

# 현장프로젝트교과 기말발표

---

7 팀

2018312021 유창연

2018312031 장성호

2020312048 박진환

# CONTENTS

---

01 Introduction

02 Crawling

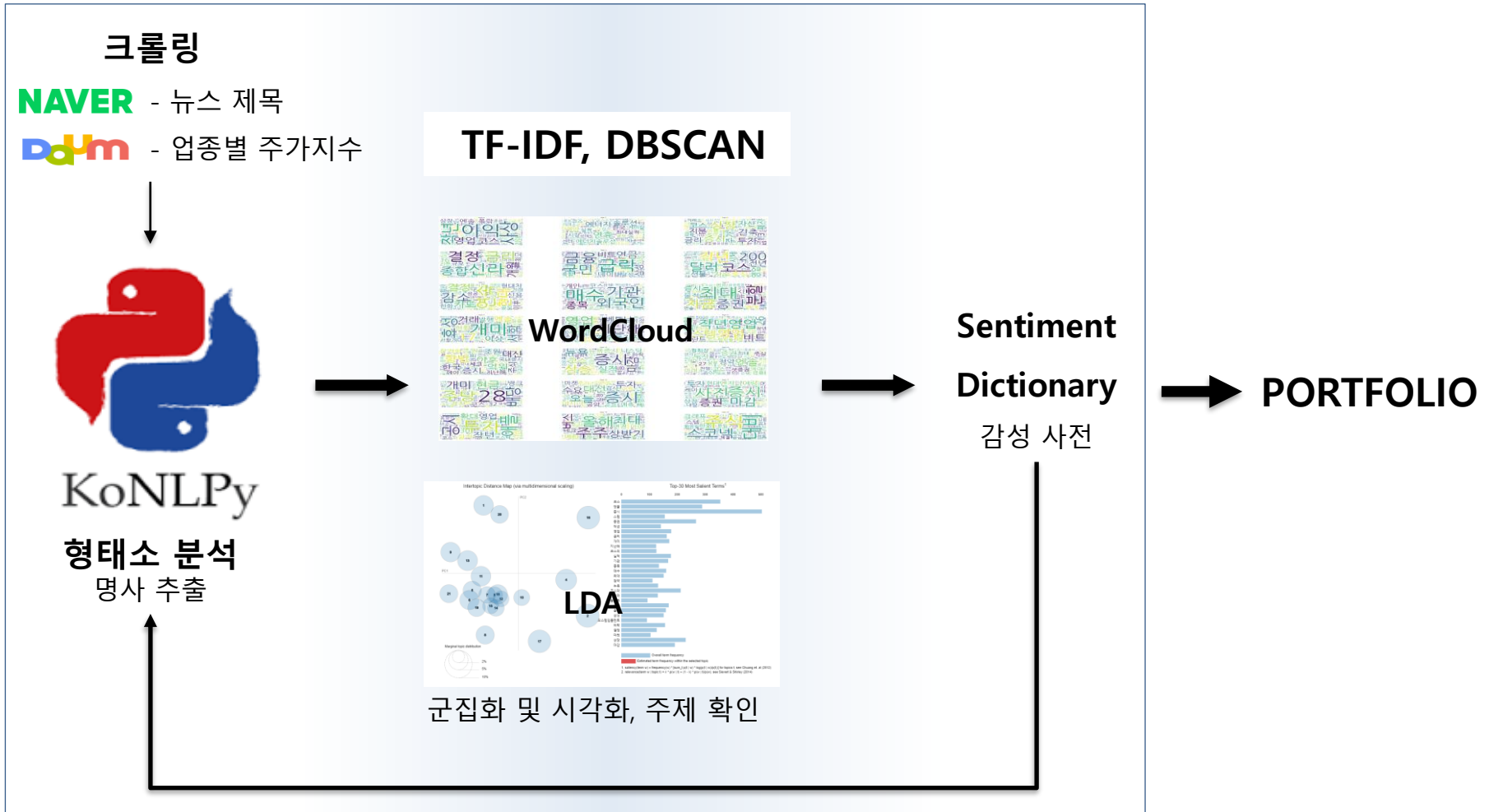
03 EDA

04 Portfolio

05 Conclusions & Future research

# 01. Text Mining을 이용한 포트폴리오 가중치 업데이트

## Flow Chart



# 01 Introduction

## 선행연구

| 저자          | 연구 방향  | 한계점   |
|-------------|--|---|
| 안성원 외(2010) | BoW, Naïve Bayesian 알고리즘 사용 후,<br>시계열 모델을 사용해 주가 예측        | 시장 전반에 대한 긍,부정 분석을 통해<br>단기 분석(19일, 1,582건)만을 진행,<br><br>모델 상승, 하락 예측 결과<br>F-Score가 55%정도로 부정확<br><br>수집된 단어를 랜덤으로 추출하여<br>감성사전 제작 |
| 조혜진 외(2013) | 한국어 역접관계를 고려한<br>오피니언 반의법 규칙 알고리즘 제안,<br>코스피 지수를 활용한 평판 분석 |   |
| 김다예 외(2018) | Word2Vec 알고리즘 사용,<br>주가 방향성 예측용 감성사전 제안                    |   |

### 한계점 해결 방안으로

- 1) 장기적인 긍,부정 분석을 통한 검증
- 2) 여러 분야에 대한 검증
- 3) 주식 가격 자체 예측이 아닌 포트폴리오에 News Score를 반영한 포트폴리오 제안

# 01 Introduction

## 업종 소개

| 업종    | 업종코드     | 업종   | 업종코드     | 업종    | 업종코드     | 업종   | 업종코드     |
|-------|----------|------|----------|-------|----------|------|----------|
| 음식료품  | D0011005 | 철강금속 | D0011011 | 유통업   | D0011016 | 금융업  | D0011021 |
| 섬유의복  | D0011006 | 기계   | D0011012 | 전기가스업 | D0011017 | 은행   | D0011022 |
| 종이목재  | D0011007 | 전기전자 | D0011013 | 건설업   | D0011018 | 증권   | D0011024 |
| 화학    | D0011008 | 의료정밀 | D0011014 | 운수창고  | D0011019 | 보험   | D0011025 |
| 의약품   | D0011009 | 운수장비 | D0011015 | 통신업   | D0011020 | 서비스업 | D0011026 |
| 비금속광물 | D0011010 |      |          |       |          |      |          |

## 02 Crawling

### 데이터 수집 - 다음 업종별 주가지수



<https://finance.daum.net/api/charts/D0011024/months?limit=200&adjusted=true>

업종코드 월봉

업종별 종목은 음식료품, 섬유 의복, 종이 목재, 증권 등 총 21개 업종 월봉 차트 2013.01~2022.12 데이터를 API를 사용해 데이터 수집

## 02 Crawling

### 데이터 수집 - 네이버 업종별 뉴스

The screenshot shows the Naver News search interface. The search term is '증권' (Finance). The filters are set to '기간' (Date) as '2021.01.01. ~ 2021.01.01.' and '유형' (Type) as '전체' (All). The search results show several news articles. Two articles are highlighted with red boxes:

- 금감원, '라임펀드' 판매사 KB증권 투자 피해자에 최대 70% 배상**  
금융감독 당국이 '라임펀드' 판매사인 KB증권의 손해배상책임을 인정하고 투자 피해자에게 투자금의 최대 70%를 배상토록 결정했다. 증권사는 고객을 조교위험성...
- KB증권, 라임펀드 투자자에 60~70% 배상**  
금융감독 당국이 '라임펀드' 판매사인 KB증권의 손해배상책임을 인정하고 투자 피해자에게 투자금의 최대 70%를 배상토록 결정했다. 증권사는 고객을 조교위험성...

[https://search.naver.com/search.naver?where=news&query=증권](https://search.naver.com/search.naver?where=news&query=증권&sort=3&nso=so%3Ar%2Cp%3Afrom20210101to20210101&start=1)  
탭 리스트      검색어  
&sort=3&nso=so%3Ar%2Cp%3Afrom20210101to20210101&start=1  
관련도 순  
시작일, 종료일      페이지

카테고리인 21개의 종목에 대한 뉴스 제목을  
2013.01.01~2022.12.31까지 Beautiful Soup을 사용하여  
1일 10개, 10(개) \* 365(일) \* 10(년) \* 21(종) = 766,500개 수집

## 형태소 분석



Kkma.pos(News) : 형태소 분석

News : 작년 12월 건설업 체감경기'개선'

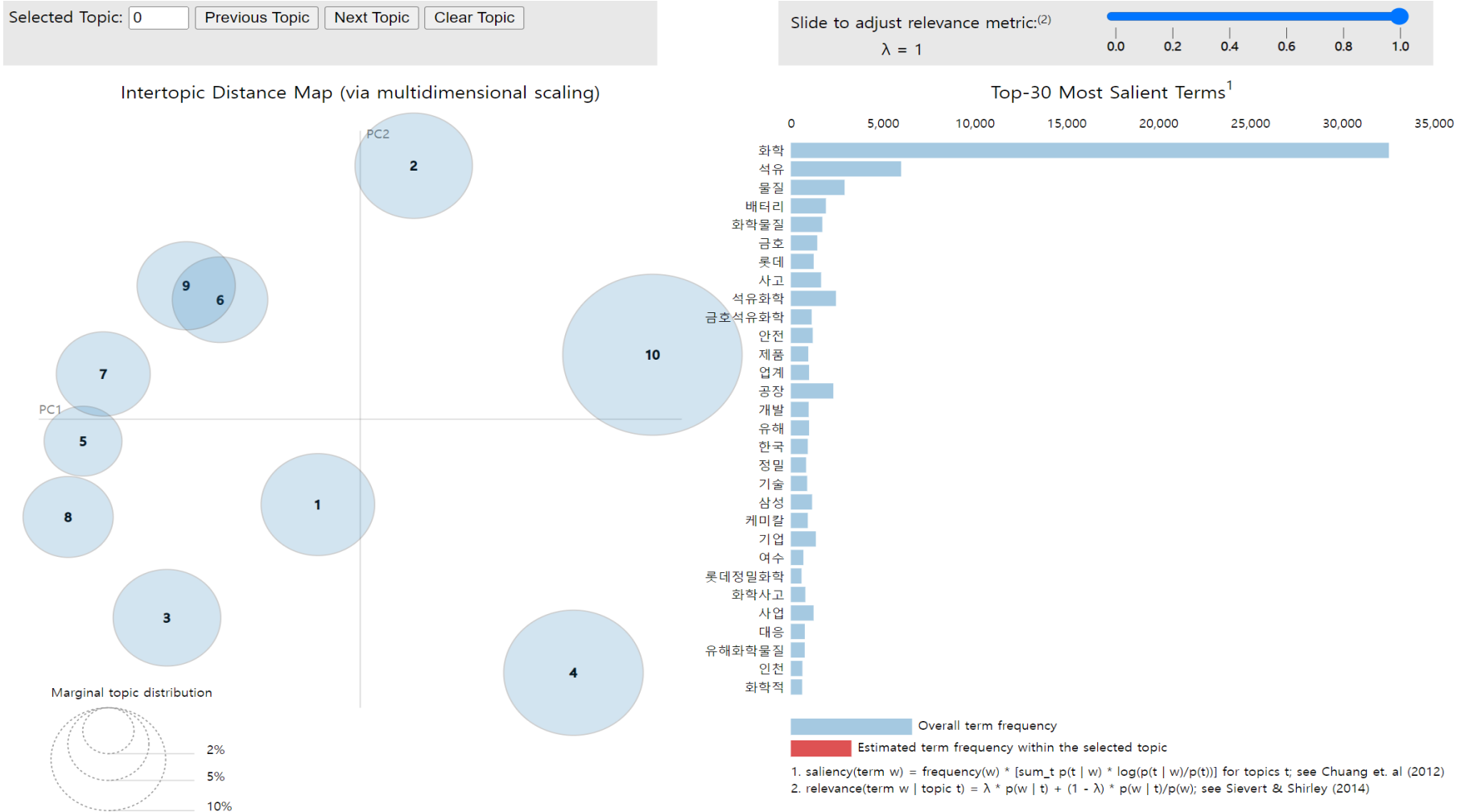
|     |                  |
|-----|------------------|
| 작년  | 'NNG' : 보통 명사    |
| 12  | 'NR' : 수사        |
| 월   | 'NNM' : 단위 의존 명사 |
| 건설업 | 'NNG' : 보통 명사    |
| 체감  | 'NNG' : 보통 명사    |
| 경기  | 'NNG' : 보통 명사    |
| ,   | 'SS' : 따옴표       |
| 개선  | 'NNG' : 보통명사     |

'NNG' 인 보통명사만 사용





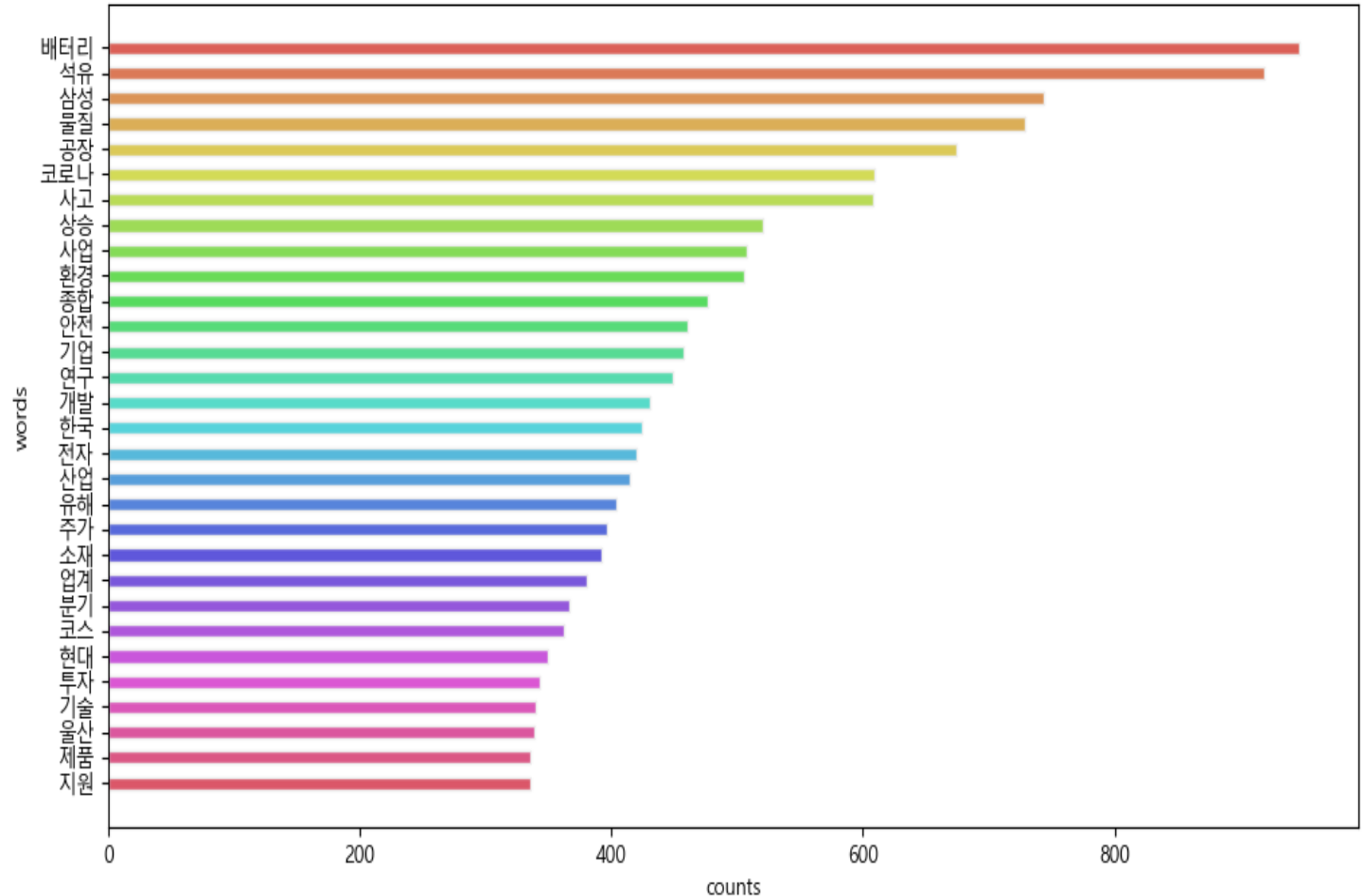
## LDA(Latent Dirichlet Allocation)



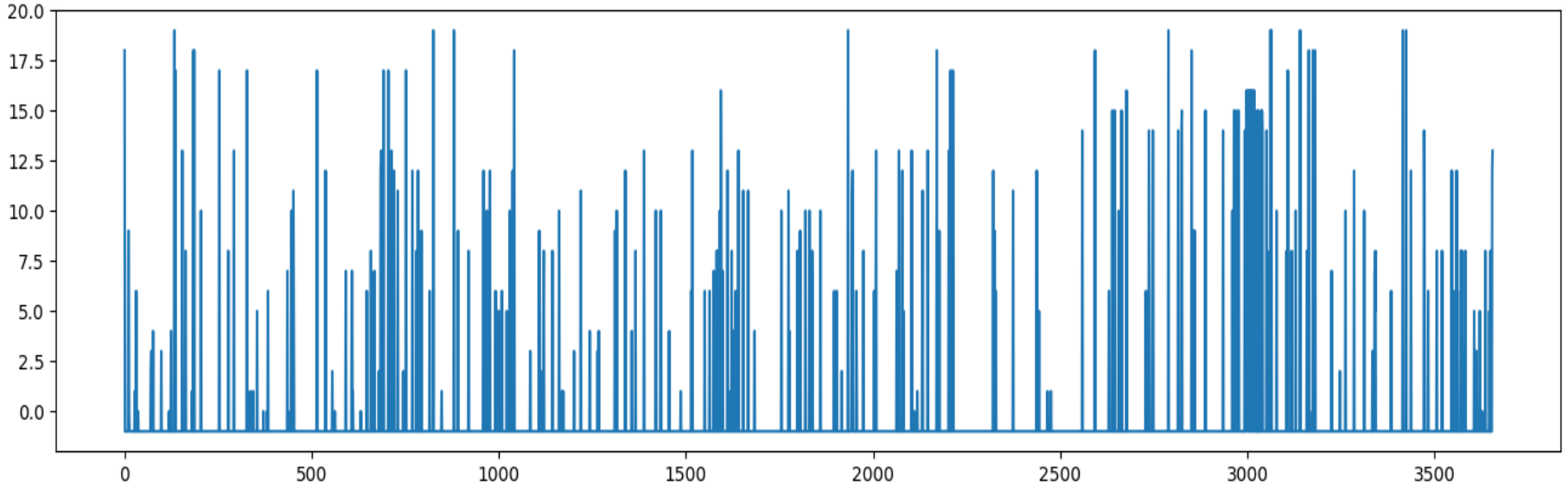
## TF-IDF

화학(2020년)

|     |            |
|-----|------------|
| 배터리 | 946.284797 |
| 석유  | 918.634142 |
| 삼성  | 743.379561 |
| 물질  | 729.223240 |
| 공장  | 674.354681 |
| 코로나 | 609.199684 |
| 사고  | 607.532828 |
| 상승  | 520.694615 |
| 사업  | 507.905758 |
| 환경  | 505.786314 |
| 종합  | 476.826615 |
| 안전  | 460.614837 |
| 기업  | 457.751140 |
| 연구  | 449.180686 |
| 개발  | 430.420172 |
| 한국  | 423.992572 |
| 전자  | 420.245882 |
| 산업  | 415.103137 |
| 유해  | 404.150170 |
| 주가  | 396.524747 |
| 소재  | 392.408191 |
| 업계  | 380.375320 |
| 분기  | 367.093384 |
| 코스  | 361.915745 |
| 현대  | 349.821954 |



## DBSCAN(밀도 기반 클러스터링)



| time           | title   | category | result |
|----------------|---|----------|--------|
| 25578 20200103 | LG화학, 전기차 배터리 분사설 '술술'...올해는 다를까?                 | 화학       | 1      |
| 25731 20200119 | LG화학, 전기차 배터리 '광복 행보'                             | 화학       | 1      |
| 25887 20200203 | LG화학, 전기차 배터리 힘입어 실적 개선 예고                        | 화학       | 1      |
| 25894 20200204 | DB금투 "LG화학, 전기차 배터리 기대감 높일 시점...매수의견"             | 화학       | 1      |
| 26161 20200302 | 中 주춤하자...LG화학, 전기차 배터리 CATL 제쳤다                   | 화학       | 1      |
| 26398 20200325 | LG화학, 전기차 배터리 '테슬라 특수' 기대... 日 파나소닉 제치고 최대 공급자 올라 | 화학       | 1      |
| 26720 20200427 | LG화학, 전기차 배터리에도 쓸 탄소나노튜브 1천200t 증설                | 화학       | 1      |
| 26723 20200427 | LG화학, 전기차 배터리 소재 탄소나노튜브 1200톤 증설                  | 화학       | 1      |
| 27037 20200528 | LG화학-SK이노, 불 붙은 전기차 배터리 소재 경쟁                     | 화학       | 1      |
| 27170 20200611 | 삼성증권 "LG화학, 전기차 배터리 흑자 임박...목표가↑"                 | 화학       | 1      |
| 27669 20200731 | LG화학, 전기차 배터리 드디어 흑자 냈다                           | 화학       | 1      |
| 28017 20200903 | [특징주] LG화학, 전기차 배터리 5달째 1위 소식에 '상승세'              | 화학       | 1      |
| 28026 20200904 | LG화학, 전기차 배터리 세계 1위 등극                            | 화학       | 1      |

| time           | title                                   | category | cluster |
|----------------|---|----------|---------|
| 25621 20200108 | 이란 '피의 보복' 개시...韓 정유·화학업계에 리스크 가중       | 화학       | 2       |
| 26106 20200225 | 정유·화학업계...LG화학 직원 코로나 검사에 '깜짝'          | 화학       | 2       |
| 26231 20200309 | 코로나19부터 유가 급락까지...'엮힌 데 뉘친' 정유·화학업계     | 화학       | 2       |
| 26242 20200310 | 유가 급락 속 마진도 줄어...정유·화학업계 '직격탄'          | 화학       | 2       |
| 26296 20200315 | [팬데믹 쇼크] 코로나·유가 '쇼크'...정유·화학업계 불확실성 최고조 | 화학       | 2       |
| 26526 20200407 | 정부, 정유·화학업계 석유수입·판매부과금 징수 유예            | 화학       | 2       |
| 26663 20200421 | "버텨야 한다"... 정유·화학업계 가동률 조정              | 화학       | 2       |
| 27467 20200710 | 정유·화학업계, 디지털 첨단 기술로 안전사고 막는다            | 화학       | 2       |
| 28285 20200930 | 정유·화학업계, 추석 맞아 '상생경영' 적극 나서             | 화학       | 2       |
| 28798 20201121 | 정유·화학업계 구조 전환 등 위기 극복 방안 모색             | 화학       | 2       |

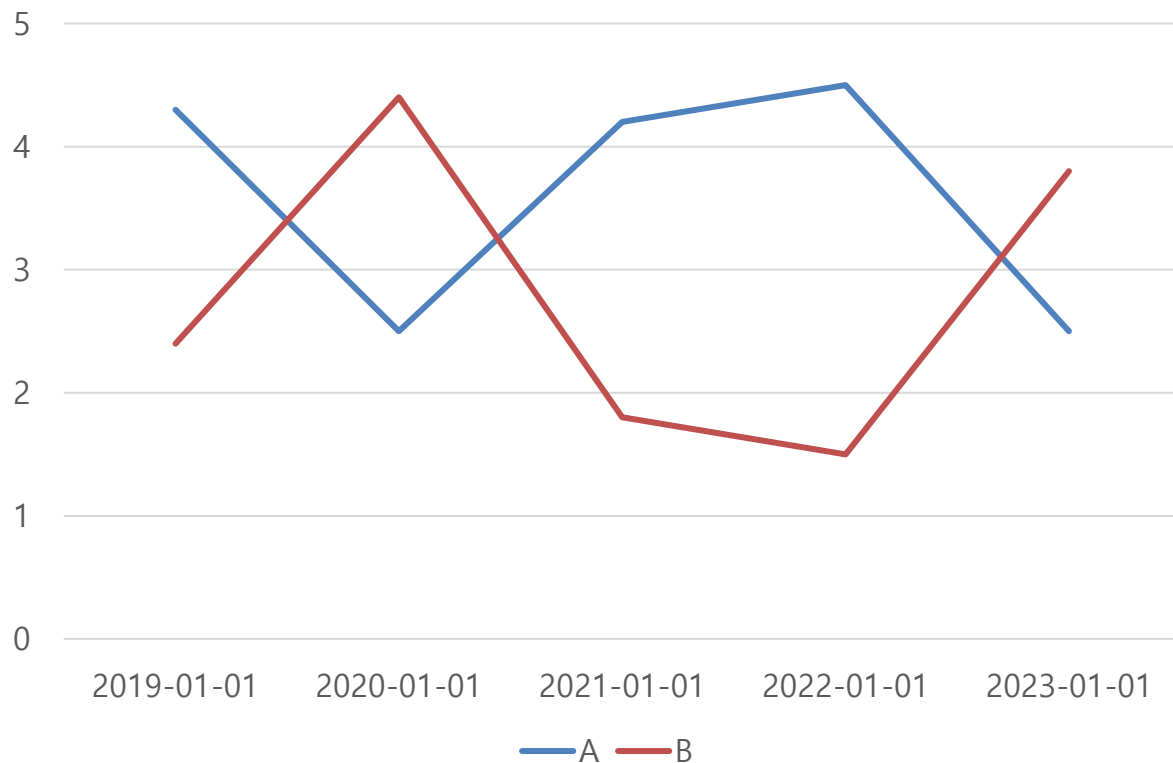
$$score = \frac{positive\ count}{positive\ count + negative\ count}$$


## 04 Portfolio

### 포트폴리오 이론

자산을 분산 투자하여 포트폴리오를 만들게 되면 분산 투자 전보다 위험을 감소시킬 수 있다.

주식 가격



주식 A와 B를 적절한 비율로 구매하면 위험성은 감소하며, A와 B의 구매 비율을 조정해 예상 수익을 조정할 수 있다.

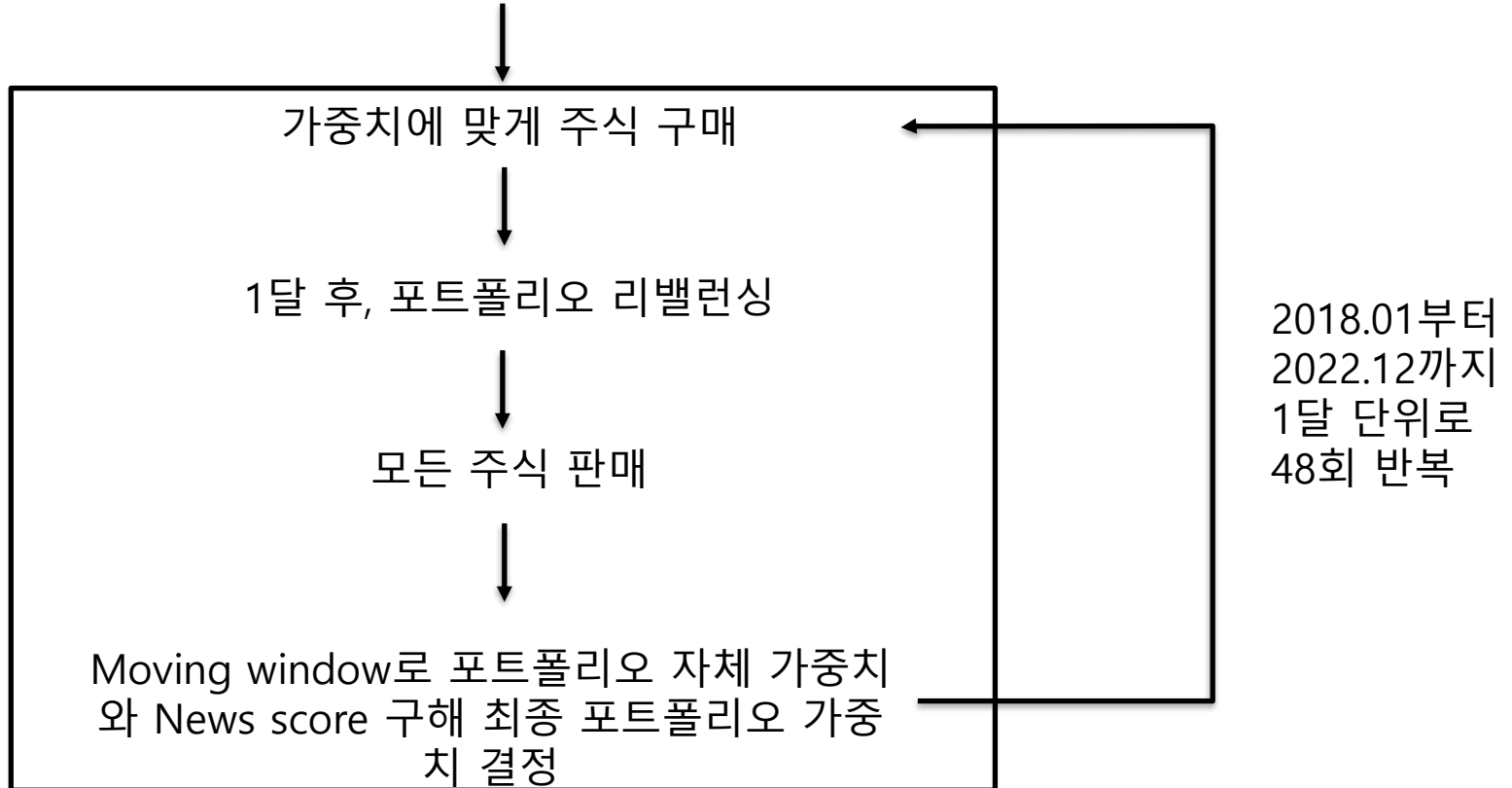
상관계수의 정도, 주식 비율에 의해 분산 효과가 결정된다.

분산 ↑ = high risk high return  
분산 ↓ = low risk low return

$$P = w_1A + w_2B$$
$$s.t. w_1 + w_2 = 1$$

### 포트폴리오 구성

2014~2018년 5년치 데이터를 사용해 공분산을 구한 후, 포트폴리오 가중치 결정



## 포트폴리오 가중치 업데이트

```
def weight_cal(news_score, weights):  
    """  
    news_score는 news 공부정 분석으로 나온 positive count/(positive count + negative count) 값입니다.  
    news_score는 list형태로 input, weights는 dict형태로 input을 받아야 합니다.  
    최종 리턴값은 seires형태로 finish weight으로 나갑니다.  
    """  
    result = []  
    for i in range(len(news_score)):  
        news_score_i = news_score[i]  
        weights_i = weights[i]  
  
        if (news_score_i >= 0.6) & (weights_i >= 0.1) & (weights_i < 0.2) : # news가 매우 긍정, 포트폴리오는 보통 가중치  
            total_weight = news_score_i*0.2 + weights_i  
  
        if (news_score_i >= 0.6) & (weights_i >= 0.2) : # news가 매우 긍정, 포트폴리오도 강한 가중치  
            total_weight = news_score_i*0.25 + weights_i  
  
        if (news_score_i >= 0.55) & (weights_i >= 0.1) & (weights_i < 0.2) : # news가 긍정, 포트폴리오는 약한 가중치  
            total_weight = news_score_i * 0.15 + weights_i  
  
        if (news_score_i >= 0.5) & (weights_i >= 0.1) : # news가 긍정, 포트폴리오는 보통 가중치  
            total_weight = news_score_i * 0.2 + weights_i * 0.8  
  
        if (news_score_i <= 0.5) & (weights_i >= 0.1) : # news가 부정, 포트폴리오는 보통 가중치  
            total_weight = news_score_i * 0.25 + weights_i * 0.75  
  
        if (news_score_i <= 0.45) & (weights_i >= 0.2) : # news가 부정, 포트폴리오는 강한 가중치  
            total_weight = weights_i - news_score_i*0.2  
  
        if (news_score_i <= 0.45) & (weights_i <= 0.1) : # news가 부정, 포트폴리오는 약한 가중치  
            total_weight = weights_i - news_score_i*0.25  
  
        if (news_score_i <= 0.4) & (weights_i <= 0.1) : # news가 매우 부정, 포트폴리오는 약한 가중치  
            total_weight = 0  
  
        if total_weight <= 0 :  
            total_weight = 0  
  
        result.append(total_weight)  
    result = pd.Series(result)  
    finish_weight = result/result.sum()  
    return finish_weight
```

포트폴리오 리밸런싱을 위해  
News와 포트폴리오의 가중치  
를 계속해서 업데이트를 진행

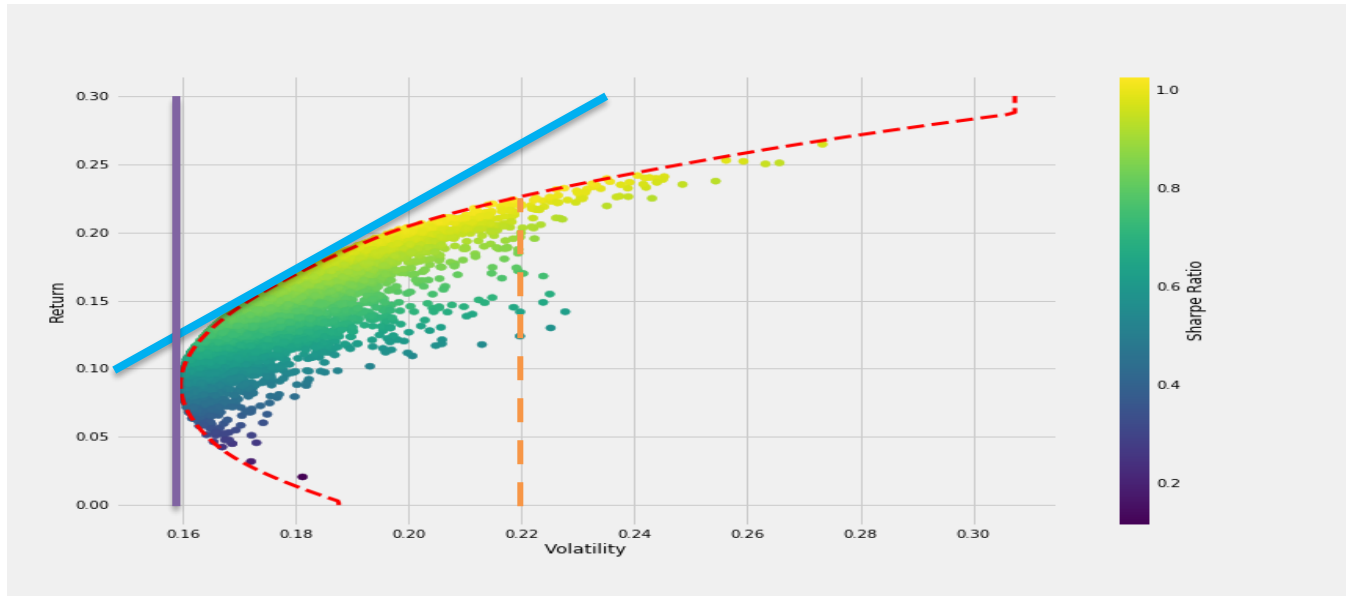
수식 정의 시, news score의  
전체 합은 22이며, 포트폴리오  
의 전체 합은 1으로 차이가 존재

스케일을 하기에는 기존의 상관관계가 깨지는 현상이 발생

∴if문을 사용해 경우의 수를 나눠  
새로운 가중치로 결정

# 04 Portfolio

## 포트폴리오 이론

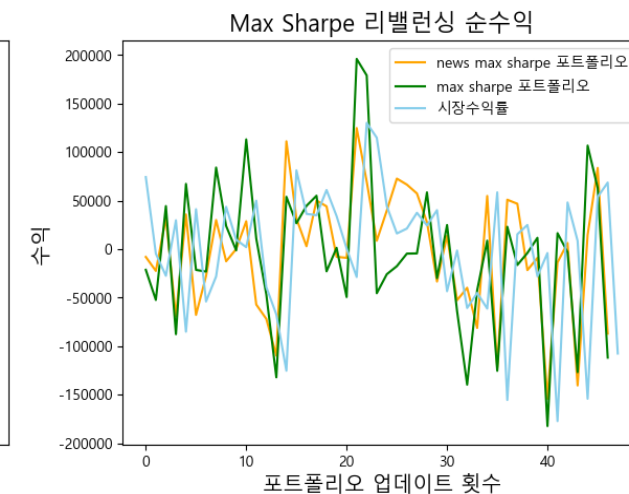
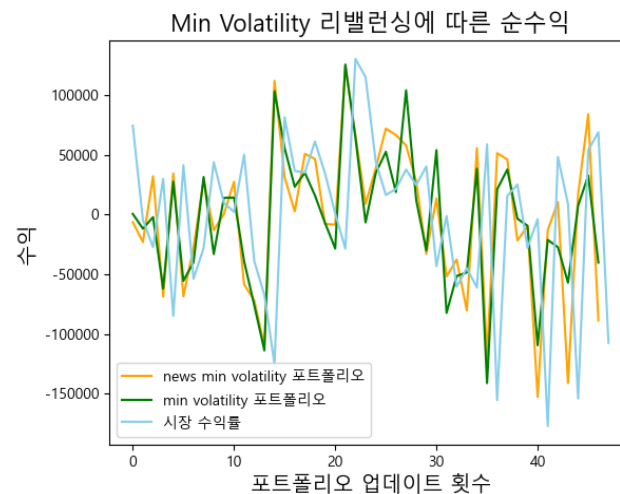
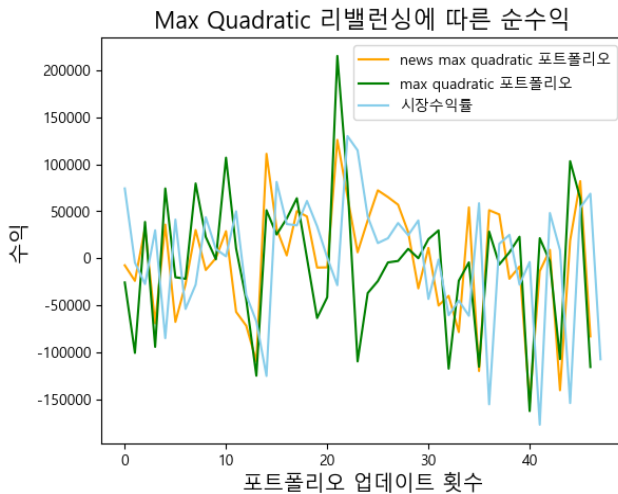
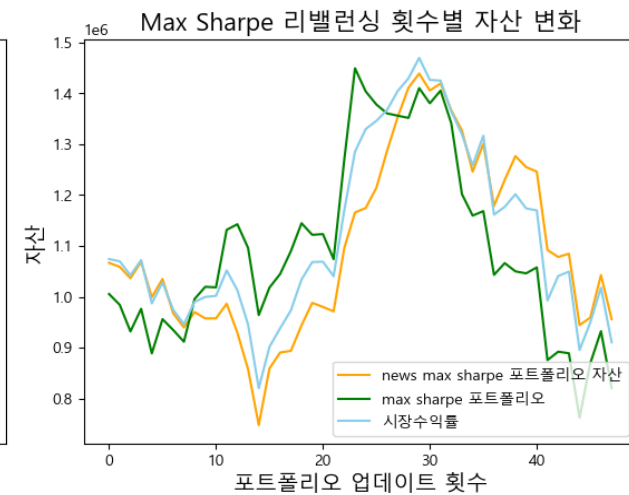
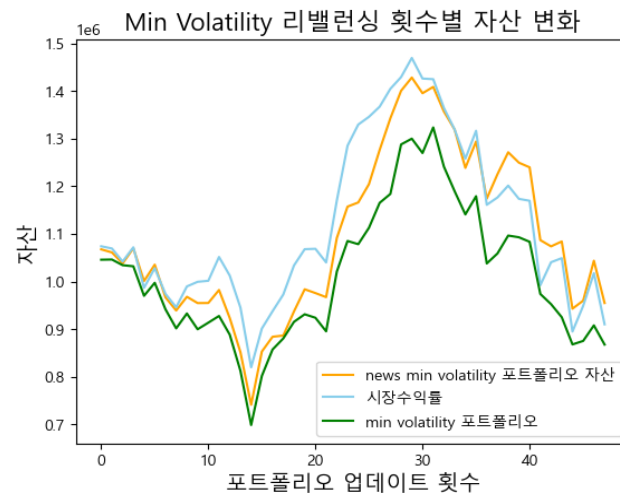
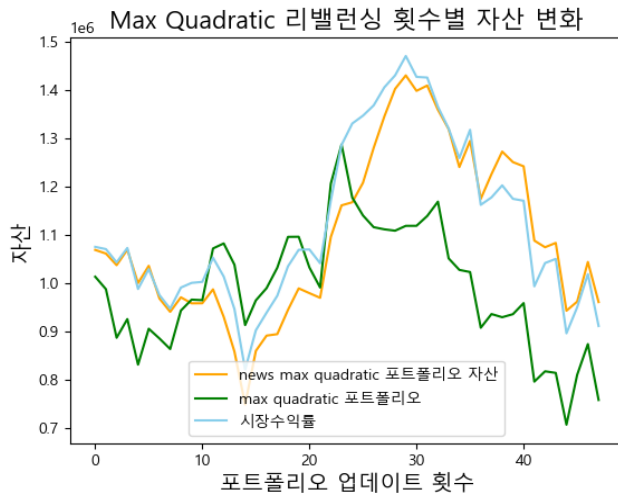


| 종류(PyPortfolioOpt)        | 설명  |
|---------------------------|---|
| Min Volatility            | 분산을 최소로 하여 안전성 추구   |
| Max Quadratic Utility     | 위험 회피도에 맞는 최적의 포트폴리오 구성<br>Risk Aversion = 1 (default)        |
| Max Sharpe(risk rate=0.2) | Sharpe 비율에 맞는 최적의 포트폴리오 구성<br>Risk Free Rate = 0.02 (default) |



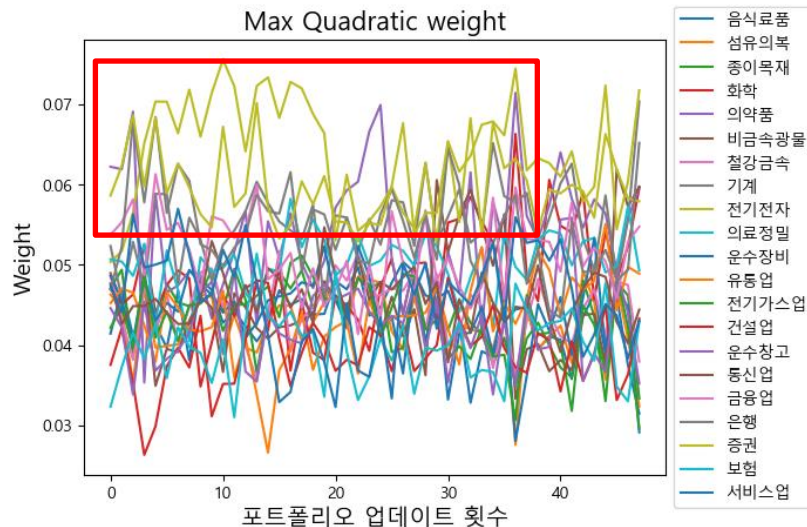
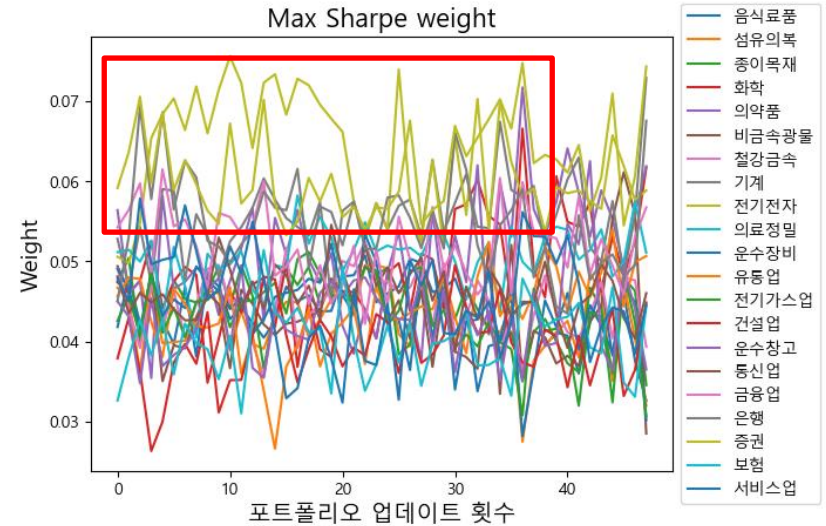
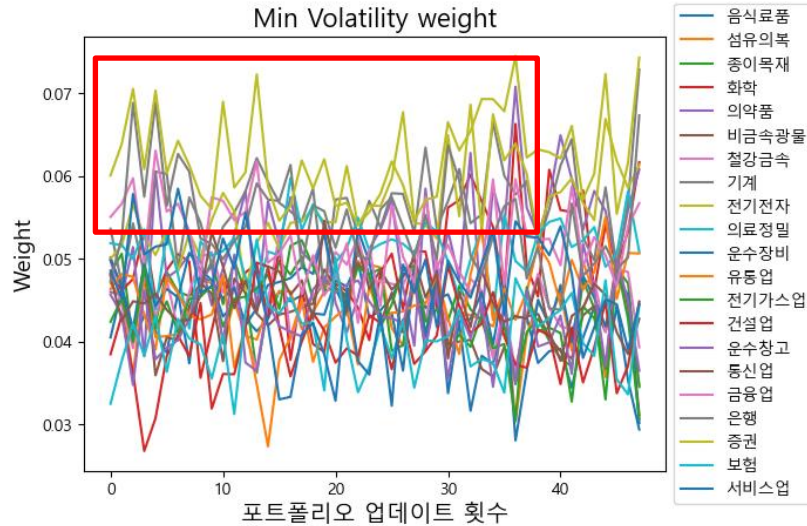
# 04 Portfolio

## 포트폴리오 결과



# 04 Portfolio

## 포트폴리오 결과



자산의 변화가 어떤 포트폴리오와 뉴스를 결합해도 동일하다는 것을 보고 가중치의 변화에 대한 확인

실제 가중치는 다르지만 자산의 변화는 동일함

가중치의 차이가 크게 존재하지 않아 안정적인 포트폴리오로 수렴했다고 판단  
체계적인 위험성이 비체계적인 위험성에 수렴

# 05 Conclusions & Future research

---

## Conclusions

- 1) News title에서 positive와 negative를 판단할 수 있는 감정사전 제작
- 2) 감성사전을 통해 Score를 계산하여 포트폴리오 가중치에 반영
- 3) 업종 전반을 표현하는 지수들로 포트폴리오를 제작한 것,  
업종별 비슷한 가중치로 인해 시장수익률과 비슷한 수익률의 포트폴리오가 제작되었음

## Future research

- 1) 업종 내부에서 1~2개의 기업을 선정하여 포트폴리오를 구성
- 2) News score를 포트폴리오 가중치에 반영하기 위한 수식 정의 및 최적화

---

감사합니다