

Web Scraping과 ML을 이용한 뉴스 데이터 분석 및 유가 예측

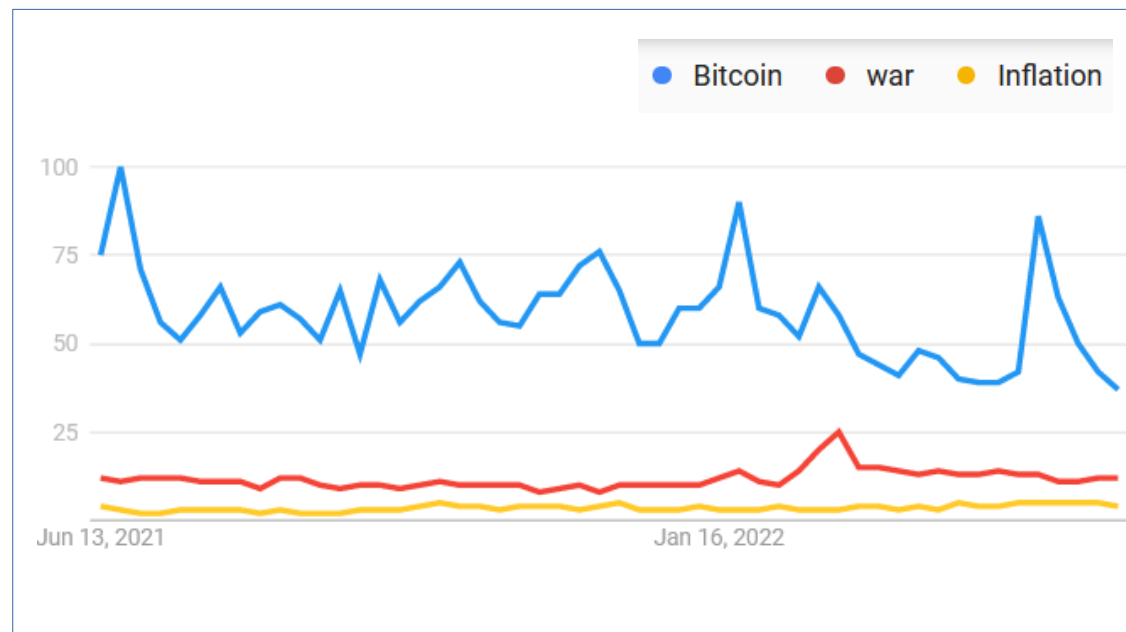


2021/06/14

0. 배경

인플레이션, 전쟁, 암호화폐

- 근래 인플레이션, 전쟁, 암호화폐에 대한 관심이 크게 증가 하였음
 - 인플레이션: 점진적으로 관심이 증가하고 있음
 - 전쟁: 우크라이나 전 발발 후 크게 관심이 증가 한 후 지금은 작년대비 조금 높은 관심 수준 유지 중
 - 암호화폐: 작년 대비 낮은 관심 수준을 유지 하고 있었으나 5월 중순 루나 폭락 후 관심 급증



source: google trend

0. 배경

인플레이션, 전쟁, 암호화폐

- 근래 인플레이션, 전쟁, 암호화폐에 대한 관심이 크게 증가 하였음
 - 인플레이션: 점진적으로 관심이 증가하고 있음
 - 전쟁: 우크라이나 전 발발 후 크게 관심이 증가 한 후 지금은 작년대비 조금 높은 관심 수준 유지 중
 - 암호화폐: 작년 대비 낮은 관심 수준을 유지 하고 있었으나 5월 중순 루나 폭락 후 관심 급증

비정형 데이터 분석의 활용

- 비정형 데이터를 활용할 경우 정형데이터로는 확인 불가능한 여러 정보(e.g. consumer/investor sentiment)를 활용하여 시장에 대한 예측 및 분석을 수행할 수 있음
- 예: 인플레이션에 대한 social sentiment와 에너지 가격 사이의 correlation (but likely not causation) 조사

1. 5월 경제 분야 뉴스 데이터 수집 결과

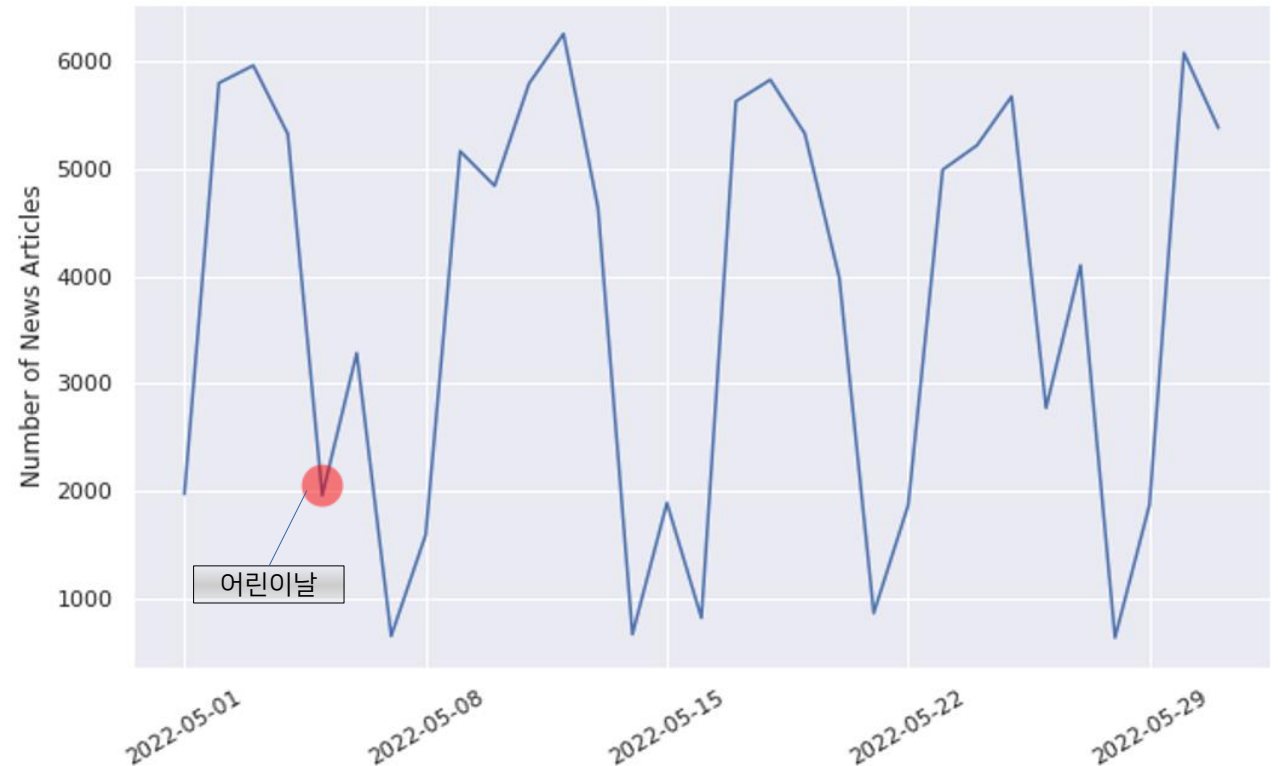
Total & Daily news

- 수집된 기사 수: 총 116,763개
- 요일에 따른 seasonality가 관측됨

요일 별 평균 기사 수

Mon	4565
Tue	5399
Wed	5647
Thu	4076
Fri	3999
Sat	706
Sun	1842

Daily News Count



2. 5월 월간 TOP 5 언론사

언론사별 News count (TOP 10)

0	뉴스1	10929
1	연합뉴스	10920
2	뉴시스	10818
3	이데일리	9562
4	파이낸셜뉴스	6852
5	머니투데이	6010
6	아시아경제	5732
7	서울경제	5153
8	한국경제	5033
9	헤럴드경제	4407

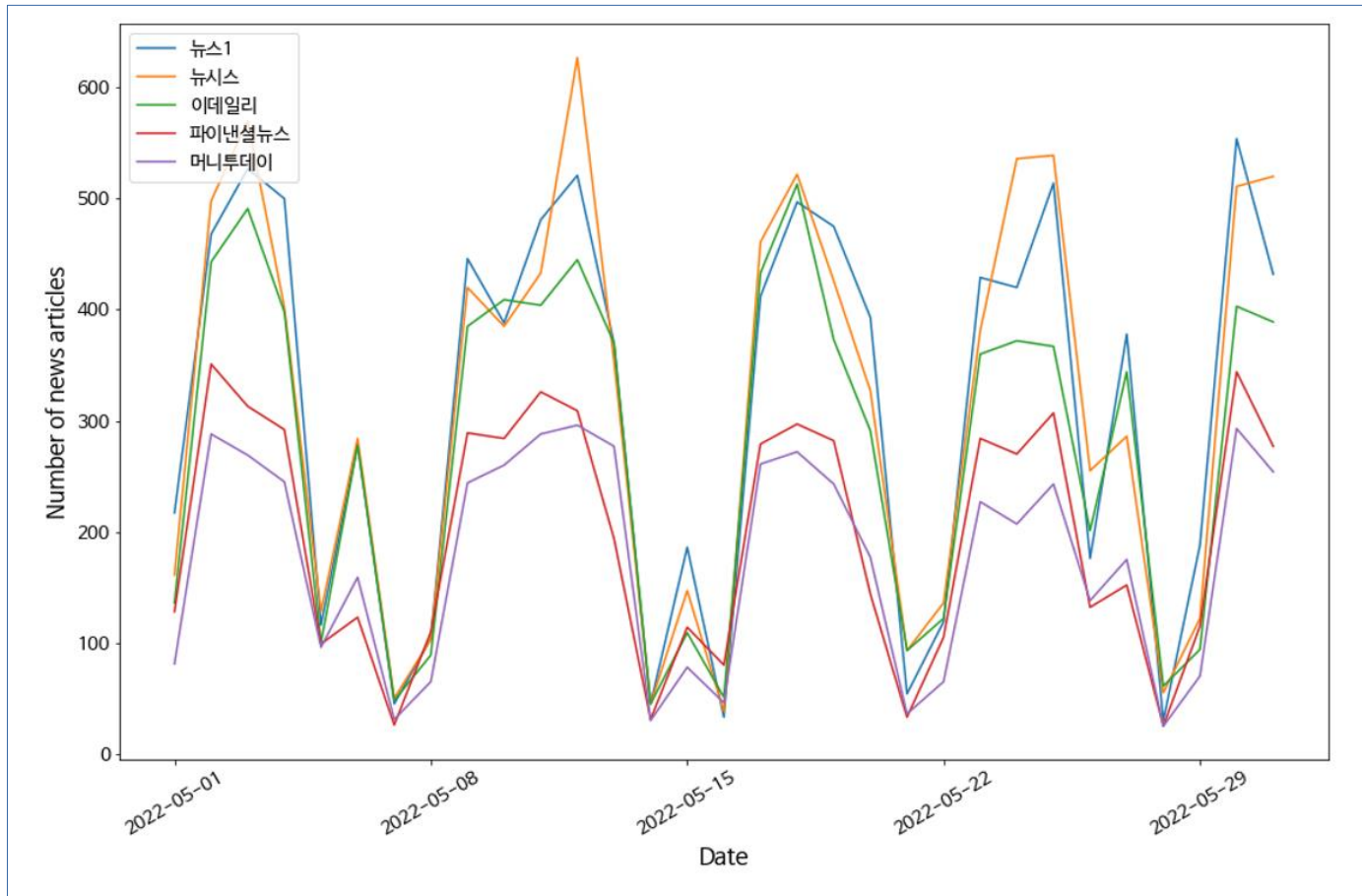
Top 5 언론사 점유율

뉴스1	9.36%
뉴시스	9.26%
이데일리	8.19%
파이낸셜뉴스	5.87%
머니투데이	5.15%

- 연합뉴스의 경우 언론사들에게 뉴스 콘텐츠를 제공하는 역할을 수행하므로 Top 5에서 제외

2. 5월 월간 TOP 5 언론사

Top 5 언론사 Daily News Count

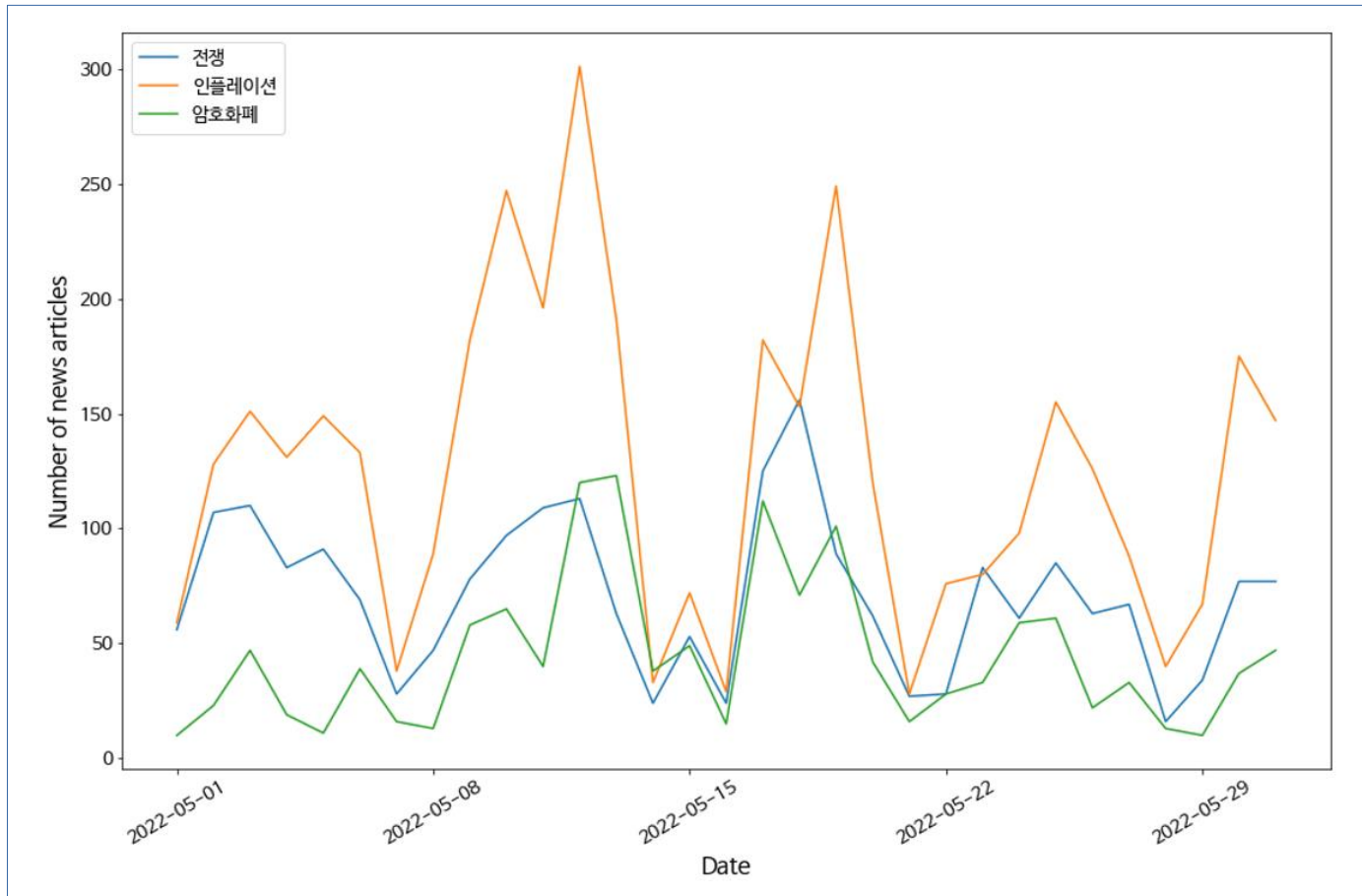


Trend Analysis

- Top 5 언론사 모두 weekly seasonal trend가 존재함
- 언론사 간 뉴스 기사 수 점유율의 차이는 요일에 관계없이 거의 일정하게 유지되고 있음

3. 주요 키워드 관련 뉴스 데이터 동향

키워드별 Daily News Count

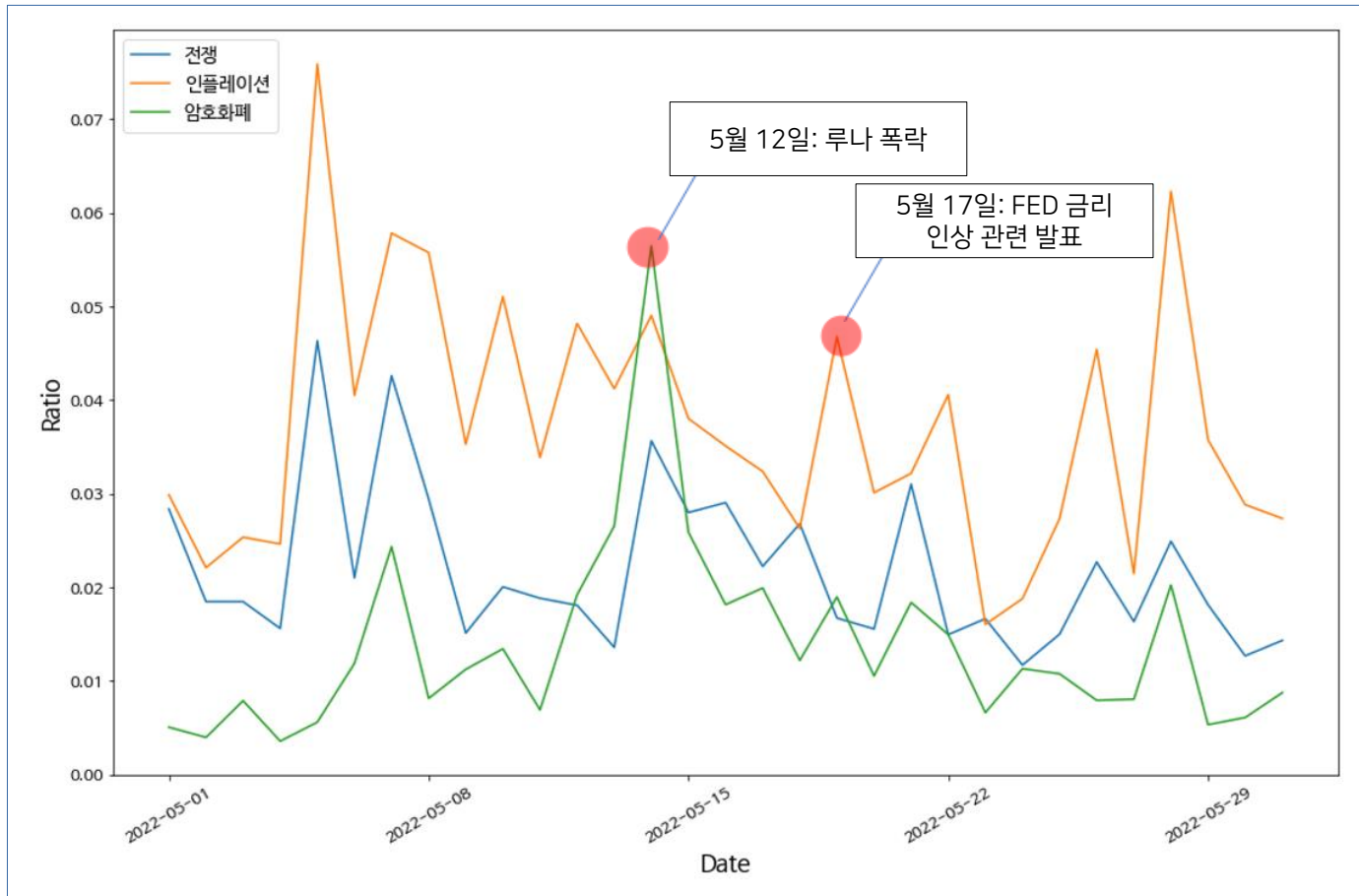


Trend Analysis

- 전반적으로 인플레이션, 전쟁, 암호화폐 순으로 기사 수가 많았음
- Weekly seasonality는 keyword와 관계없이 항상 존재했음

3. 주요 키워드 관련 뉴스 데이터 동향

키워드별 뉴스 비율





- 전쟁: 1.2~4.5%
- 인플레이션: 1.6~7.6%
- 암호화폐: 0.4~5.6%

4. 일간 TOP 3 이슈

2022-05-01

- #1: 중기부, 해외규격인증획득지원사업 참여 기업 모집
- #2: 에너지 가격 급등에 4월 무역수지 또 적자...수출은 12.6% ↑
- #3: 4월 수출 576.9억 달러...무역수지는 적자

 인플레이션 관련

 암호화폐 관련

2022-05-02

- #1: [인사] 한양증권
- #2: 코스피 하락 출발
- #3: 신한금융투자, 이마트24와 '열어보니 미국주식' 이벤트

2022-05-03

- #1: 홍남기 부총리, 물가관계장관회의의 마지막 주재
- #2: 4월 소비자물가 4.8% ↑ ... 13년 6개월 만에 최고 상승률
- #3: 요기요, 가정의 달 맞아 '감사 릴레이' 할인 이벤트

4. 일간 TOP 3 이슈

2022-05-04

- #1: 美 FOMC 하루 앞두고 코스피 사흘째 하락
- #2: 카카오뱅크, 예·적금 금리 최대 0.4%p 인상
- #3: [인사] 동양생명


2022-05-05

- #1: GS리테일 '어바웃펫', 네이버서 100억원 투자 유치
- #2: 포스코건설, 700여개 협력사 대상 '동반성장 프로그램' 운영
- #3: 한지로 만든 환상적인 야경

2022-05-06

- #1: 코스피 1.23% 하락 마감
- #2: 삼성·LG전자, 美 '2022 에너지스타상' 최고상 수상
- #3: 머스크, 트위터 인수 위해 9조원 투자 유치

 인플레이션 관련

 암호화폐 관련

4. 일간 TOP 3 이슈

2022-05-07

- #1: 51년 된 여의도 시범아파트, 최고 60층 재건축 추진
- #2: 뉴욕증시, 긴축 우려 등에 또 하락...나스닥 1.40% ↓
- #3: 서울 아파트 매수심리 회복했다...새 정부 출범 앞두고 '반등'



인플레이션 관련



암호화폐 관련

2022-05-08

- #1: 20대 가계대출 감소했지만...제2금융권 대출·다중채무자는 증가
- #2: 한경연, 올해 경제성장률 전망 2.9%→2.5%로 하향 조정
- #3: 비트코인, 4천530만원대로 하락...2개월여 만의 최저

2022-05-09

- #1: '코로나 초기'로 돌아간 코스피
- #2: CJ대한통운, 1분기 영업이익 757억원...전년比 57.2% ↑
- #3: [인사] IBK투자증권


4. 일간 TOP 3 이슈

2022-05-10

#1: 코스피 2600선 붕괴

#2: 국내증시 골치

#3: 작년 공공기관 47% 적자...한전 5조8천억 적자

 인플레이션 관련

 암호화폐 관련

2022-05-11

#1: 새 정부 국정과제 '탄소 중립'...구미 제조업체 80% "대응 못해"

#2: 5월 1~10일 수출 28.7% 증가...무역적자 37억달러

#3: 4월 은행 가계대출 1.2조 ↑ ...5개월만에 상승 전환

2022-05-12

#1: 하락세 지속 코스피, 원 달러 환율은 연고점 경신

#2: 쿠팡, 1분기 매출 '사상 최대'

#3: 휘발유·경유價 둘 다 오른다..."하반기까지 지속될수도"

4. 일간 TOP 3 이슈

2022-05-13


- #1: 쌍용차 새주인 후보 'KG그룹 컨소시엄'
- #2: LGU+, 1분기 영업이익 2612억원...전년比 5.2% ↓
- #3: 롯데케미칼, 1분기 영업이익 826억...전년비 86.8% ↓


2022-05-14

- #1: 삼성, 직무적성검사 오늘부터 이틀간 온라인으로 진행
- #2: 전 세계 주요 거래소, 루나 · 테라 거래중단 · 상장폐지
- #3: 피자쿠치나, 불향 가득 신메뉴 '불타오르네' 선보여

2022-05-15

- #1: 휘발유보다 비싼 경유
- #2: 중대재해법 시행 100일...기업 69% "법 이해 못해 대응 어려워"
- #3: 인도 밀 수출 중단

 인플레이션 관련

 암호화폐 관련

4. 일간 TOP 3 이슈

2022-05-16

#1: CJ, 1분기 영업익 4978억... 전년比 11.2% ↑

#2: 테라 측 "비트코인 8만개 팔았다" 남은 코인 313개 뿐

#3: 대우조선해양 1분기 영업손실 4701억...적자 지속

 인플레이션 관련

 암호화폐 관련

2022-05-17

#1: '타운홀미팅' 취임식에 시장 방문까지...잇따른 소통 행보 나선 秋

#2: 지속되는 비트코인 약세장

#3: 치솟는 경유 가격...내달부터 보조금 리터당 50원씩 더 준다

2022-05-18

#1: [1Q 코스피][표]개별 순이익 상위20사

#2: 루나-테라 사태...금융당국, 코인 규제로 선회하나

#3: KDI, 올해 성장률 2.8%로 하향...물가상승률은 4.2% 전망

4. 일간 TOP 3 이슈

2022-05-19


- #1: 현대캐피탈, 인천해송고등학교서 '1사1교 금융교육' 실시
- #2: 하락 출발한 코스피
- #3: 직방, 미국 법인 설립...글로벌 가상오피스 '소마' 출시


2022-05-20

- #1: 상승 개장 코스피
- #2: IBK기업은행, '신용회복 성실상환자' 카드 2만7000명 발급
- #3: 바이든 방한 효과...넷플릭스 자회사, 한국에 1억 달러 투자

2022-05-21

- #1: 현대중공업-현대미포조선, 현대중공업그룹 4기 기술연수생 모집
- #2: 머스크 `성추행 의혹`에 테슬라 주가 6%대 추락...`육백슬라` 됐다
- #3: 한·미 외환시장 안정 협력...상설 통화스와프 힘 실리나

 인플레이션 관련


 암호화폐 관련

4. 일간 TOP 3 이슈

2022-05-22

- #1: 정의선 "로보틱스·자율주행 등 미국에 50억달러 추가 투자"
- #2: 바이든 "美 선택한 현대차에 감사...실망시키지 않을 것"
- #3: '루나' 권도형 "한국에 미납세금 없다...숨길 것 없어" 주장

 인플레이션 관련

 암호화폐 관련

2022-05-23

- #1: 해양생물자원관, 해양교육 구독서비스 6월부터 운영
- #2: 삼성 생활가전 '최고 브랜드' 연이어 선정...유럽 사로잡아
- #3: 다올저축은행, 정기예금 금리 최대 0.4%p 인상...연 최고 3.3%

2022-05-24


- #1: GC녹십자, 고혈압·고지혈증 '4제 복합제' 식약처 허가
- #2: 대한항공, 올해 만료 마일리지 유효기간 1년 연장
- #3: 카카오페이, ESG 추진위원회 신설

4. 일간 TOP 3 이슈

2022-05-25

- #1: 중기부, 중소기업 탄소중립 대응 지원 본격화
- #2: 힐스테이트 불로 포레스트, 분양
- #3: [가상화폐 시세] 비트코인, 이 시간에 사면 손해본다?

 인플레이션 관련

 암호화폐 관련

2022-05-26

- #1: 무협 '한·중 기업인 협력 교류회' 성료... "코로나19 이후 협력 강화"
- #2: 노노스창업교육학원, '희망리턴패키지' 정부지원 무료교육 과정
- #3: [김대호의 경제읽기] 한은, 기준금리 1.75%로... 추가 인상 전망은?

2022-05-27

- #1: 우미그룹, 2022년 상반기 대졸 신입사원 공채
- #2: 인도, 밀·설탕 이어 쌀 수출 금지할수도... "시간 문제"
- #3: 신한은행, 예적금 금리 최고 0.4%p 인상

4. 일간 TOP 3 이슈

2022-05-28

- #1: [지표로 보는 경제] 5월 28일
- #2: 기준금리 인상이 내 삶에 미치는 세가지 영향
- #3: "김포-하네다 노선 6월 15일 운항 재개 검토"

 인플레이션 관련

 암호화폐 관련

2022-05-29

- #1: 수입 육류가격지수 1년새 39% 올라...소고기 56%·돼지고기 14% ↑
- #2: 밀리의 서재, 코스닥 상장 절차 돌입
- #3: 삼성·SK·LG·현대차 전략회의 개최...경영 환경 점검

2022-05-30

- #1: 6월 1일부터 화물차·택시·버스 경유보조금 확대
- #2: 보유세 부담 2020년 수준으로 인하
- #3: 소상공인 371만 명에 최대 1천만 원...오늘부터 손실보전금 지급

4. 일간 TOP 3 이슈

2022-05-31

- #1: 롯데백 부산 광복점 내일부터 영업 중단
- #2: 일본 패키지 관광 허용, 일본 노선 활성화 기대
- #3: 굿스터프이터리 국내 1호점 개점



인플레이션 관련



암호화폐 관련

Top 3 이슈 요약

- 인플레이션과 관련된 기사가 앞서 사용했던 3가지 keyword 중 가장 많이 주요 기사로서 선정되었음
- 전쟁 관련 뉴스는 5월 간 단 하루도 TOP3 이슈에 들지 못했음
- 암호화폐 관련 주요 기사는 주로 LUNA 폭락과 관계되어 있고 따라서 5월 12일 이후 집중 되어있음
- 기존 키워드 외 자주 언급된 주요 주제로는 금리, 무역수지, 주가, 기업 동향 등이 있었음

5. 키워드 관련 주요 이슈 : “물가”

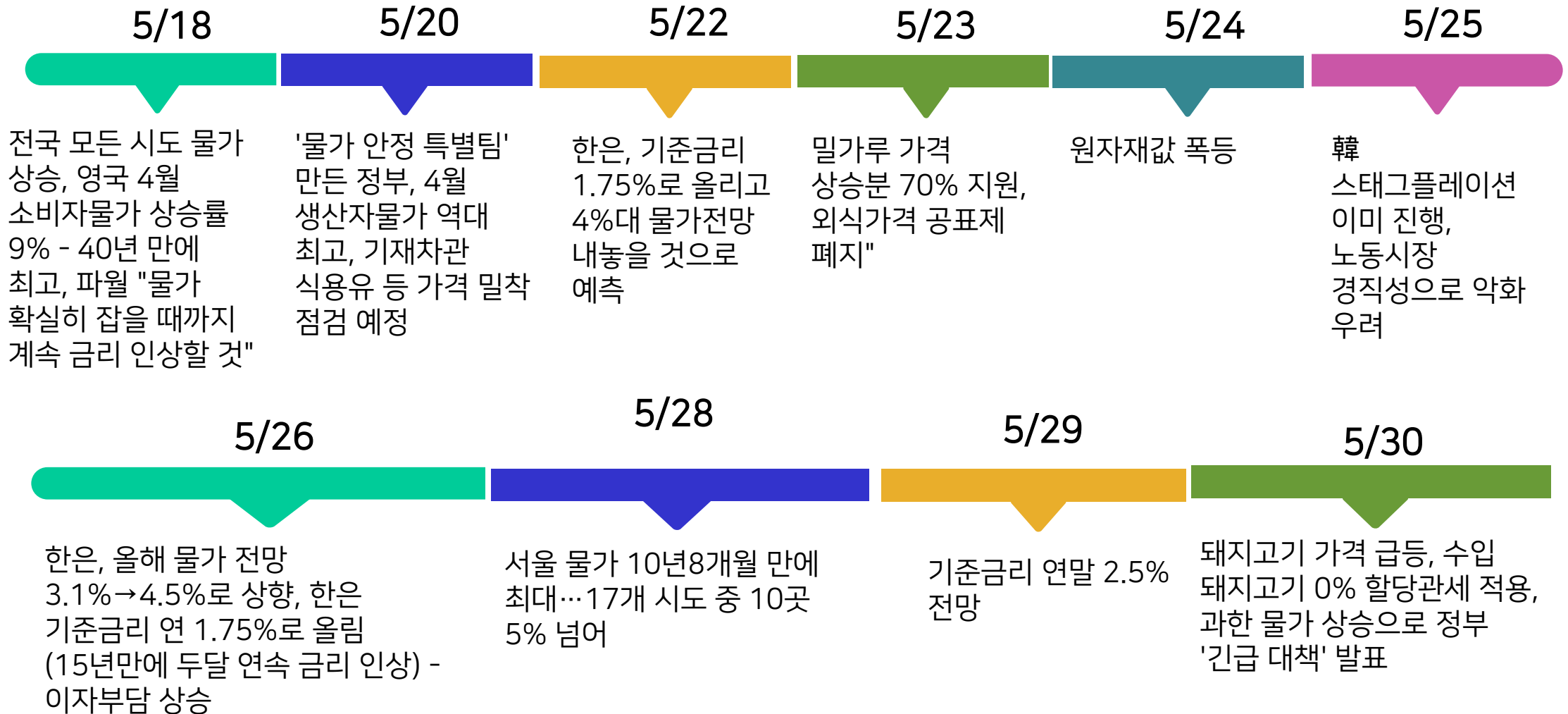
2022-05-01
5월 증시도 '고물가·긴축' 경계..."코스피 바닥 2,600"
바닥 밑에 지하실' 충격의 뉴욕증시...월가 "5월에 팔고 11월 이후에 사라"
추경호 "국가채무 엄격한 관리 필요...상속세 인적공제 확대"
[전기료 폭등전야②]한전 독점 전력시장 열린다...요금 오를까 내릴까
5월 증시도 쉽지 않다...증권사 "코스피 바닥 2600, 붕괴 위험도"
2022-05-02
치솟는 물가 속...한우도 킹크랩도 '30% 할인'에 우르르
[증시전략] 연준 5월 FOMC 회의에 쏠린 눈...'빅스텝' 시작하나
한덕수 "서민경제 안정 최우선 과제... 모든 정책수단 총동원할 것"
홈플러스 '물가 안정 프로젝트' 100일...온라인 매출 24%↑
추경호 '민간 주도 성장' 박차...법인세 인하로 경쟁력 키운다
2022-05-03
4월 소비자물가 4.8% ↑ ... 금융위기 이후 13년 반 만에 최고
4월 소비자물가 4.8%로 5% 육박...2008년 금융위기 이후 13년6개월만에 최
홍남기 "당분간 물가 상승압력 지속...마지막까지 안정 총력"
[포토]금융위기 이후 최고 수준으로 물가 상승률 최고 수준

- 각 날자 별로 '물가'와 가장 관련성 높은 기사를 clustering을 통해 5개씩 추출함
- CSV 파일로 정리 후 일주일 단위로 가장 중요해 보이는 기사를 직접 골라 event timeline 작성

5. 키워드 관련 주요 이슈 : “물가”



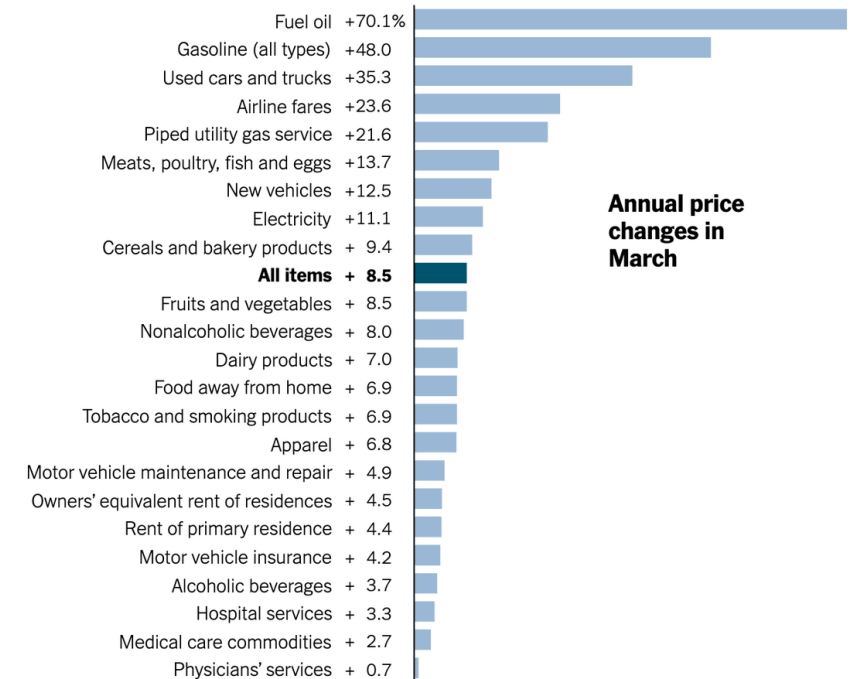
5. 키워드 관련 주요 이슈 : "물가"



6. 뉴스 데이터를 이용한 유가 예측

주제 선정 이유 및 키워드 설명

- 최근 휘발유 값의 급등과 관련해 자주 뉴스를 접할 수 있었음
- 인플레이션과 유가, 전쟁과 유가가 관련이 있을 것이라는 가설을 세우고 주제를 선정하게 됨
- 앞서 수집한 전쟁, 인플레이션, 환율 기사에 유가 관련 기사를 추가적으로 수집해서 유가의 등락을 예측해 보고자 함.
- 키워드에 해당하는 기사:
 - 키워드 1. 유가: 유가, 기름 값, 휘발유
 - 키워드 2. 전쟁: 전쟁, 러시아, 우크라이나
 - 키워드 3. 인플레이션: 인플레이션, 금리, 물가
 - 키워드 4. 환율: 환율, 달러



Year-over-year changes in a selection of categories of the Consumer Price Index • Source: Bureau of Labor Statistics • By The New York Times

6. 뉴스 데이터를 이용한 유가 예측

회귀분석 식

$$Y = a + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + \varepsilon$$

Y = daily oil price

네이버 금융의 휘발유 가격 (5월 간)

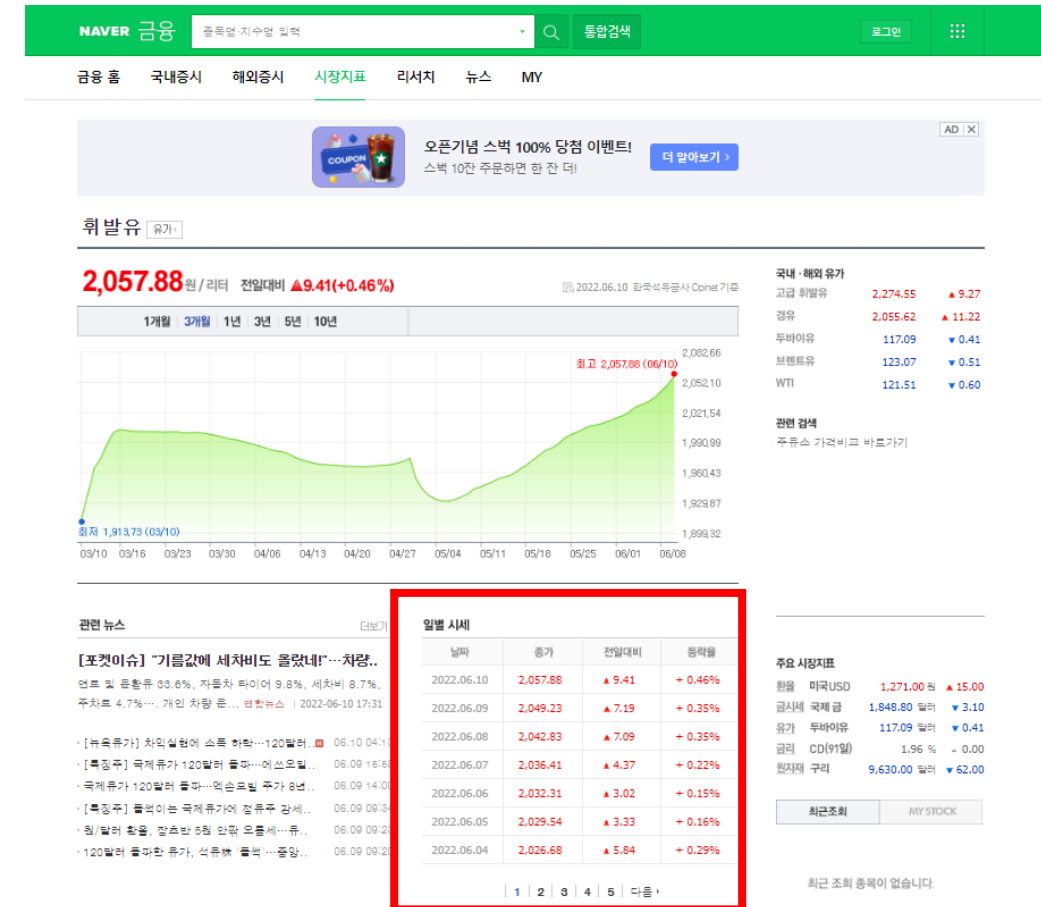
X₁ = 유가 기사 daily sentiment average

X₂ = 전쟁 기사 daily sentiment average

X₃ = 인플레이션 기사 daily sentiment average

X₄ = 환율 기사 daily sentiment average

Sentiment analysis를 이용한 daily sentiment average와 daily oil price의 multiple linear reg. regression

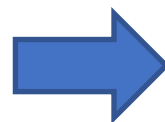


6. 기사로 유가의 등락을 예측할 수 있을까?

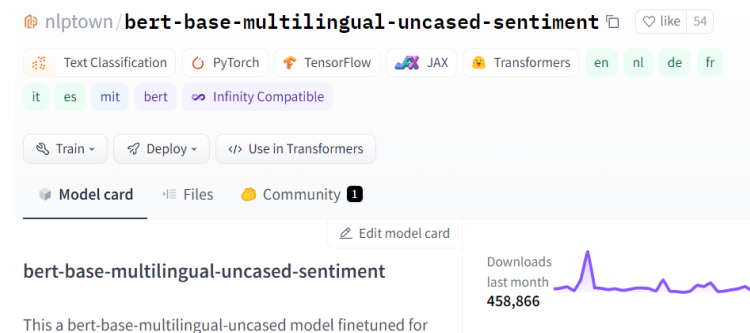


감성분석 모델 선정

- 기사 감성 분석을 위해 Huggingface에서 Korean sentiment 11개 모델 선정
- 자체 테스트 진행 후 모델 2개 최종 선정
- 모델 리스트:
 - 'snunlp/KR-FinBert-SC'
 - 'monologg/koelectra-small-finetuned-sentiment'
 - 'monologg/koelectra-base-finetuned-sentiment'
 - 'monologg/koelectra-small-finetuned-goemotions'
 - 'monologg/bert-base-cased-goemotions-original'
 - 'monologg/bert-base-cased-goemotions-ekman'
 - 'monologg/bert-base-cased-goemotions-group'
 - 'monologg/koelectra-base-finetuned-nsmc'
 - 'beomi/kcbert-base'
 - 'nlptown/bert-base-multilingual-uncased-sentiment'
 - 'philschmid/distilbert-base-multilingual-cased-sentiment-2'



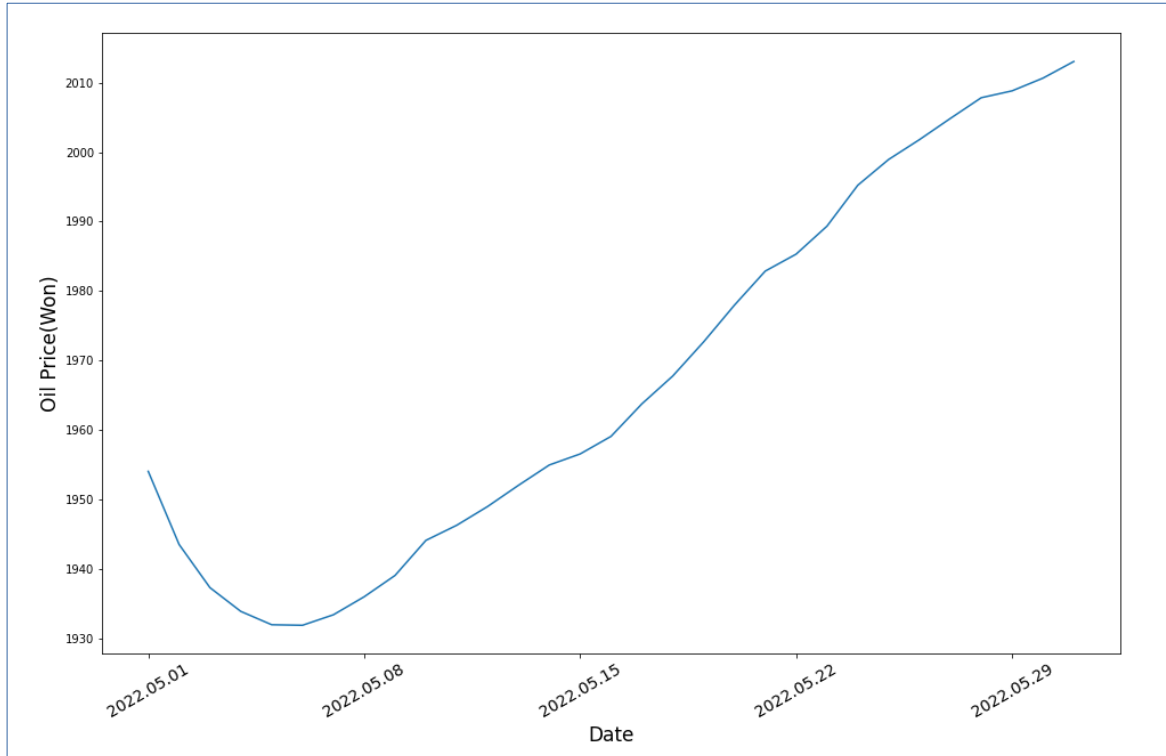
- 최종 선정 모델
 1. 'snunlp/KR-FinBert-SC'
 2. Multilingual based인 'nlptown/bert-base-multilingual-uncased-sentiment'



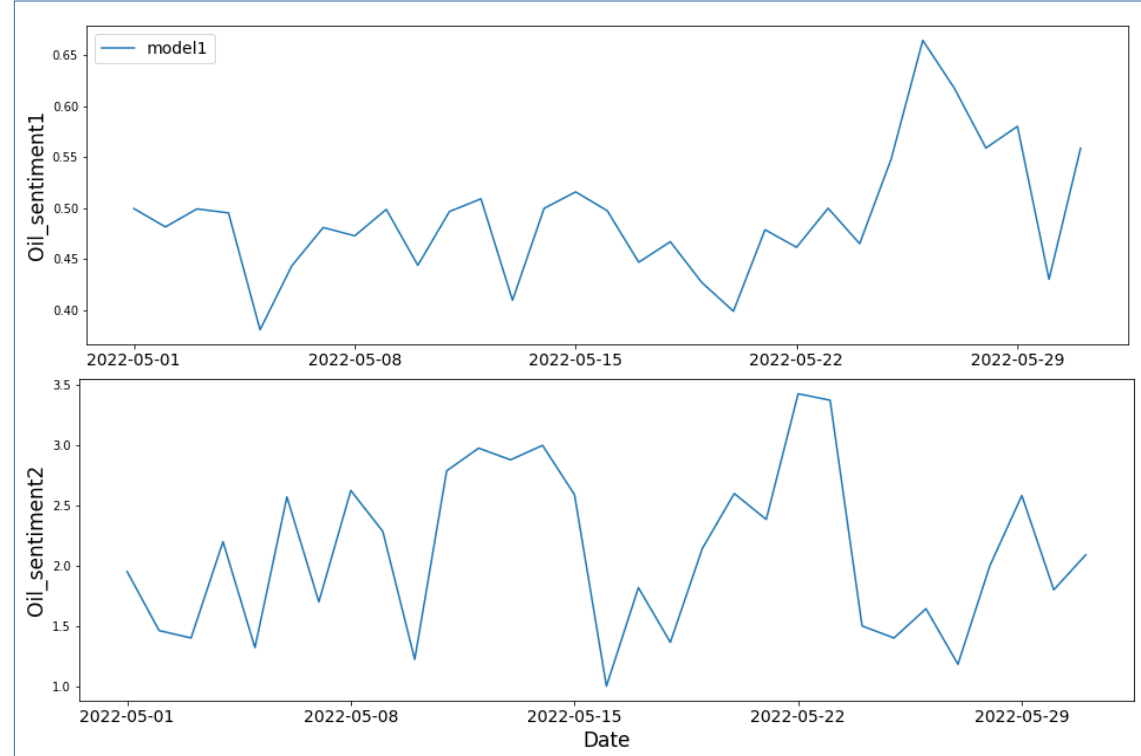
- 타 모델들은 긍부정 구분이 잘 안 되거나 한 쪽으로 쏠리는 경향을 보임
- 소셜미디어, 리뷰데이터 분석에 더 적합

6. 뉴스 데이터를 이용한 유가 예측

유가 Time Trend



유가 Daily Avg. Sentiment (sentiment model 1, 2)

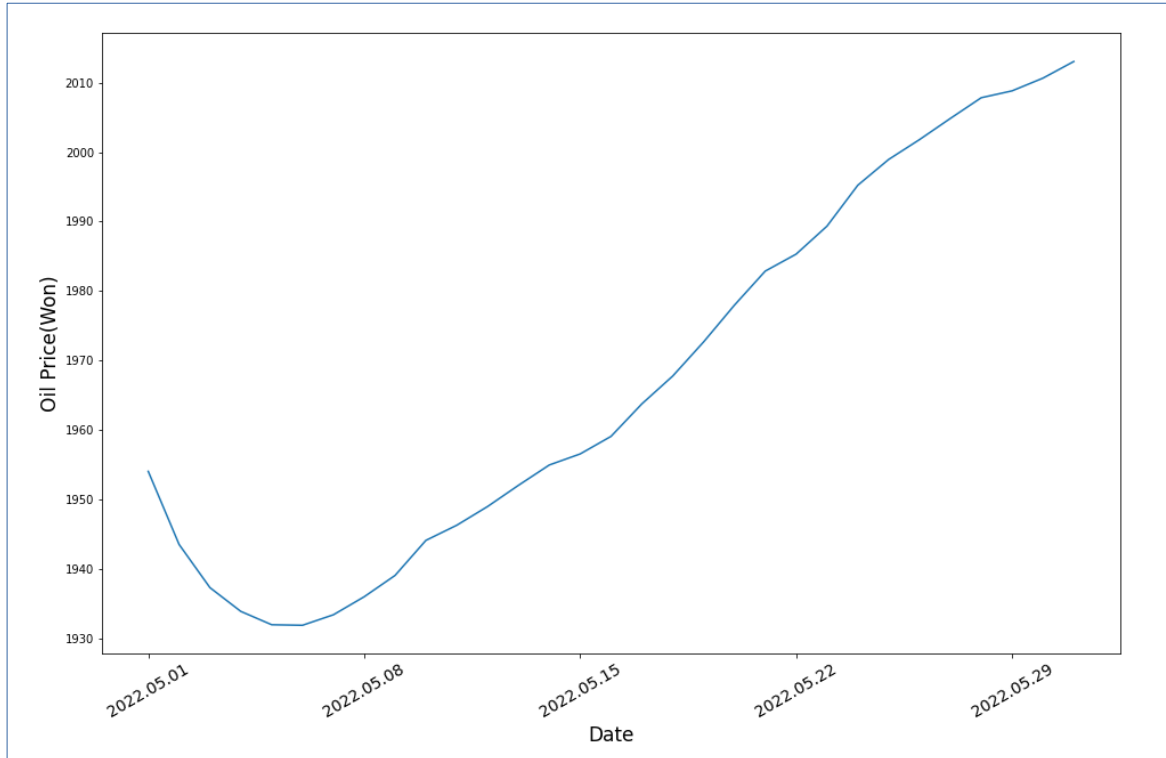


주안점

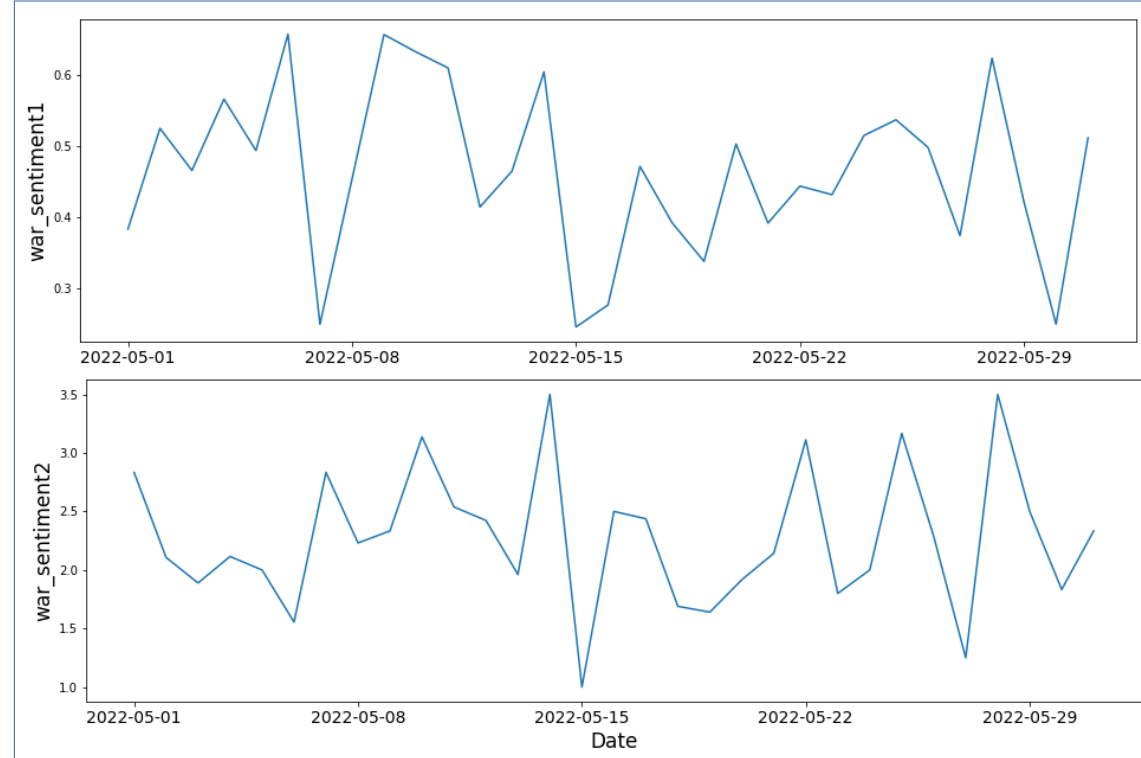
- Daily average sentiment의 trend가 모델에 간에 어느정도 유사성을 띄나 동일하지는 않음
- 같은 data를 사용했음에도 사용된 sentiment 모델에 따라 daily average sentiment 값이 다르게 나옴
- 유가, 인플레이션 sentiment의 경우 유가가 5월간 지속적으로 상승하고 있음에도 무조건 적으로 악화하는 trend는 보이지 않음

6. 뉴스 데이터를 이용한 유가 예측

유가 Time Trend



전쟁 Daily Avg. Sentiment (sentiment model 1, 2)

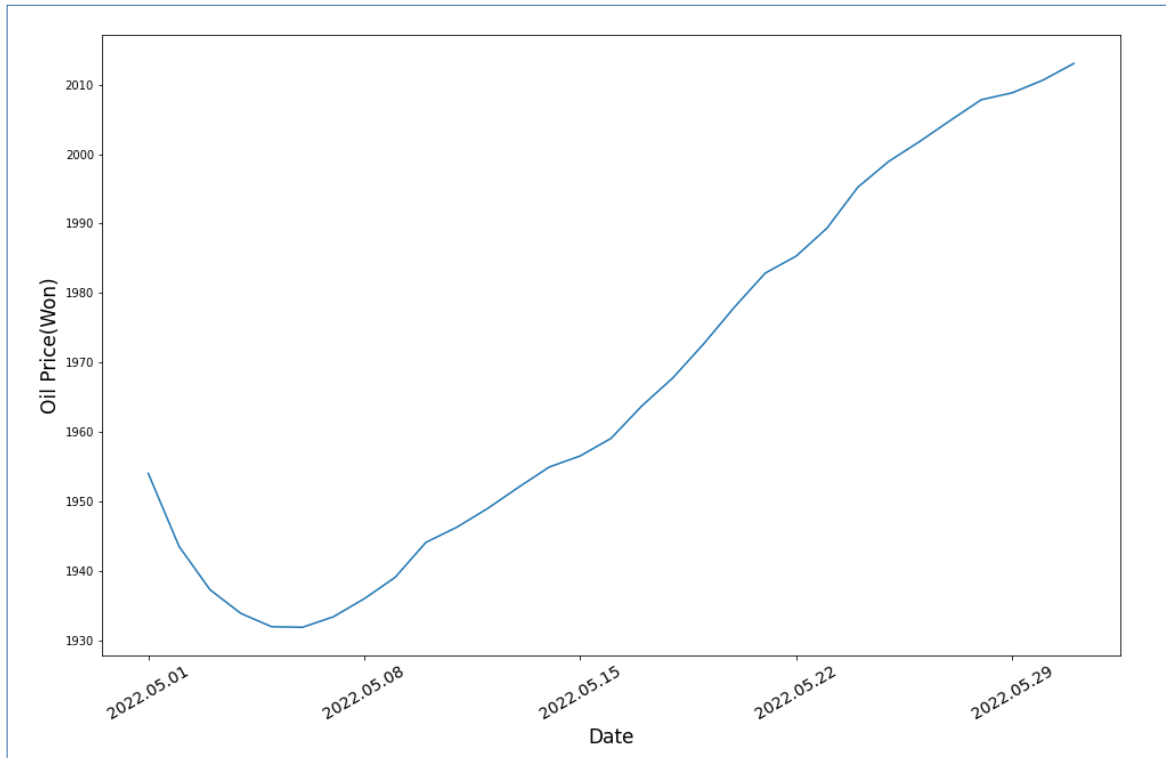


주안점

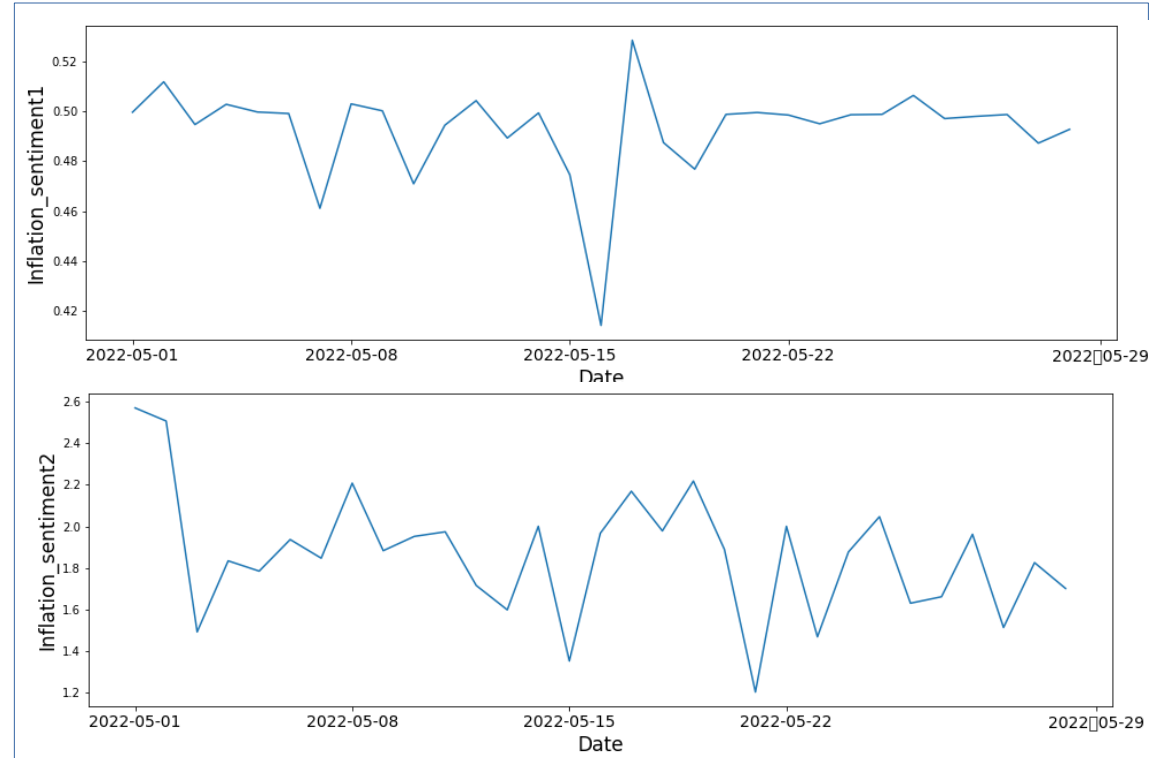
- Daily average sentiment의 trend가 모델에 간에 어느정도 유사성을 띄나 동일하지는 않음
- 같은 data를 사용했음에도 사용된 sentiment 모델에 따라 daily average sentiment 값이 다르게 나옴
- 유가, 인플레이션 sentiment의 경우 유가가 5월간 지속적으로 상승하고 있음에도 무조건 적으로 악화하는 trend는 보이지 않음

6. 뉴스 데이터를 이용한 유가 예측

유가 Time Trend



인플레이션 Daily Avg. Sentiment (sentiment model 1, 2)

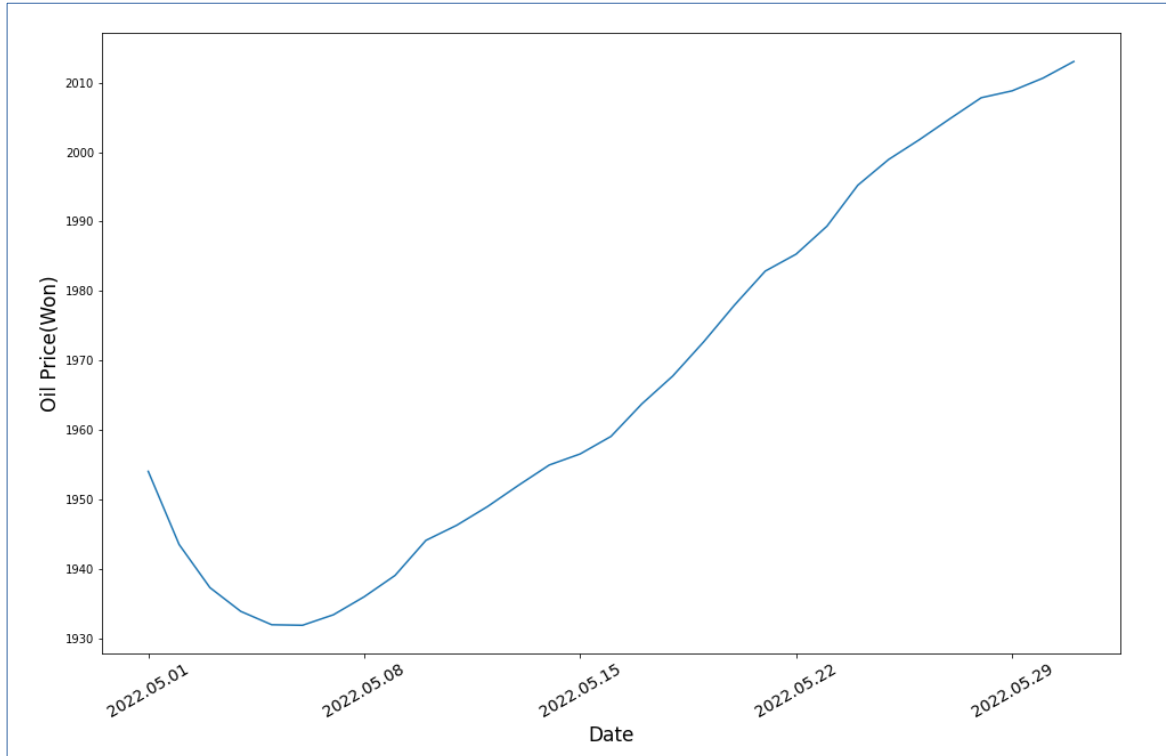


주안점

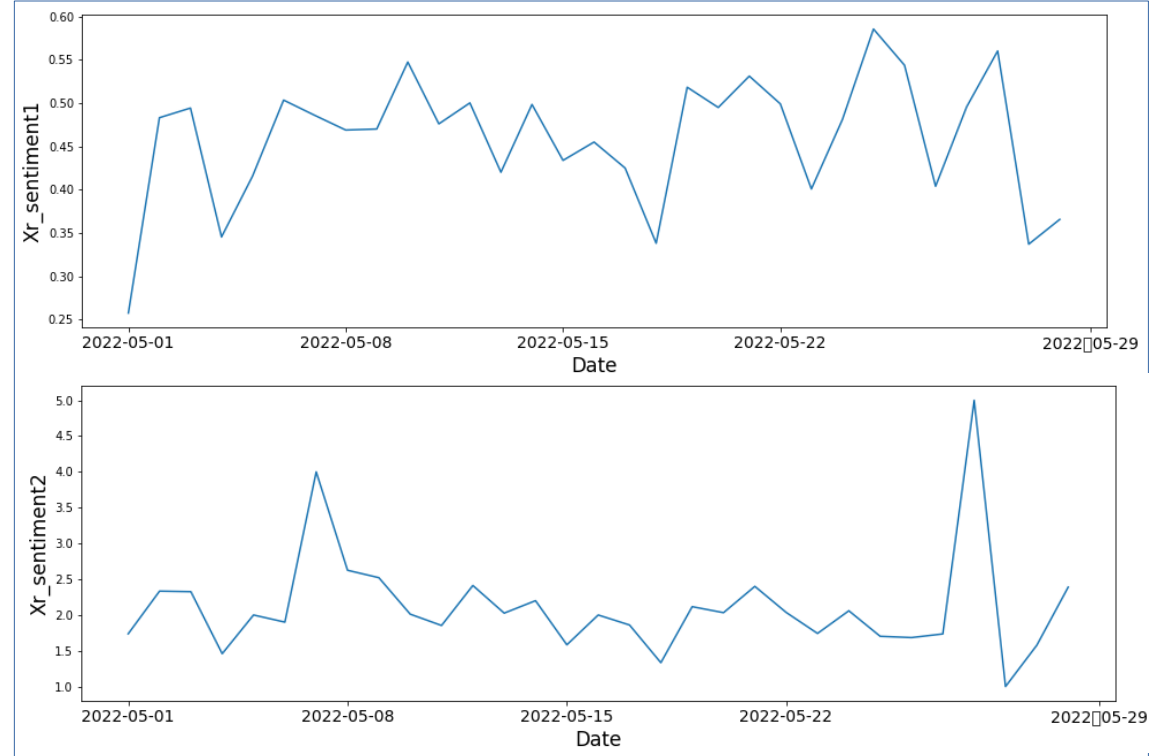
- Daily average sentiment의 trend가 모델에 간에 어느정도 유사성을 띄나 동일하지는 않음
- 같은 data를 사용했음에도 사용된 sentiment 모델에 따라 daily average sentiment 값이 다르게 나옴
- 유가, 인플레이션 sentiment의 경우 유가가 5월간 지속적으로 상승하고 있음에도 무조건 적으로 악화하는 trend는 보이지 않음

6. 뉴스 데이터를 이용한 유가 예측

유가 Time Trend



환율 Daily Avg. Sentiment (sentiment model 1, 2)



주안점

- Daily average sentiment의 trend가 모델에 간에 어느정도 유사성을 띄나 동일하지는 않음
- 같은 data를 사용했음에도 사용된 sentiment 모델에 따라 daily average sentiment 값이 다르게 나옴
- 유가, 인플레이션 sentiment의 경우 유가가 5월간 지속적으로 상승하고 있음에도 무조건 적으로 악화하는 trend는 보이지 않음

6. 뉴스 데이터를 이용한 유가 예측

OLS Regression Results			
Dep. Variable:	유가	R-squared:	0.296
Model:	OLS	Adj. R-squared:	0.188
Method:	Least Squares	F-statistic:	2.736
Date:	Sat, 11 Jun 2022	Prob (F-statistic):	0.0504
Time:	11:17:10	Log-Likelihood:	-141.00
No. Observations:	31	AIC:	292.0
Df Residuals:	26	BIC:	299.2
Df Model:	4		
Covariance Type: nonrobust			

	coef	std err	t	P> t	[0.025	0.975]
const	1744.1445	133.807	13.035	0.000	1469.101	2019.188
유가_sentiment1	209.5842	76.039	2.756	0.011	53.285	365.884
전쟁_sentiment1	-68.9113	46.271	-1.489	0.148	-164.023	26.200
인플레이션_sentiment1	284.2691	274.133	1.037	0.309	-279.220	847.758
환율_sentiment1	26.1629	65.930	0.397	0.695	-109.357	161.683
Omnibus:	2.283	Durbin-Watson:		0.544		
Prob(Omnibus): 0.319 Jarque-Bera (JB): 1.365						
Skew:	0.211	Prob(JB):		0.505		
Kurtosis:	2.063	Cond. No.		93.7		

결과

- snunlp/KR-FinBert-SC을 사용하여 추출한 sentiment를 통한 regression 결과임
- 다소 낮은 R^2 값(0.296)이 관측 되었음
- Coefficient 중 유가_sentiment를 제외하고는 weak statistical significance를 보여주고 있음
- (t-value < 2)
- 이 결과만 봤을 경우 생각보다 전쟁이나 기타 에너지 가격과 직접적으로 연관되어 있지 않은 주제에 대한 sentiment는 유가prediction에 큰 도움이 되지 않는 것으로 판단됨

6. 뉴스 데이터를 이용한 유가 예측

OLS Regression Results

Dep. Variable:	유가	R-squared:	0.111
Model:	OLS	Adj. R-squared:	-0.026
Method:	Least Squares	F-statistic:	0.8125
Date:	Sat, 11 Jun 2022	Prob (F-statistic):	0.529
Time:	11:17:51	Log-Likelihood:	-144.62
No. Observations:	31	AIC:	299.2
Df Residuals:	26	BIC:	306.4
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	2026.0125	42.290	47.908	0.000	1939.085	2112.940
유가_sentiment2	-5.2184	7.927	-0.658	0.516	-21.512	11.075
전쟁_sentiment2	11.4442	10.468	1.093	0.284	-10.072	32.961
인플레이션_sentiment2	-31.2267	18.719	-1.668	0.107	-69.703	7.250
환율_sentiment2	-20.0374	26.677	-0.751	0.459	-74.873	34.798
Omnibus:	5.008	Durbin-Watson:	0.240			
Prob(Omnibus):	0.082	Jarque-Bera (JB):	1.847			
Skew:	0.145	Prob(JB):	0.397			
Kurtosis:	1.840	Cond. No.	35.5			

결과

- nlptown/bert-base-multilingual-uncased-sentiment를 사용했을 시 결과
- 낮은 R^2 값(0.111)이 관측 되었음
- 인플레이션 관련 sentiment의 coefficient만 90%이상의 level of significance를 지니고 있음
- 앞에서 다른 sentiment analysis model을 사용했을 때와 다른 결과가 도출되었으며 인플레이션의 coefficient가 negative인 점에서 결과 값에 의문이 있음.

6. 뉴스 데이터를 이용한 유가 예측

OLS Regression Results

Dep. Variable:	유가	R-squared:	0.111
Model:	OLS	Adj. R-squared:	-0.026
Method:	Least Squares	F-statistic:	0.8125
Date:	Sat, 11 Jun 2022	Prob (F-statistic):	0.529
Time:	11:17:51	Log-Likelihood:	-144.62
No. Observations:	31	AIC:	299.2
Df Residuals:	26	BIC:	306.4
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	2026.0125	42.290	47.908	0.000	1939.085	2112.940
유가_sentiment2	-5.2184	7.927	-0.658	0.516	-21.512	11.075
전쟁_sentiment2	11.4442	10.468	1.093	0.284	-10.072	32.961
인플레이션_sentiment2	-31.2267	18.719	-1.668	0.107	-69.703	7.250
환율_sentiment2	-20.0374	26.677	-0.751	0.459	-74.873	34.798
Omnibus:	5.008	Durbin-Watson:	0.240			
Prob(Omnibus):	0.082	Jarque-Bera (JB):	1.847			
Skew:	0.145	Prob(JB):	0.397			
Kurtosis:	1.840	Cond. No.	35.5			

한계 및 개선 점

- 5월 한달 간의 짧은 기간 내의 데이터를 사용하여 regression analysis를 진행하였기 때문에 여러 유의미한 control variable들을 (e.g. GDP, interest rate, etc) 모델에 활용할 수 없었음
- 사용한 Sentiment Analysis Model에 따라 상이한 결과 값이 도출되었음 → 비정형 데이터 ML tool을 다룰 때 결과의 신뢰도 면에서 유의할 필요가 있음

감사합니다