

텍스트 마이닝 프로젝트

메타버스 감성분석 및 공포-탐욕 지표 제안

손우진, 김홍민, 장주현

1. Introduction
2. experimental method
3. Data
4. Indicators
5. Fear And Greed Indicator

1. Introduction

불안한 시장 정세

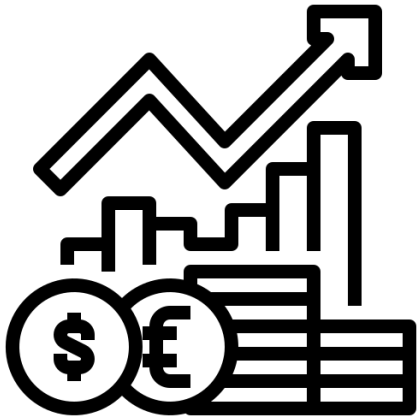
- 1986년 일본 “버블 경제 사태”
- 2차 세계대전 이후 최악, 최대규모의 미국 금융 위기로 불렸던 “서브프라임 모기지 사태”
- 최근 세계 3위 암호화폐 거래소 FTX의 “뱅크런”
- 이 밖에도 연속되는 인플레이션 현상
- 자꾸만 들려오는 전쟁에 대한 우려 등

국내 금융시장의 특징

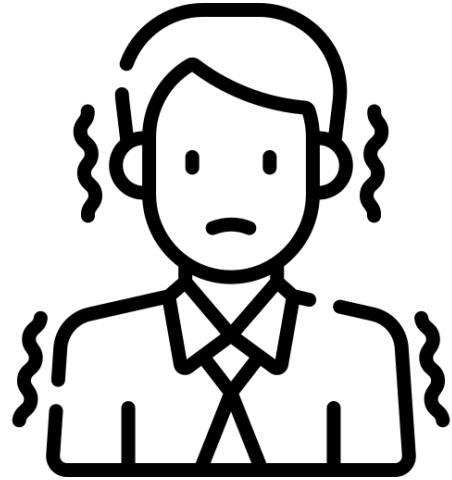
- 해외 시장에 비해 외부적 요소에 크게 휘둘리는 시장
- 코로나 이후 한탕주의성 신규 투자자 증가

1. Introduction

금융 시장에는 많은
거시적 지표 존재



국내 소비자들의
감성을 나타내는 지표가 X



국내 시장에 대한
공포심과 탐욕감을 보여주는
Fear And Greed 지표 제안

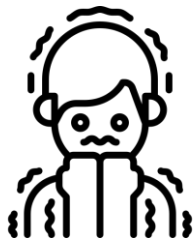
1. Introduction

Fear And Greed ?

- 단순한 감성분석 기반의 결과는 선행 지표로 활용하기엔 부족하다 생각
- 공포-탐욕 지수는 감성분석 결과와 거시적 지표들의 움직임을 결합해 최종 산출

0에 가까우면 -> 공포스러움

100에 가까우면 -> 탐욕적임



0



50



100

0. 주제 선정

본 연구는 “공포-탐욕 지표”를 제안하면서 최근 시장에서 많은 관심을 받은 **메타버스**를 메인 키워드로 실험을 제안할 예정 (관심이 쏠리는 곳에 자금의 유입이 따라오기 마련)

1. 데이터 수집

- “빅카인즈”에서 제공하는 뉴스 데이터
- “네이버 종목토론방” 제목 내용
- 파이썬 Finance DataReader 패키지 기반 메타버스 관련 주 + 거시적 지표 사용

2. experimental method

2. 감성 사전 구축

개별 종목 별 뉴스 데이터의 제목과 본문의 키워드 토큰에 대한 감성 사전을 구축

3. Fear And Greed 지표 산출

개별 종목 별로 구축된 감성 사전을 기반으로

당일 뉴스기사와 종목토론방의 감성 Score 평균에 거시적 지표를 결합하여 최종 지표 산출

3. Data

1. 빅카인즈 뉴스 데이터

연관 키워드가 속한 뉴스 데이터 중에서 중복된 기사를 제외하고 1년간의 데이터 : 31,523개

2. 종목 별 등락 여부 데이터

전일 대비 상승 : 1 / 하락 : 0

3. 종목 별 일일 거래량 데이터

당일 거래량 / 유통 주식 수 : 실제 상장 주식 수 - 특수 관계인의 보유 주식 수

4. 거시적 지표 데이터

코스피, S&P500 VIX, 환율

4. Indicators (감성 스코어)



뉴스의 제목과 본문의 키워드에 대한
감성 사전 구축

감성사전 단어 수

제목 : 13,914개

본문 : 25,275개

4. Indicators (감성 스코어)

Example : 칩스앤미디어

```
'코인': 4.122999627837984,  
'연대': 0.8405284704131002,  
'전략': 1.7714923706735715,  
'마케팅': 3.6455154447339027,  
'학회': 2.926870115370302,  
'산학': -0.7785634536657988,  
'협력': -8.431149981391917,  
'프로젝트': 4.953293636025292,  
'시시비비': -1.4784145887606994,  
'디지털': -18.677893561592427,  
'경쟁력': 1.5591737997767026,  
'정책': -11.292519538518818,  
'커머스': 9.964458503907686,  
'노미': 5.245813174544096,  
'공공기관': 3.2026423520655003,  
'교육': -1.626163007071124,  
'메타': 5.126907331601111,  
'버스': 5.74190547078958,  
'근로': 2.2945664309639007,  
'복지': 4.2026423520654985,  
'공단': 1.178265723855602,  
'가상': 21.614625976926494,  
'인재': -7.695757350204721,  
'개발': 0.7679568291775298,  
'구축': 4.7847041310010505,
```

개별 키워드에 대한 Score는 전일 대비 등락 여부에 따라 계산

(예시)

1. 새로운 뉴스 등장
2. 해당 종목은 전일 대비 상승
3. 해당 뉴스에서 생성된 토큰은 긍정적 토큰으로 간주
4. 반대로 전일 대비 하락의 경우 부정적 토큰으로 간주
5. 1~4번의 과정을 모든 데이터에 대해 진행하여 최종 사전 구축

4. Indicators (감성 스코어)

Example : 칩스앤미디어

date	title	content	칩스앤미디어Change	title_score_칩스앤미디어
20221111	우영우와 울산고래 등 제페토로 새로 만나는 여행지들	우리나라의 보석같은 지역 관광지들이 메타버스로 태어났다 전국 개 인구감소지역을 포...	1.0	2.197242
20221111	무릎관절통 한방에 날리는 방법은	우리는 일상생활을 통해 생명활동에 필요한 생체에너지를 생성하고 순환시키며 균형 조절...	1.0	0.992835
20221111	힘 실린 증시 바닥론 성장주 날았다	지난달 미국 소비자물가지수 상승률이 시장 예상치를 밑돌며 증시 바닥론이 고개를 들자...	1.0	-0.264091
20221111	고객 돈도 위험 베팅에 대출 조사 착수	파산 위기에 몰린 글로벌 대형 암호화폐 거래소 가 고객 돈을 빼내 알라메다리서치를 ...	1.0	0.005737
20221111	베일 벗는다 오늘 개최 특별 포인트 넷	가 뜨거운 관심 속에 개최된다 문화체육관광부와 한국연예제작자협회는 일과 일 일산...	1.0	-0.043931

전일 대비 상승 데이터: 0.5013026648801548
전일 대비 하락 데이터: 0.49869733511984515

사전 구축 시 설명력을 위해
데이터 분포를 맞춤

감성사전 기반 Score

4. Indicators (거래 회전율)

종목 별 일일 거래회전율 데이터 : 당일 거래량 / 유통 주식 수

유통 주식 수 : 실제 상장 주식 수 - 특수 관계인의 보유 주식 수

Example : Naver

가. 최대주주 및 특수관계인의 주식소유 현황

(기준일 : 2022년 09월 30일)

(단위 : 주, %)

성 명	관 계	주식의 종류	소유주식수 및 지분율				비고
			기 초		기 말		
			주식수	지분율	주식수	지분율	
국민연금공단	본인	보통주	0	0	0	0	-
국민연금기금	특수관계인	보통주	14,670,489	8.94	13,179,405	8.03	장내매도
계		보통주	14,670,489	8.94	13,179,405	8.03	장내매도
		-	-	-	-	-	-

상장 주식 수 : 164,049,085개
특수 관계인 보유 수 : 14,670,489개

유통 주식 수 : 149,378,596개

5. Fear And Greed

Fear And Greed

1. 감성 사전 기반으로 산출된 Score
2. 일일 거래 회전율
3. Kospi 200, S&P500 VIX, 환율(KRW / USD)

위 지표들을 결합하여 최종 Fear And Greed Score를 산출 할 예정