

감성 분석이 결합된 MVO 전략: 심층 분석과 적용 가능성

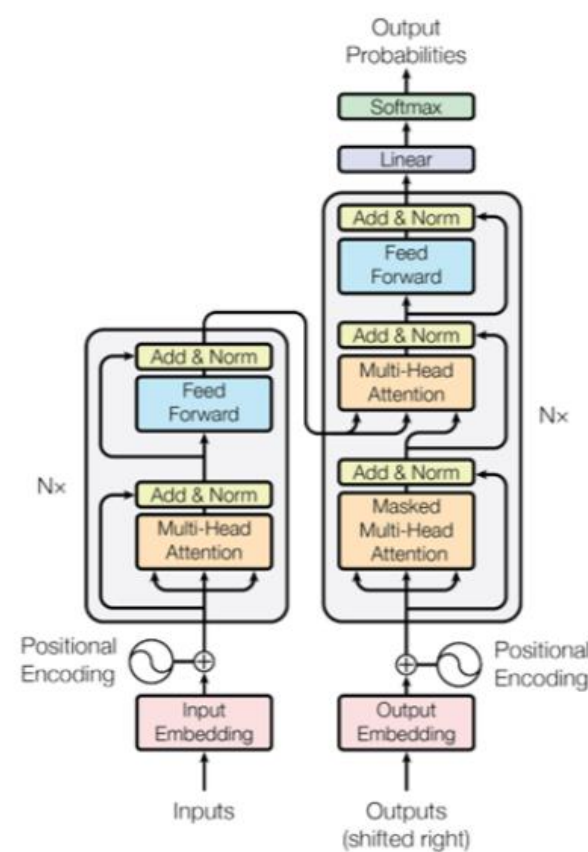
목차



<u>PART 01</u>	연구 방법론
<u>PART 02</u>	Sentiment Analysis
<u>PART 03</u>	연구 및 분석
PART 04	결론

: 연구방법론

1
자연어 처리



2
감성 분석



일반적으로(NLP: Natural Language Processing), 텍스트 마이닝 (text mining), 기계 학습(ML: Machine Learning) 등의 기술을 사용하여 문장 또는 문서에서의 표현이 긍정, 부정 혹은 중립인지를 판별하는 기술

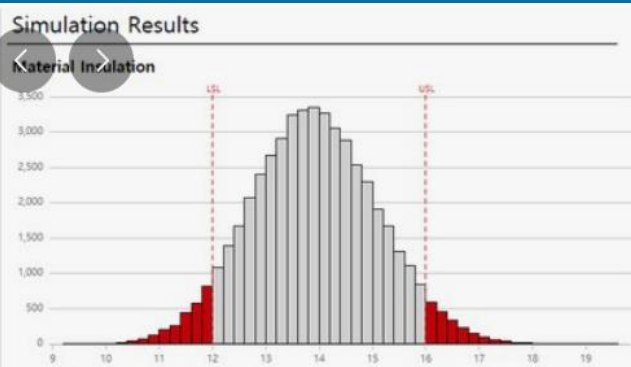
3
MVO

Minimize: $\frac{1}{2}w^T \Sigma w - \lambda \mu^T w$

- w: 각 자산에 대한 투자 비중의 벡터
- Σ: 자산 간의 공분산 행렬
- μ: 각 자산의 기대 수익률 벡터
- λ: 투자자의 위험 회피 성향을 나타내는 조정 변수

포트폴리오 이론의 핵심 개념 중 하나로, 투자자의 위험 회피 성향을 고려하여 주어진 기대수익률을 최대화하거나 주어진 리스크 수준에서 수익률을 최대화하는 포트폴리오를 구성하는 방법

4
몬테카를로 시뮬레이션



- 1) 입력 변수의 확률 분포를 정의한다.
- 2) 해당 확률 분포에 따라 무작위로 샘플을 생성한다.
- 3) 생성된 샘플을 바탕으로 모델을 여러 번 반복 실행하여 결과를 도출한다.
- 4) 결과를 분석하여 통계적 특성을 추정한다.

무작위 샘플링을 통해 복잡한 시스템이나 과정을 모형화하고 분석하는 기법

: 데이터 설명

01 - 블룸버그 서비스 제공 기사

주로 해외 매크로 지표(PCE, 금리인하 등)뉴스를 다루고 있음.
2020.11.17~2024.06.05 일간의 데이터를 수집하였으며, 총
4045(809*5)개의 기사를 기반으로 추출

02 - 한국 경제 사이트 제공 코스피 태그 데이터

코스피에 집중적으로 관련이 있는 데이터만 모아놓음. 기사는
사이트에서 제공하고 있는 2016.07.08 ~ 2024.05.31 까지 총
7583개의 기사를 기반으로 추출

03 - 매일경제 사이트 코스피 검색 데이터

한국 경제에서 제공하는 코스피 태그 데이터 보다 많은 데이터를
다루고 있지만 다소 노이즈와 같은 기사도 포함되어 있음. 기사는
2016-03-22 ~ 2024.06.05 까지 총 25577개의 기사를 기반으로 추출

[정규화]

$$\text{monthly sentimental score} = \frac{\sum \left(\frac{\sum(\text{daily senti score})}{\text{num daily news}} \right)}{\text{num monthly news}}$$

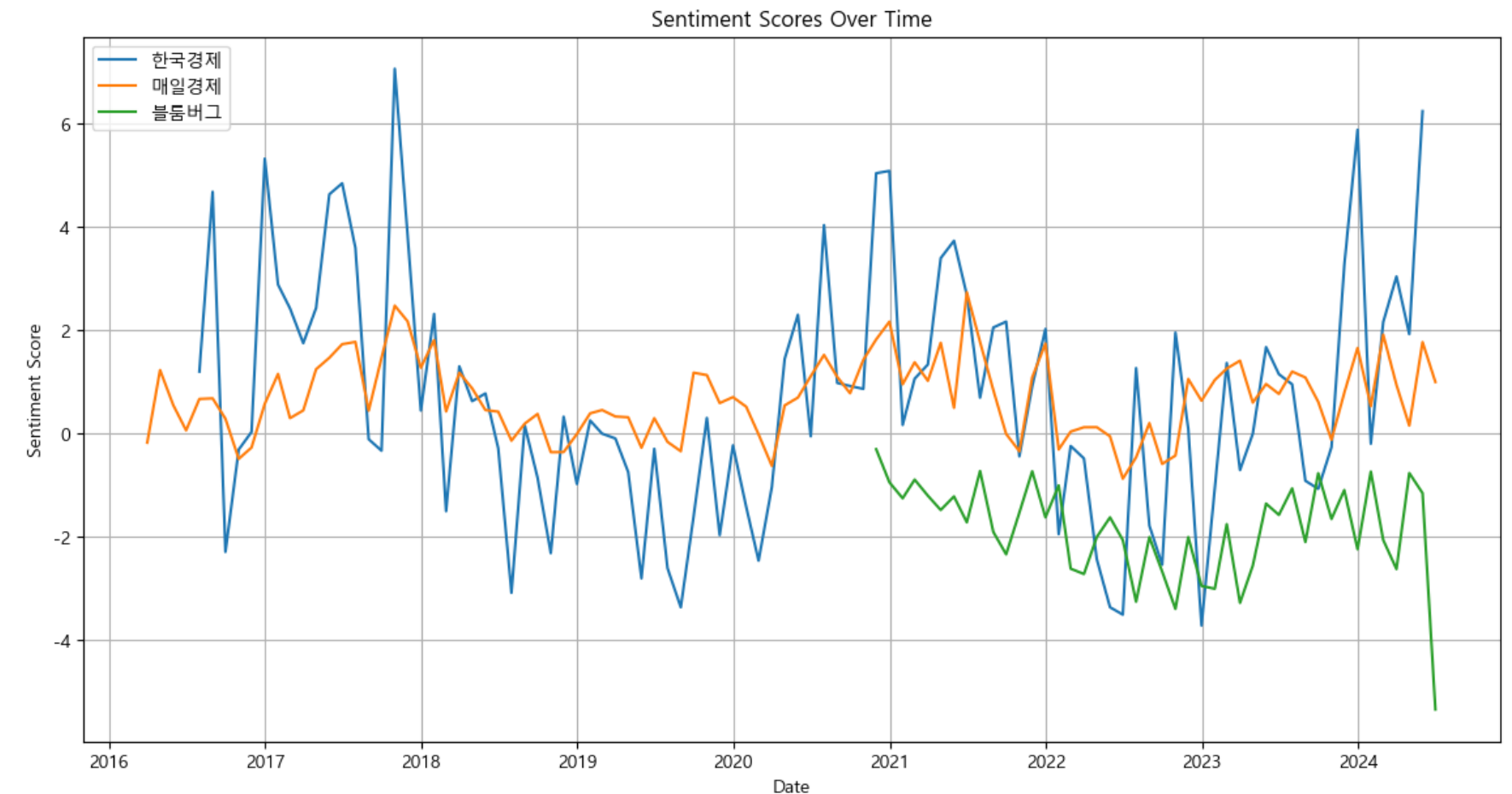
$$\text{min_senti_score} = \min(\text{monthly sentimental score}_{12} \sim \text{monthly sentimental score}_1)$$

$$\text{max_senti_score} = \max(\text{monthly sentimental score}_{12} \sim \text{monthly sentimental score}_1)$$

$$\text{adjusted senti score} = 0 : \text{if } (\text{senti score} < \text{min_senti_score})$$

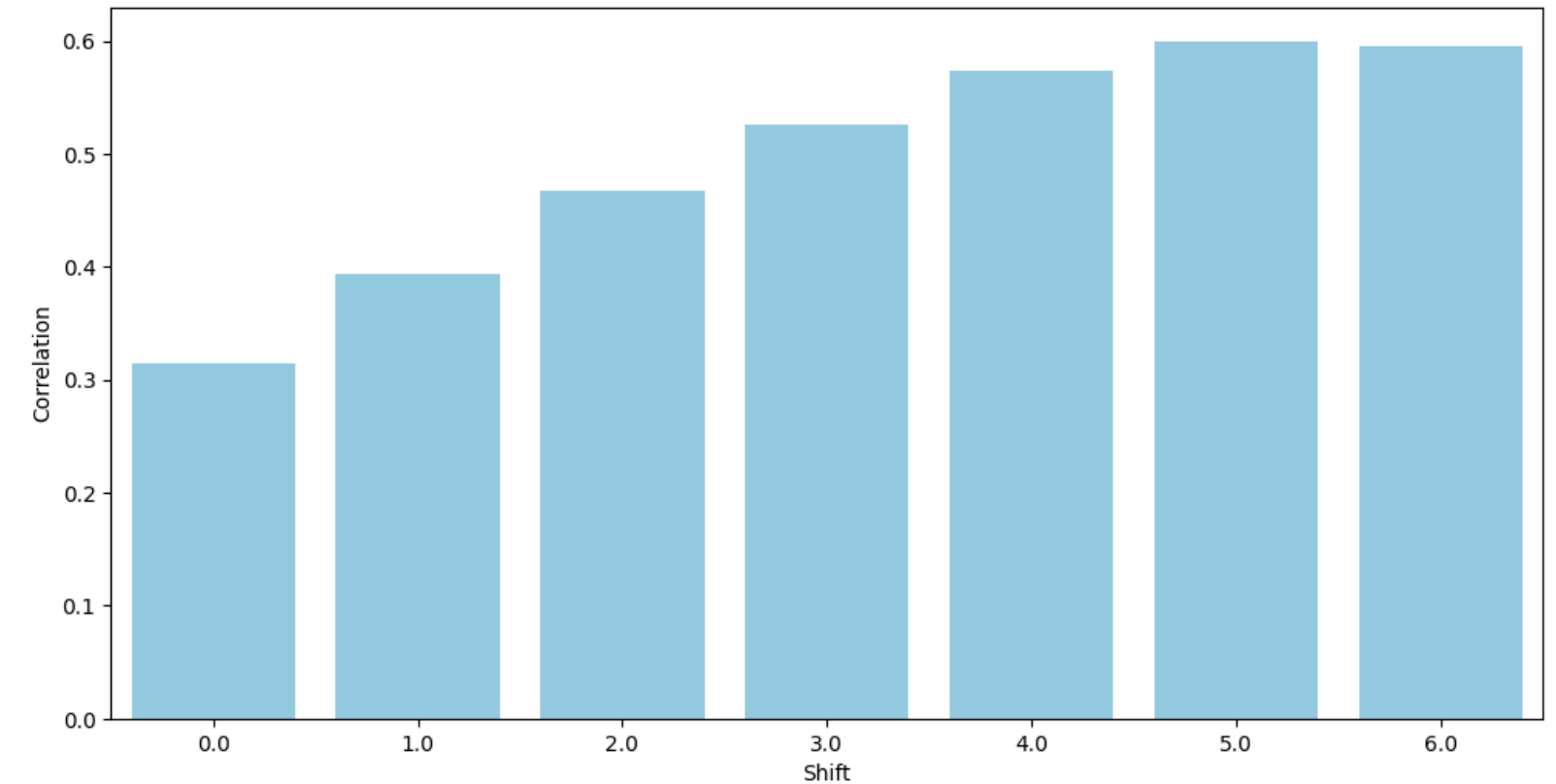
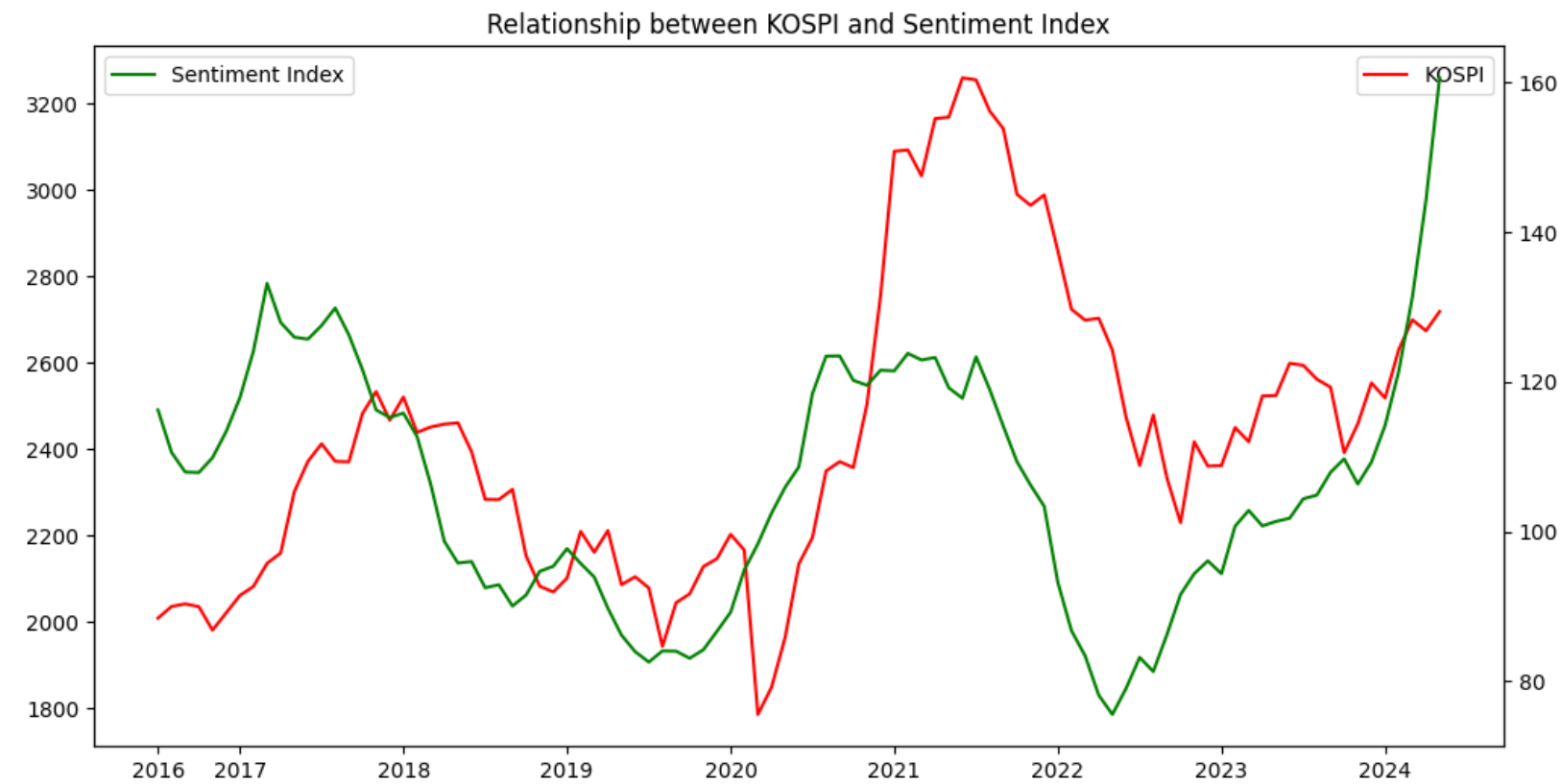
$$\text{adjusted senti score} = 1 : \text{if } (\text{senti score} > \text{max_senti_score})$$

$$\text{adjusted senti score} = (\text{senti score} - \text{min senti score}) / (\text{max senti score} - \text{min senti score})$$



• Corr

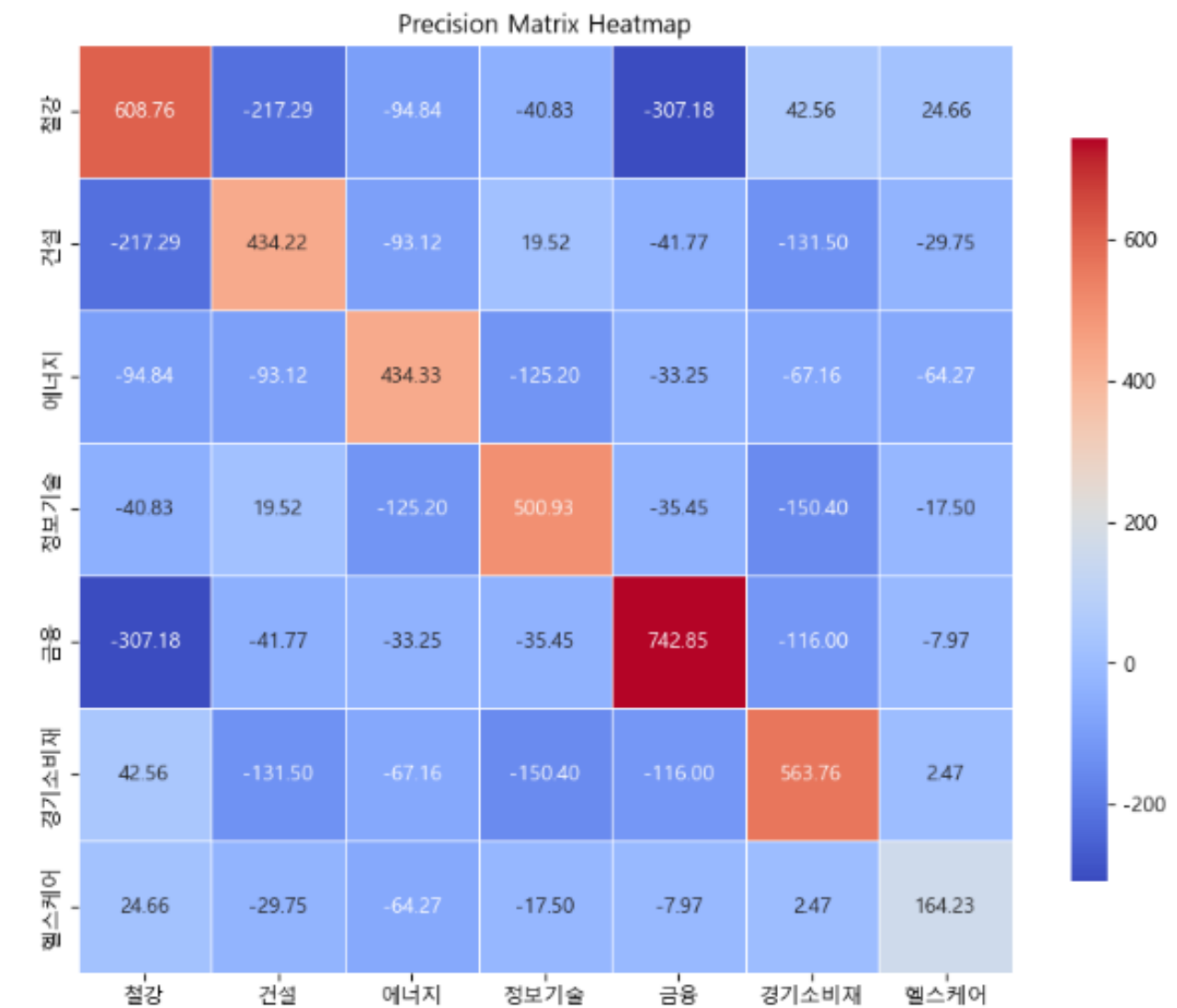
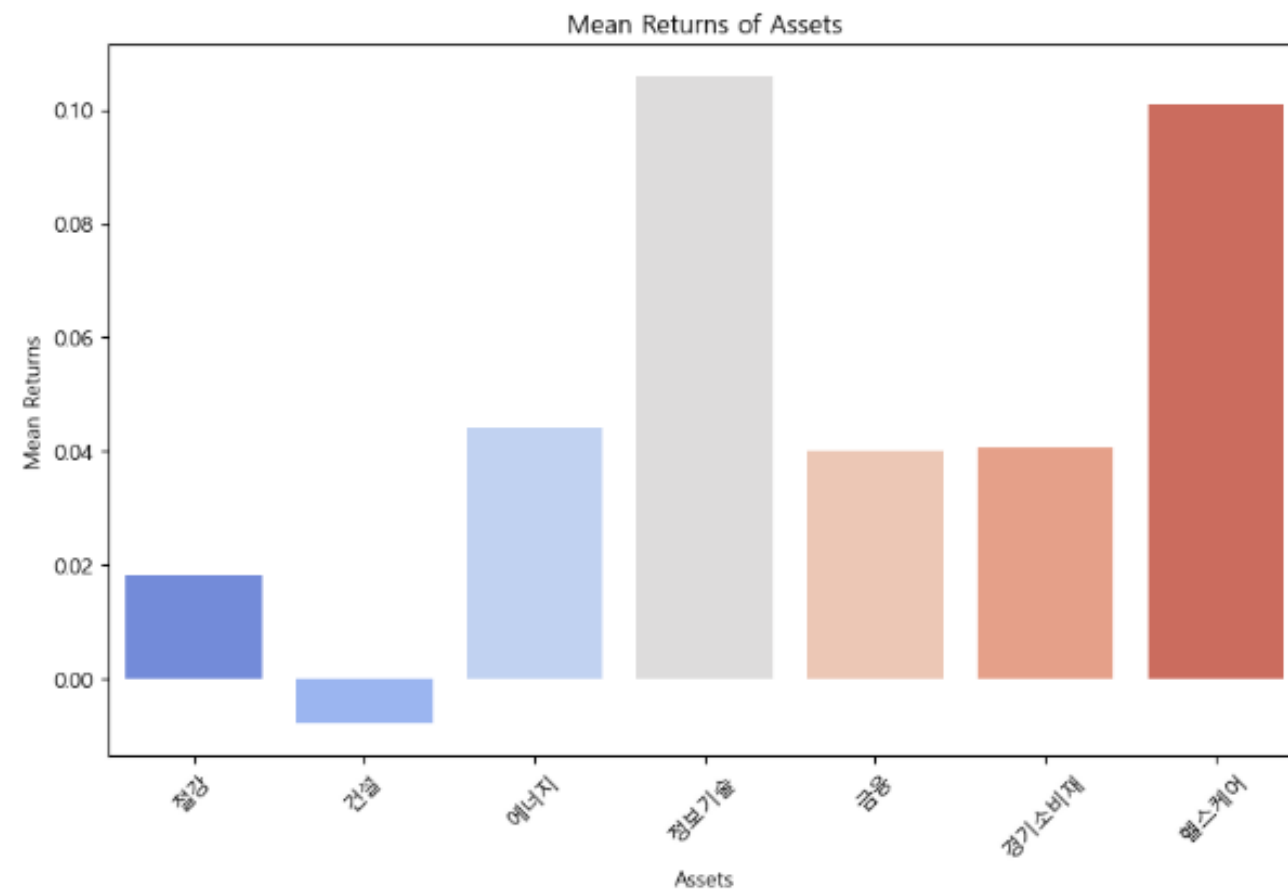
$$\text{Sentiment Index} = \frac{\sum_d \text{Positive} - \sum_d \text{Negative}}{\sum_d \text{Positive} + \sum_d \text{Negative}}$$



: Granger Causality

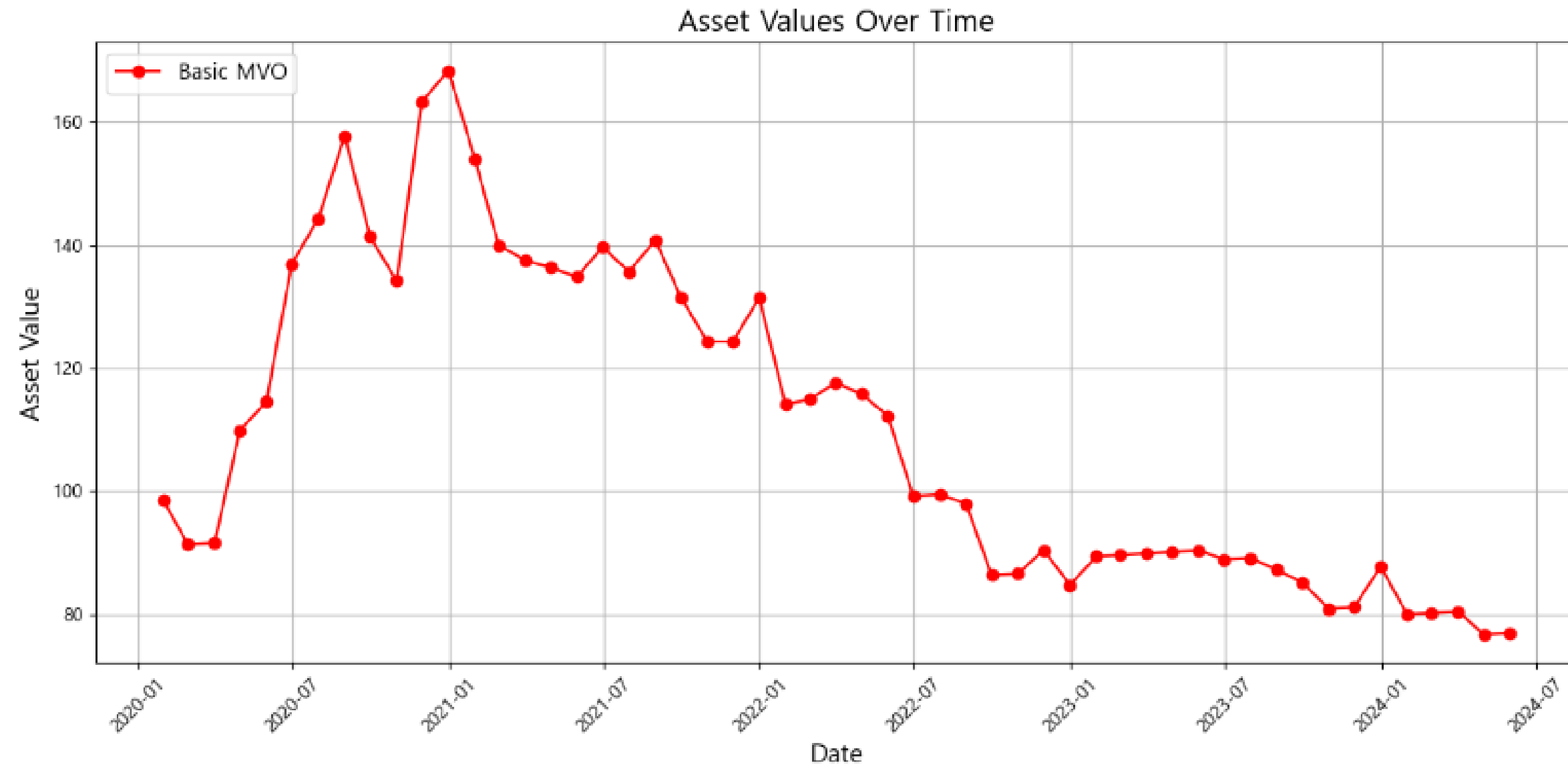
시차	경기선행지수	M2	BTC	sent-index
1	0.5699	0.707	0.6506	0*
2	0.0383*	0.8049	0.347	0*
3	0.0678	0.4673	0.5562	0*
4	0.2748	0.1211	0.5762	0.0001*
5	0.3913	0.0204*	0.4863	0.0001*
6	0.0609	0.0423*	0.5569	0*
7	0.0199*	0.0591	0.5448	0*
8	0.0536	0.0117*	0.5725	0.0005*
9	0.0469*	0.0066*	0.6216	0.0008*
10	0.0515	0.0136*	0.7625	0.0001*
11	0.0549	0.0106*	0.5006	0*
12	0.0597	0.0062*	0.1439	0*

: Basic Mean-Variance Optimization



: Basic Mean-Variance Optimization

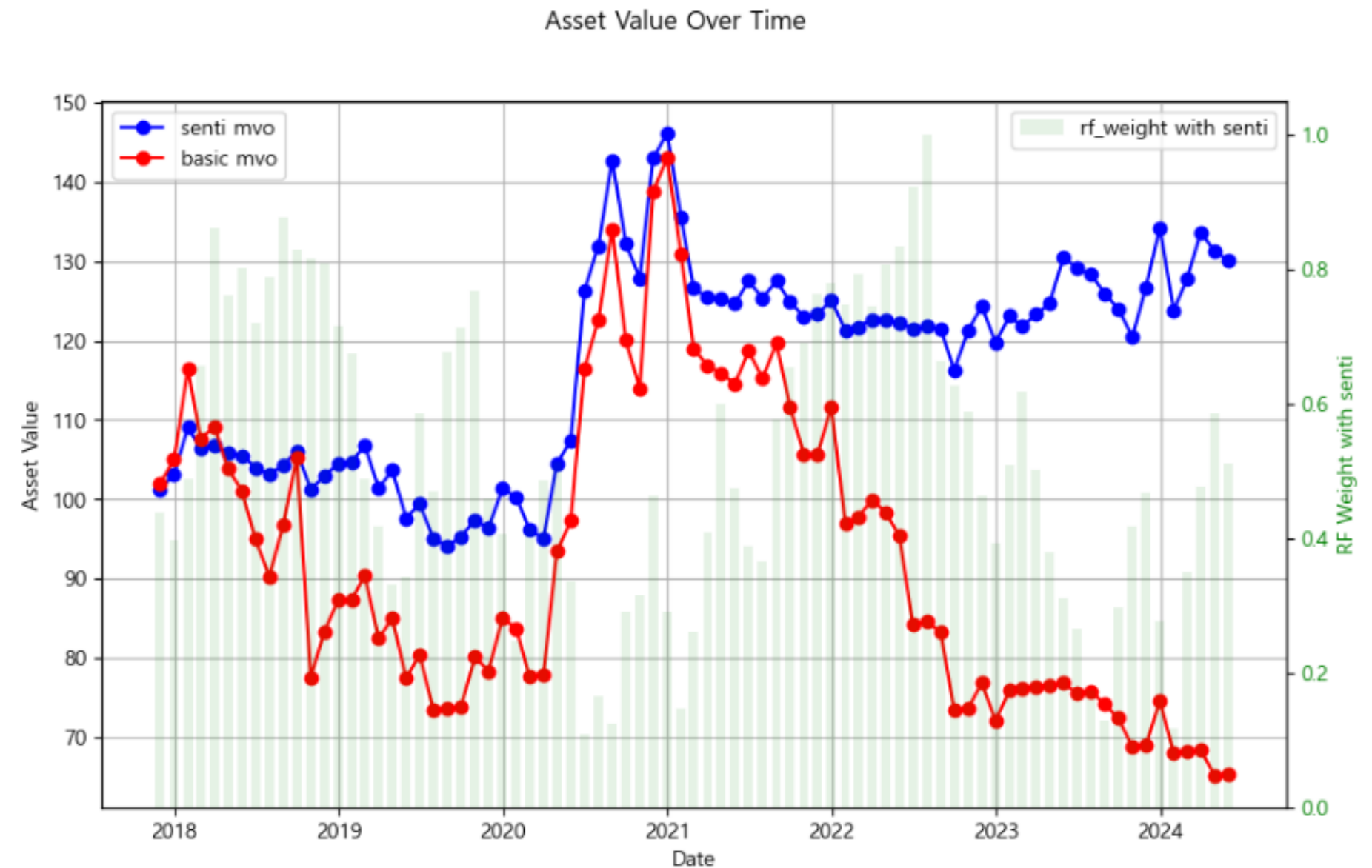
[벡테스트를 통한 월별
수익률]



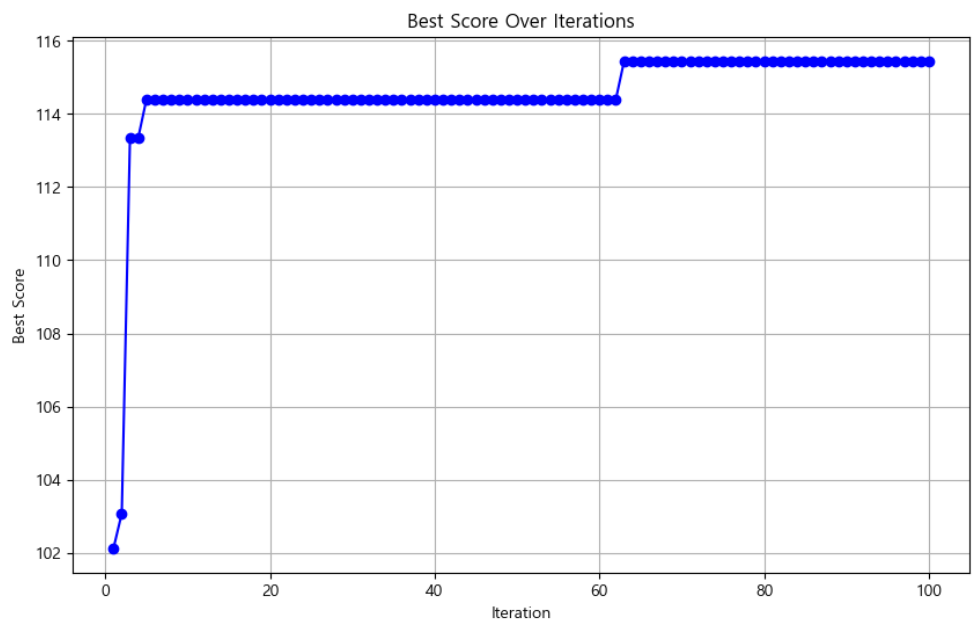
: Sentimental Data 추출 방식

[Sentimental Data 추출 방법]

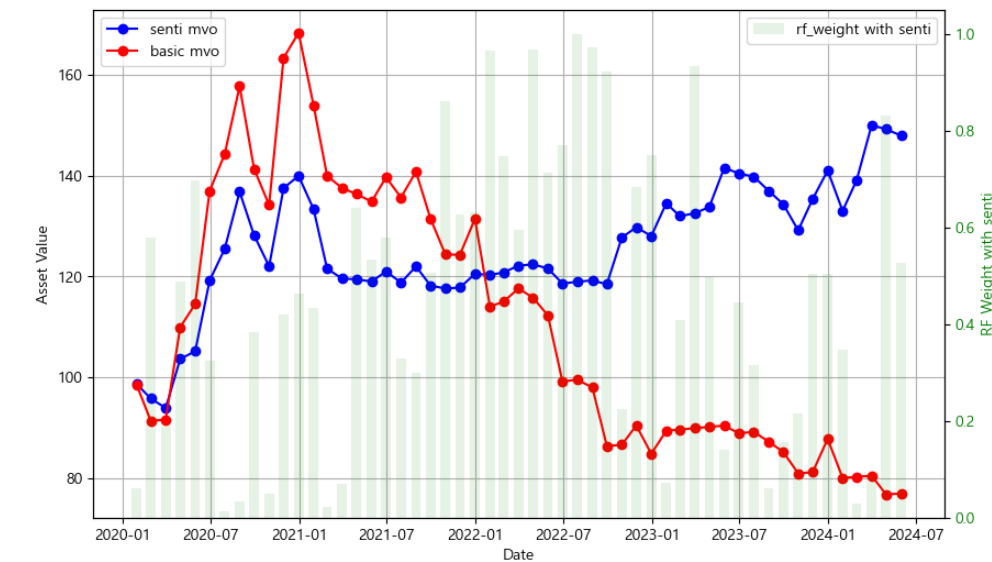
1. 무위험 자산 없이 mvo weight 를 뽑는다.
2. $(1 - \text{한달전 senti score} * 0.4 + \text{두달전 senti score} * 0.3 + \text{세달전 senti score} * 0.3)$ 한 만큼의 weight 를 무위험 자산에 투자한다.
3. 한달마다 리밸런싱을 진행하며 최종 수익률을 계산한다.



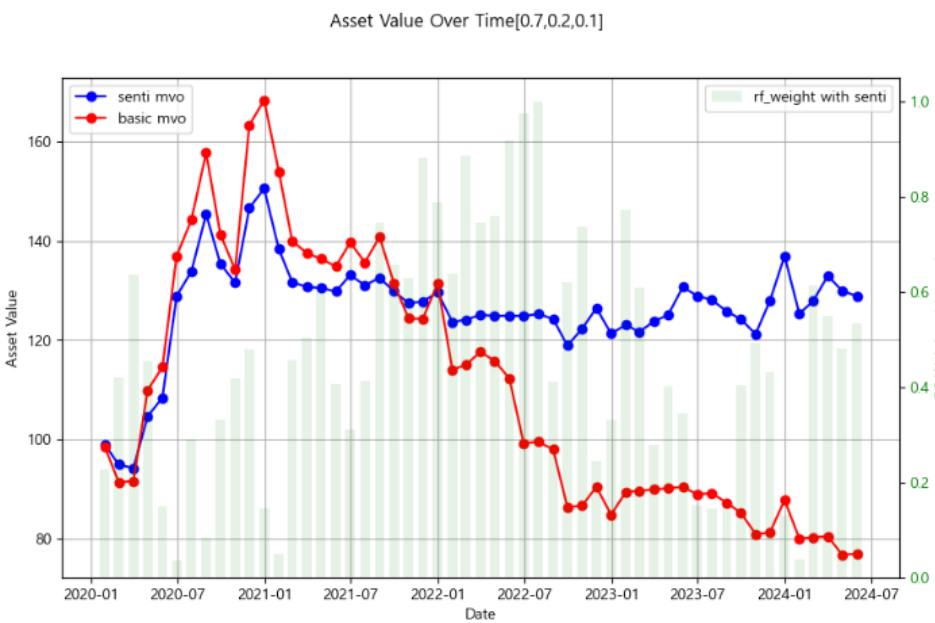
: Monte-Carlo 활용한 최적 weight 판단



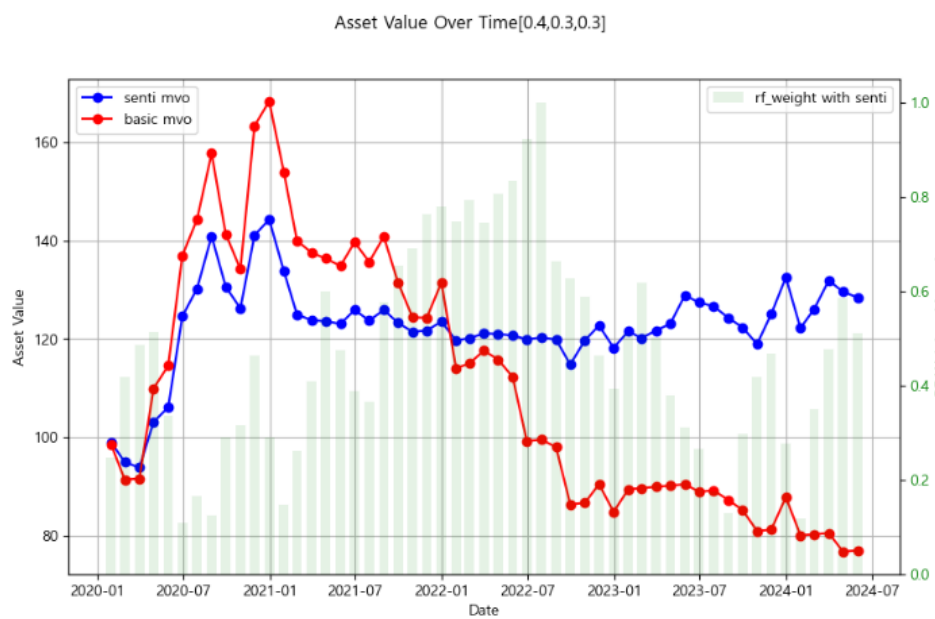
Asset Value Over Time[몬테카를로]



[senti 1]



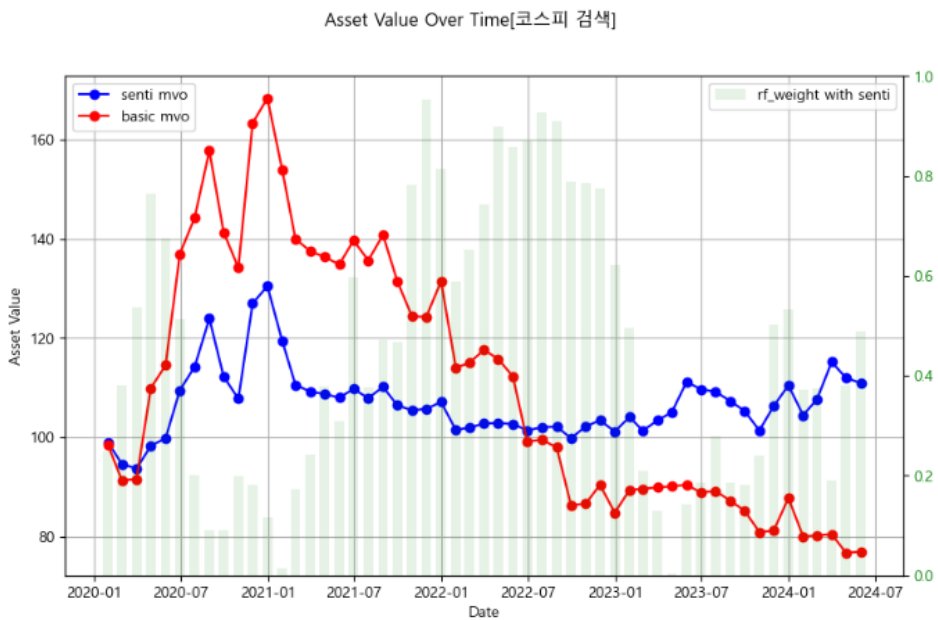
[senti 2]



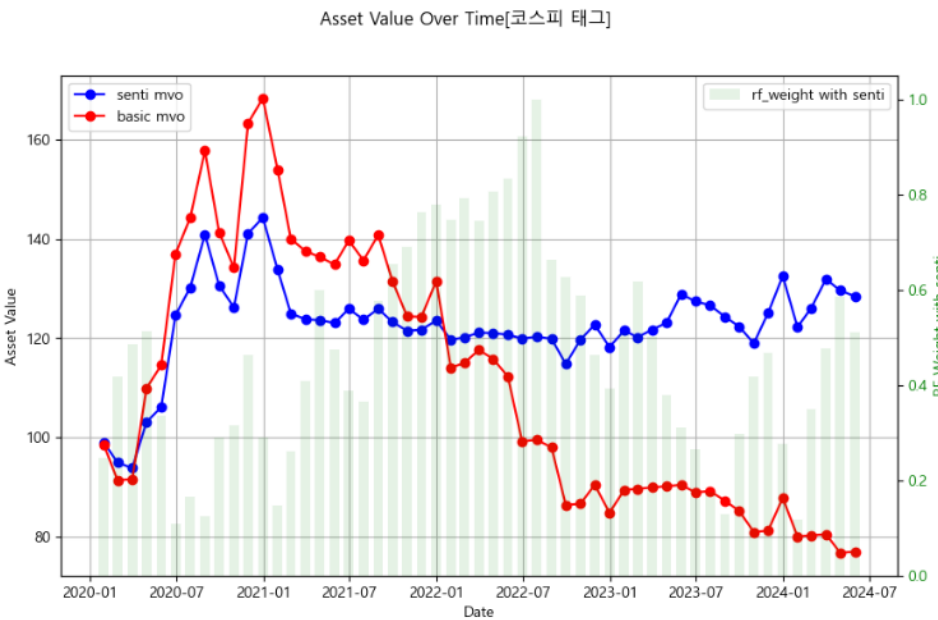
[senti 3]

	Senti1 (M.C)	Senti2 (0.7,0.2,0.1)	Senti3 (0.4,0.3,0.3)	Basic Mvo
Sharpe Ratio	0.6512	0.4301	0.4354	-0.1245
MDD	-0.1599	-0.2046	-0.2104	-0.5445
Sortino Ratio	1.3679	0.8782	0.8812	-0.2293
Information Ratio	0.9012	0.8278	0.7919	Nan

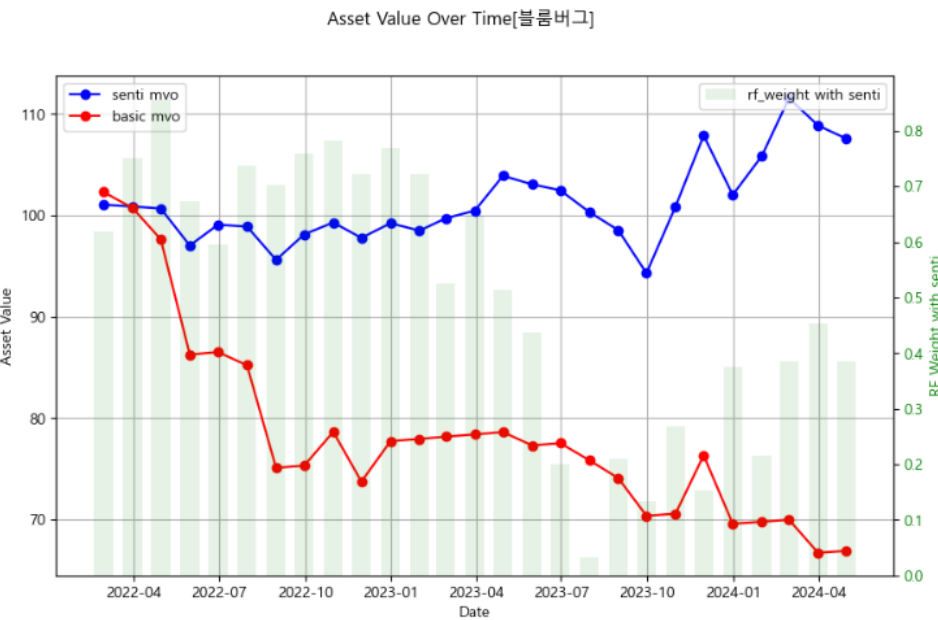
: Sentimental Data 변경



[코스피 검색 데이터]



[코스피 태그 데이터]



[해외 매크로 데이터 이용]

	Senti(코스피 검색)	Senti(코스피 태그)	Senti(해외 매크로)	Basic MVO
Sharpe Ratio	0.2265	0.4354	0.3331	-0.1245
MDD	-0.2351	-0.2104	-0.0924	-0.5445
Sortino Ratio	0.4162	0.8812	0.6840	-0.2293
Informatio Ratio	0.4654	0.7919	2.0457	Nan

· 결과

KR-finbert 모델을 활용해서 코스피와 관련된 sentimental score를 뽑아보았다.

그 결과 sentiment score가 실제 코스피 200의 지수에 선행하는 경향이 있음을 확인할 수 있었다.

Corr은 0.6과 같이 나왔으며 약 5개월 전 senti가 가장 코스피 지수를 잘 예측한다고 볼 수 있었다.

또한, granger 인과검정을 통해서 코스피와 기타 다른 지표들간의 인과관계를 비교해 보았을 경우 M2, 경기선행지수, 비트코인 가격 등 기존에 알려진 주가와 선행관계를 띠는 모델보다 senti score가 기존에 알려진 주가에 선행하는 지표보다 주가를 예측하는데 더 도움이 됨을 확인할 수 있었다.

이를 바탕으로 senti score를 활용해서 mvo weight를 최적화 해 보았다. 그 결과 sharp ratio가 기본 mvo 대비 $-0.12 \rightarrow 0.43$, MDD가 $-0.54 \rightarrow -0.2$ 로 크게 상승한 것을 관측할 수 있었다. sentimental data의 경우 코스피 태그를 통해 직접적으로 코스피와 관련된 데이터가 다른 해외 매크로 데이터 혹은 코스피 검색 데이터 보다 성능이 더 뛰어남을 관측할 수 있었다.

·
·

감사합니다

한 학기동안 수고많으셨습니다.

끝!