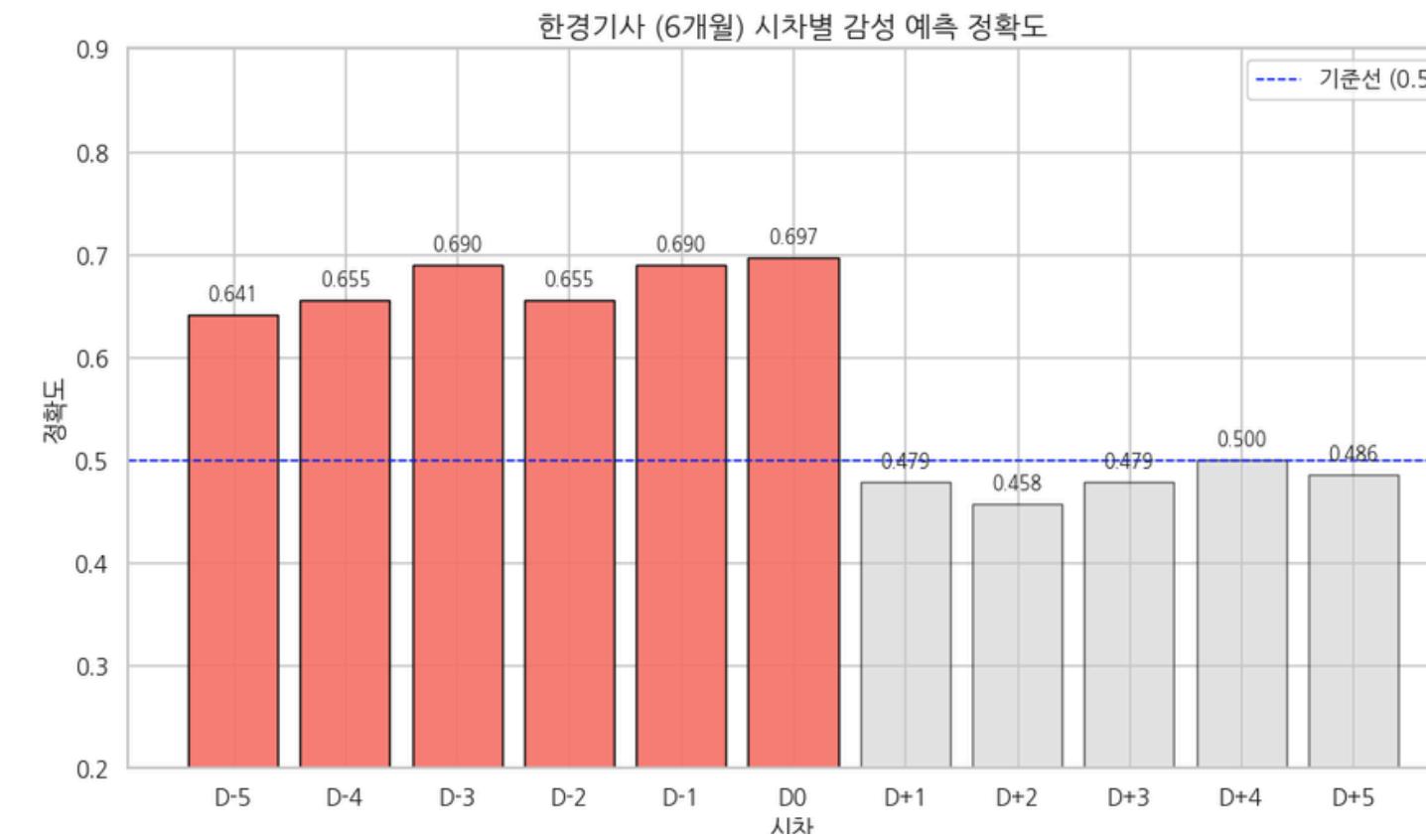
<한국경제 국내 뉴스>

<한국경제tv 유튜브>

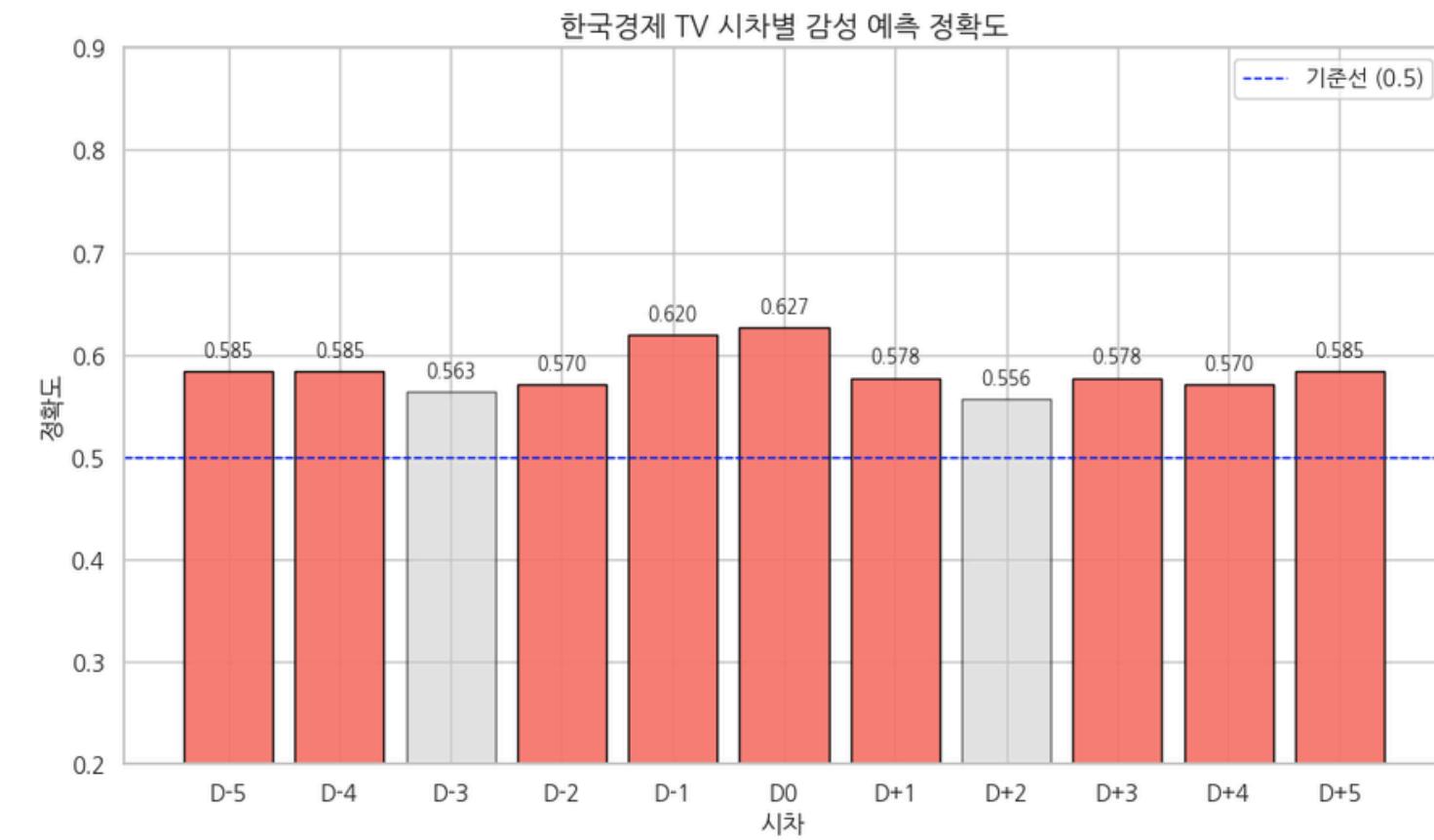
12月 한 달 간 한경 채널들의 감성 분석 비교 결과,
리포트, 국내뉴스 보다 유튜브의 선행성이 더 뚜렷함

5. 전통매체 vs 유튜브 비교

한경tv 유튜브-금융리포트, 뉴스기사의 비교



<한국경제 국내 뉴스>

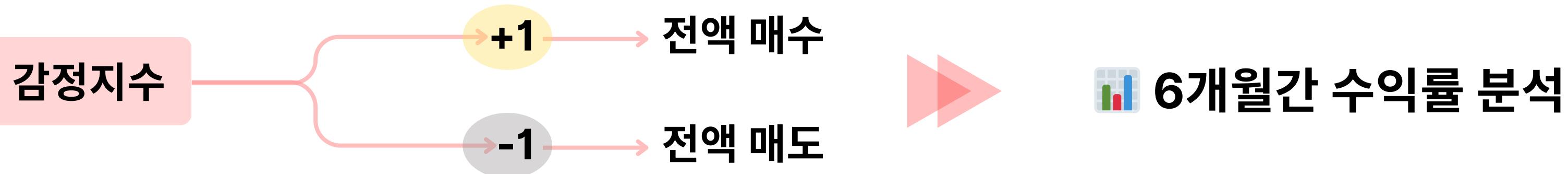


<한국경제tv 유튜브>

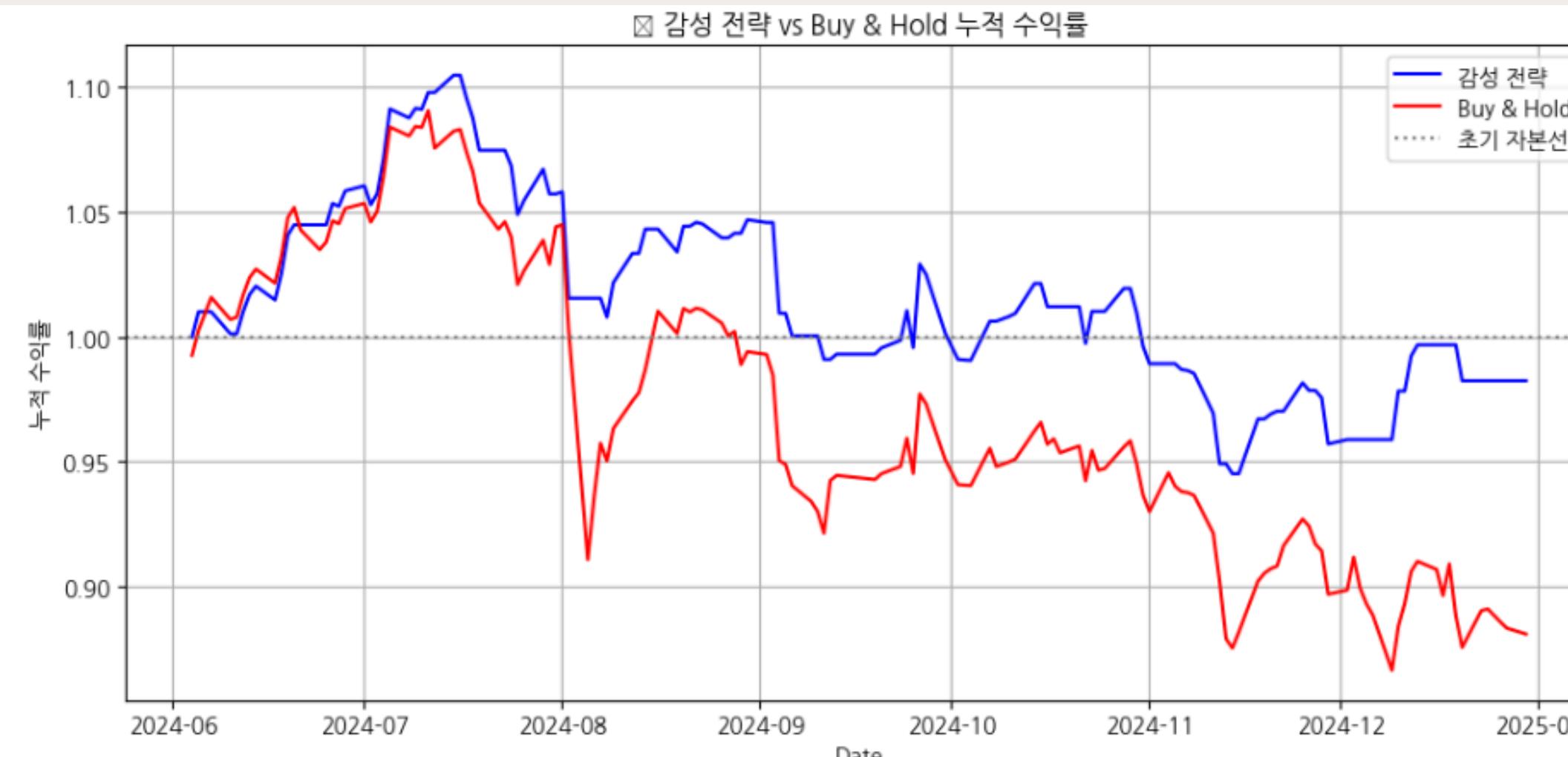
6개월간 비교했을 때도 비슷한 경향성

6. 감정지수 기반 백테스팅

한경tv 감정지수 기반 6개월간 백테스팅



[교차표 - D+1]	코스피_상승	상승	하락
감성	49	37	
긍정	49	37	
부정	23	32	



전 구간에서 수익률 강세
특히 '하락장'에 수익률 강세

감정지수 기반 투자전략의
실효성 확인!

6. 투자 보조 서비스 제안



주식시장 분석

감정과 함께 투자하기
10만 조회수 · 1일 전

감성분석 결과

매수 추천!

해당 채널의 1년간 적중도

적중률
65%

평균 수익률
7%

“감정과 함께 투자하기”

서비스 예시 흐름

1. 유튜브 영상 감성분석 결과 요약

- 예: “매수 추천” or “주의 요망” 등으로 요약

2. 채널 신뢰도 지표 제공

- 채널별 감성의 시장 예측 적중률
- 수익률 기반 백테스트 결과: 예측에 따라 투자 시 수익률 등

3. 개인화 추천

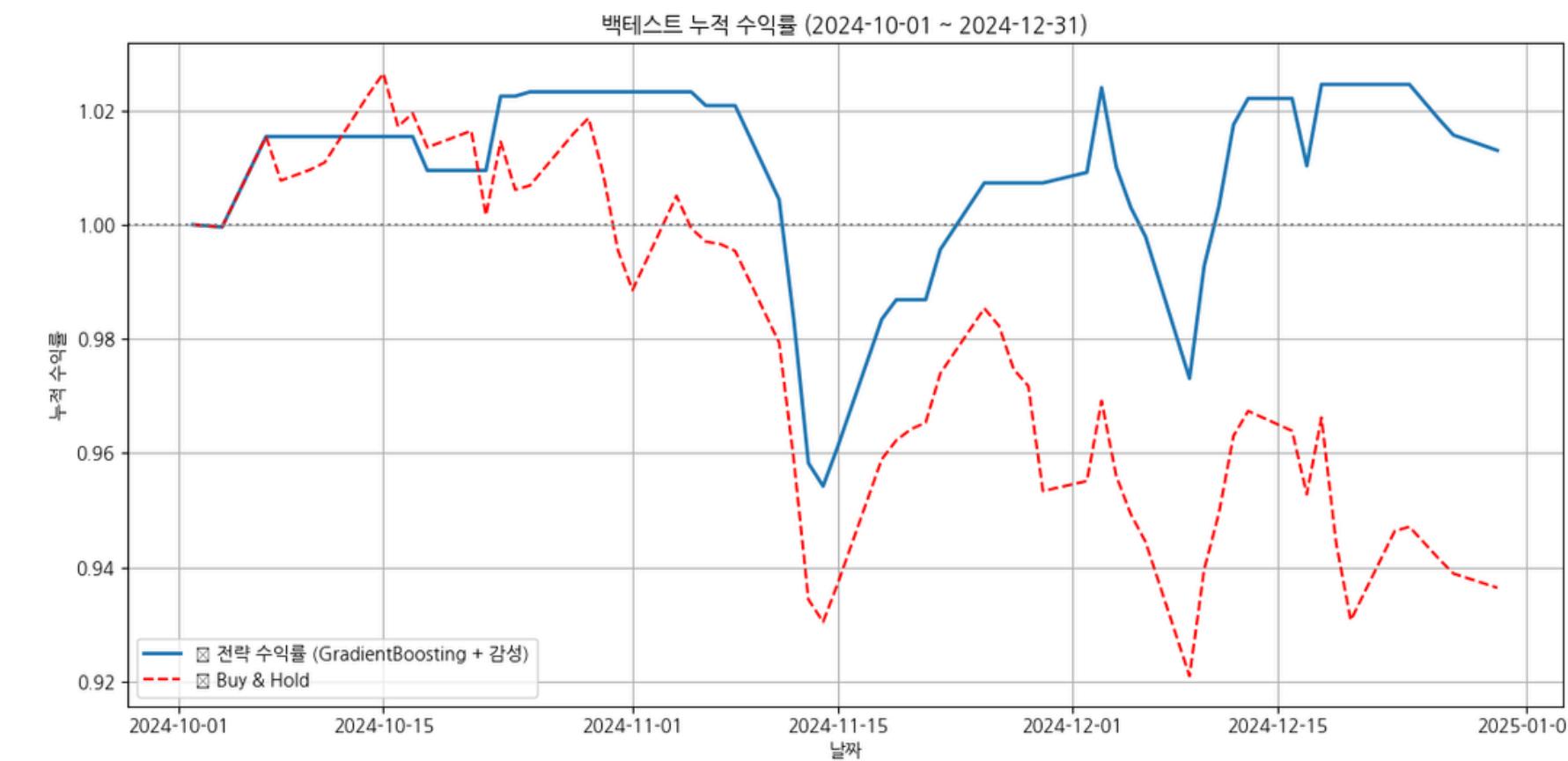
- 유저가 즐겨보는 유튜브 채널 기반, 감성 점수 피드백 제공
- “이 채널은 지난 1년간 예측력이 ○○%였습니다”

7. 한계점

구분	한계 내용	설명 요약	해결 방안
변수 누락	주가는 다양한 외부 요인의 영향을 받음	감성 점수만으로 주가 변동 설명에 한계 존재	추가적 데이터 수집 및 모델 고도화 필요
모델 한계	FINBERT는 금융 뉴스/ 문서 / 댓글 기반으로 학습됨 토큰 수 제한	유튜브 영상의 구어체와 비정형 표현에 완전 들어맞진 않았다.	추후 관련 감성사전 학습 및 파인튜닝 LLM 기반 감성분석
감성분석 왜곡	감성의 강도가 고려되지 않고 -1, 1만으로 도출		LLM 활용 감성분석으로 수치형 변수로 보완
편향 가능성	채널별 정치적, 정서적 편향이 결과에 영향 가능	특정 시청층 대상 메시지로 왜곡될 수 있음	여러 채널의 데이터 합쳐서 분석

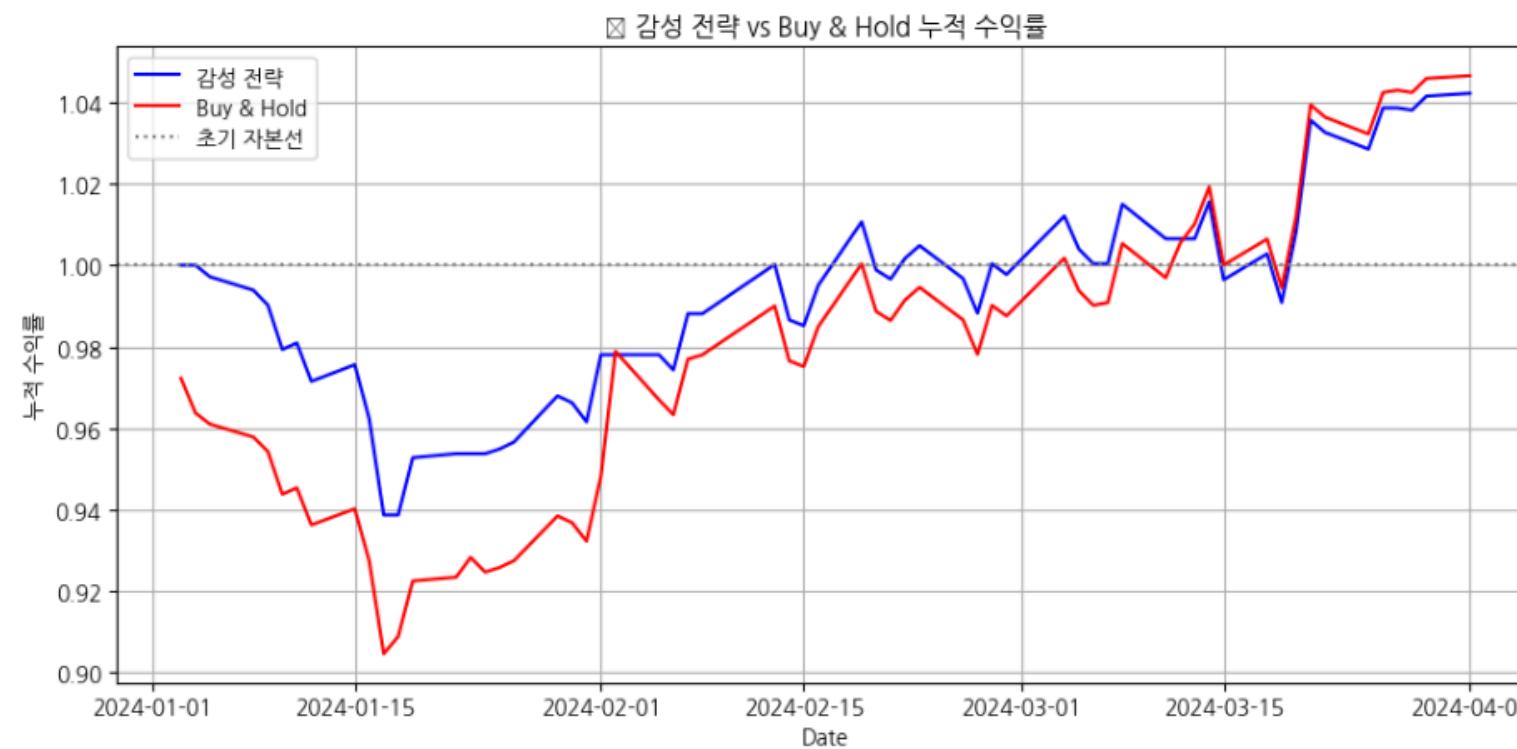
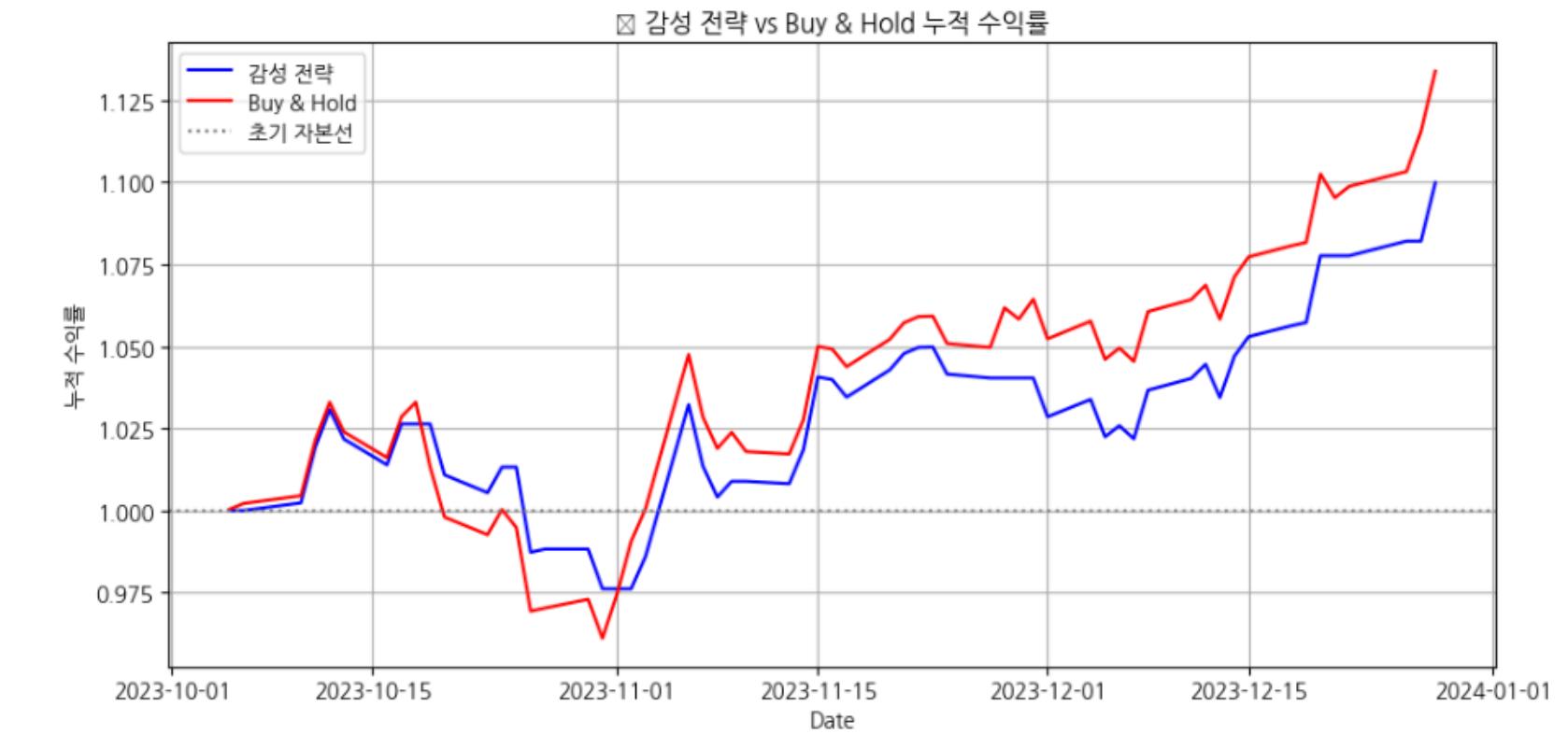
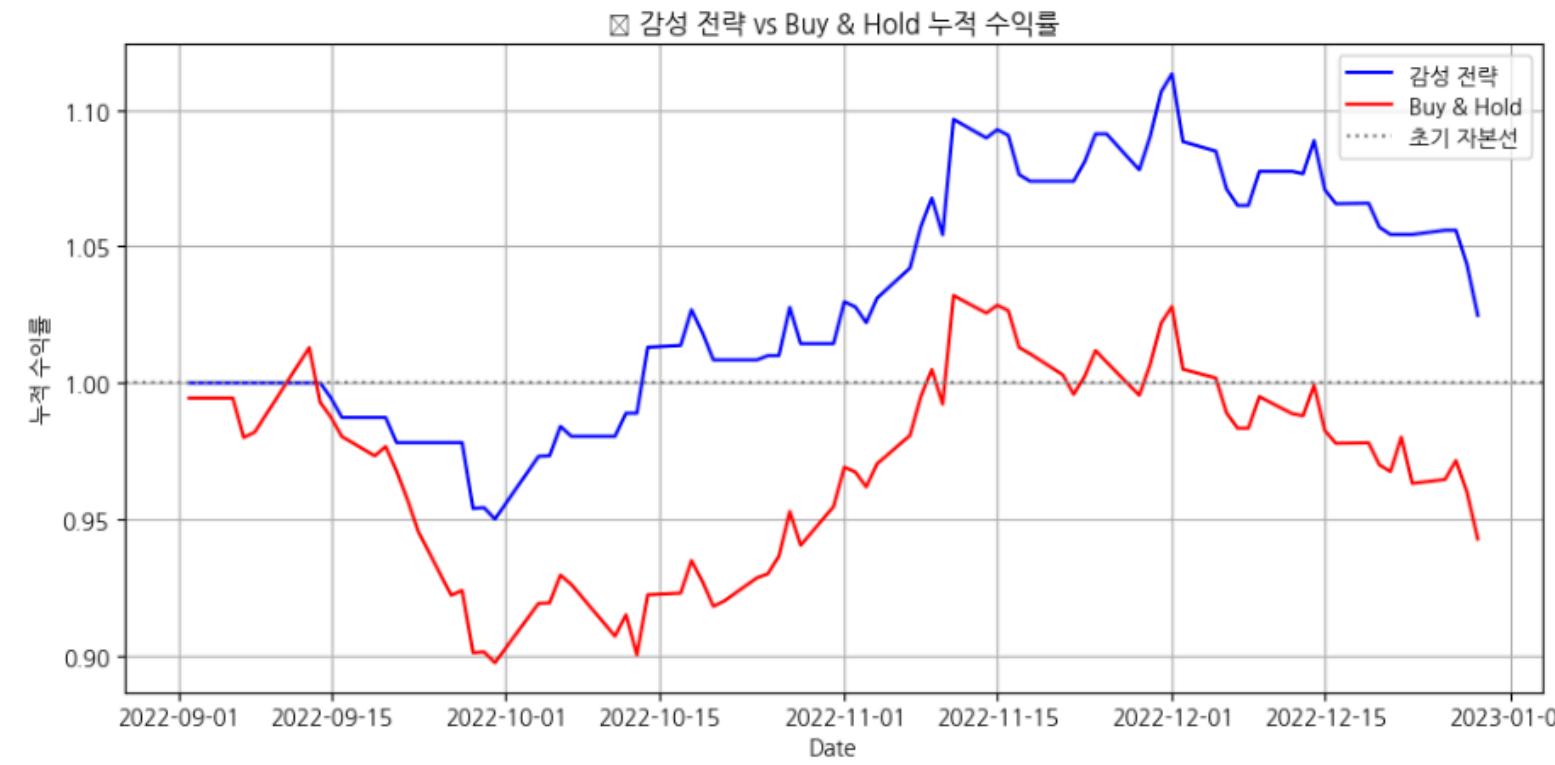
8. Appendix

Feature	미포함	포함
base	0.500000	0.493421
base+usd	0.486842	0.493421
base+allmacro	0.532895	0.559211
base+allmacro+5d	0.546053	0.565789



feature: 시가, 최저가, 최고가, 종가, 거래량, 환율, VIX, 국제유가,
감정분석결과, 1일 전~5일 전 종가 상승여부

8. Appendix





감사합니다

Thank you

谢谢 (xièxìè)

ありがとう

Gracias

Merci

Danke

Grazie

Спасибо

شكرا

ধন্যবাদ

Cảm ơn

ខូចគុណ

·

·

·

금융A팀 | 주식전문채널 DARtv