

Miraeasset AI R&D Team Report

강상인 : [Linkedin](#)

sik247@nyu.edu: [Github](#)

0. 프로젝트 목표

포트폴리오 최적화 프로젝트는 S002에서 S514까지의 다양한 주식 범주에 대한 포트폴리오 할당 전략을 최적화하는 데에 Long Short-Term Memory (LSTM) 신경망을 활용하고 있습니다. LSTM은 특히 금융 시계열 데이터에 적합한 딥러닝 모델로서, 시계열 데이터의 순차적인 특성을 잘 이해하고 처리할 수 있는 능력이 있습니다.

LSTM의 최적성은 시계열 데이터의 복잡한 동적 패턴을 학습하고 예측하는 데 있습니다. 특히 금융 시장에서는 주식 가격의 흐름, 변동성, 상관관계 등과 같은 다양한 동적인 특성이 존재합니다. LSTM은 이러한 특성을 잘 캡처하여 최적의 포트폴리오 할당을 찾아내는 데 도움을 줍니다.

LSTM을 선택한 이유 중 하나는 그 유연성과 강력한 예측 능력에 있습니다. 이 모델은 금융 시계열 데이터의 비선형적인 패턴을 학습할 수 있으며, 장기적인 의존성 및 기억력을 갖고 있어 포트폴리오 최적화와 같은 복잡한 작업에 효과적입니다.

따라서 LSTM은 주식 시장에서의 동적이고 불안정한 특성을 고려하여 포트폴리오 최적화에 적합한 선택이라고 할 수 있습니다.

1. 데이터 로딩

```
raw_stock = pd.read_csv('data/raw_data_ts.csv')
monthly_treasury_bond = pd.read_csv('data/TB3MS.csv')
daily_treasury_bond = pd.read_csv('data/TB3MS_daily.csv')
```

1. **raw_stock**: 주가 데이터로, 'data/raw_data_ts.csv'에서 가져온 것입니다. 주식 종목(S002 - S514)에 대한 월간 데이터로 예상됩니다.

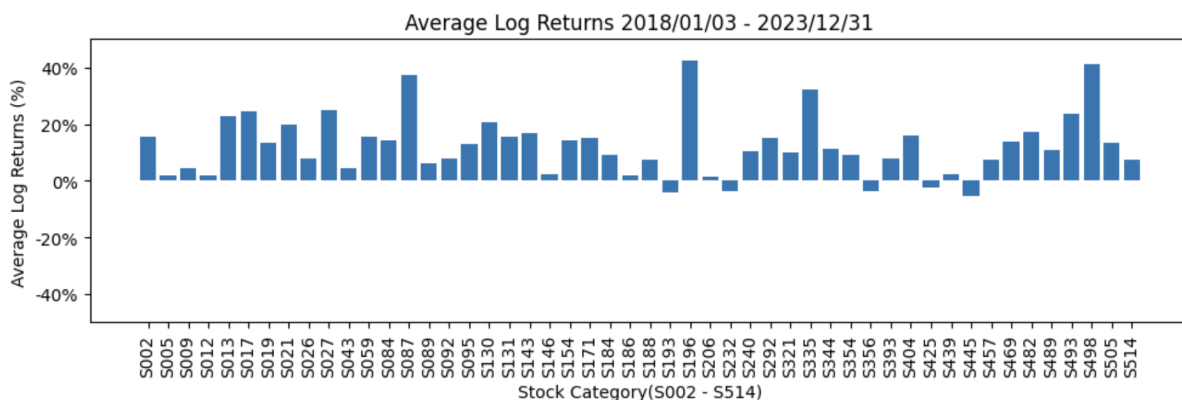
2. **monthly_treasury_bond**: 월간 국채 3개월물 금리 데이터입니다. 'data/TB3MS.csv'에서 가져온 것으로 예상됩니다.
3. **daily_treasury_bond**: 일일 국채 3개월물 금리 데이터입니다. 'data/TB3MS_daily.csv'에서 가져온 것으로 예상됩니다.

2. 날짜 파싱 및 데이터 로딩

- **날짜 파싱 함수**: 정확한 날짜 표현을 보장하기 위해 사용자 정의 날짜 파싱 함수를 도입하였습니다.
- **데이터 로딩**: Raw 주식 데이터가 로드되며, 'eval_d' 열이 datetime으로 변환되어 인덱스로 사용됩니다.

3. 수익률 계산(로그)

- **티커 정의**: 후속 분석을 위해 고유한 티커가 식별되었습니다.
- **날짜 범위 정의**: 분석을 위해 시작 및 종료 날짜가 지정되었습니다.
- **날짜 범위 처리**: 데이터프레임의 인덱스에 종료 날짜가 존재하는지 확인하기 위한 조정이 이루어졌습니다.
- **로그 수익률 데이터프레임 초기화**: 주식 수익률 정보를 캡처하기 위해 로그 수익률 데이터프레임이 초기화되었습니다.
- **각 티커에 대한 계산**: 각 주식에 대한 로그 수익률의 반복적인 계산이 수행되었습니다.
- **데이터 부재 처리**: 특정 주식에 대한 데이터가 없는 경우를 처리하기 위한 확인이 구현되었습니다.



4. 공분산 행렬 계산

- **공분산 행렬:** 로그 수익률을 기반으로 공분산 행렬이 계산되었습니다.
- **다중공선성 처리:** 공분산 행렬에서 발생할 수 있는 다중공선성에 대한 고려사항이 있습니다.

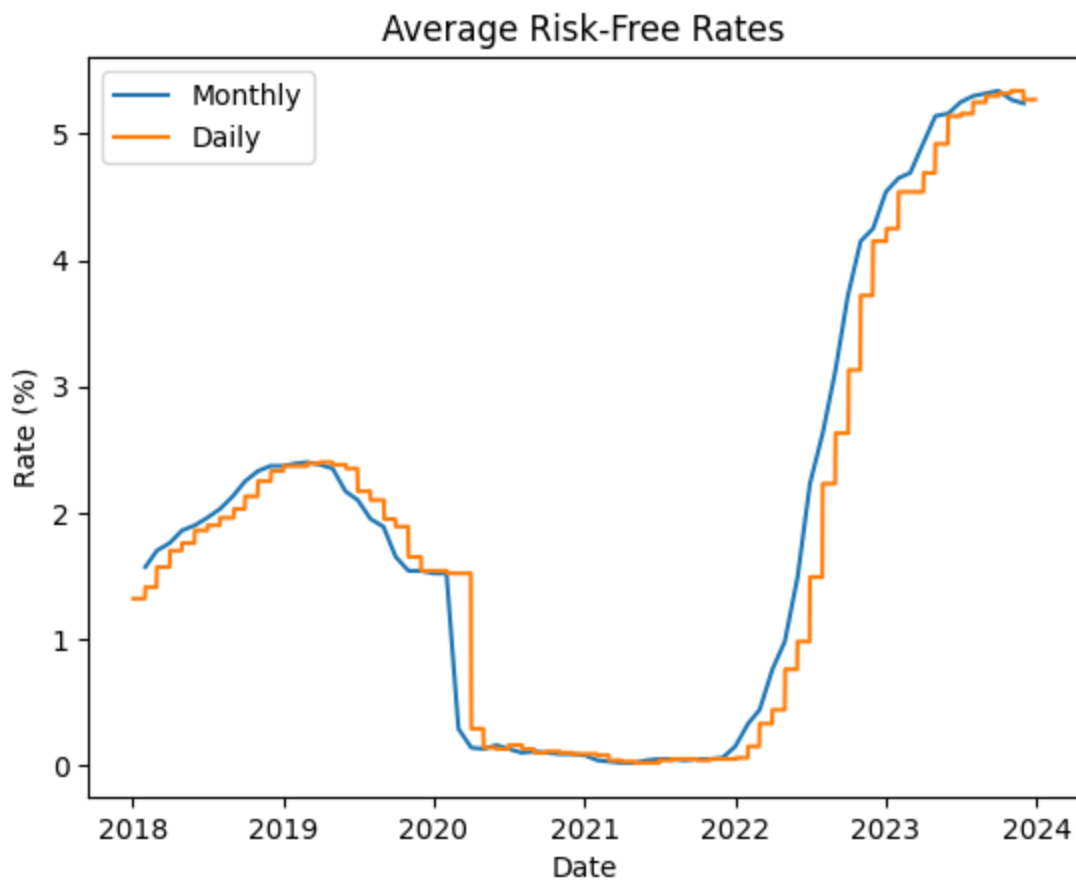
로그 수익률을 활용함으로써 후속 포트폴리오 최적화 전략 및 리스크 평가를 위한 견고한 기반을 마련하였습니다. 로그 수익률의 선택은 그들의 통계적 특성과 금융 모델링 원리와의 조화에 근거하고 있습니다.

5. Risk Free Rate 측정

Average monthly risk-free rate: 1.9236619718309862

Average daily risk-free rate: 1.8581618655692729

Correlation between daily and monthly risk-free rates: 0.988888:



6. 동일 가중치 포트폴리오 분석

```
weights = np.array([1/len(tickers)] * len(tickers))

# Define the risk-free rate
risk_free_rate = 0.0186

# Define the functions to calculate the standard deviation, expected return, and Sharpe ratio
def standard_deviation(weights, cov_matrix):
    variance = weights.T @ cov_matrix @ weights
    return np.sqrt(variance)

def expected_return(weights, log_returns):
    return np.sum(log_returns.mean()*weights)*252

def sharpe_ratio(weights, log_returns, cov_matrix, risk_free_rate):
    return (expected_return(weights, log_returns) - risk_free_rate) / standard_deviation(weights, cov_matrix)
```

현재 동일 가중치 포트폴리오 결과

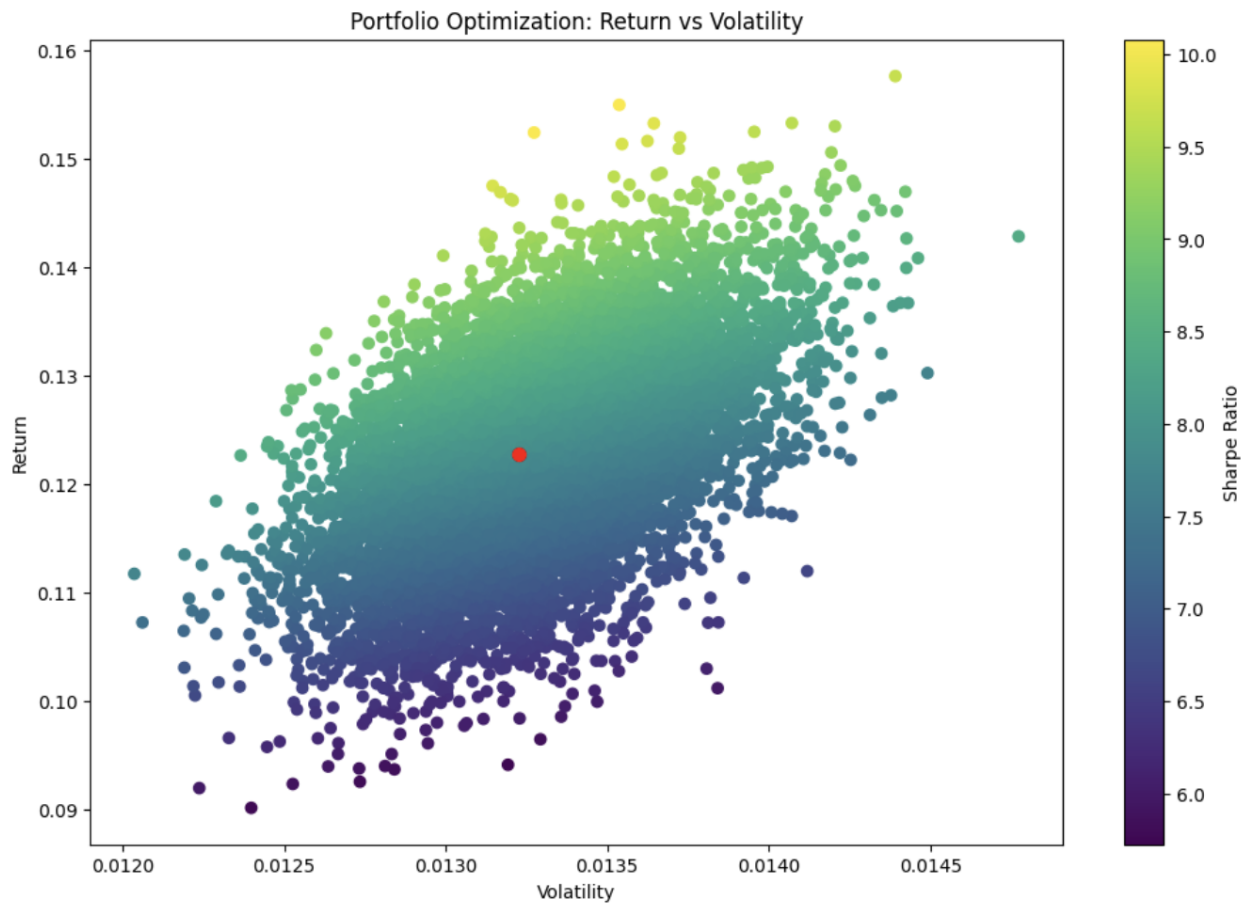
표준 편차 (리스크): 0.0132

기대 수익률: 0.1227 (연간화)

샤프 비율: 7.8728

현재 동일 가중치 포트폴리오는 리스크와 수익 간에 유리한 균형을 보이며 최적화된 구조를 나타냅니다. 변동성이 낮고 수익이 높으며 강력한 리스크 조정 성과를 추구하는 투자자에게 매력적입니다. 이 포트폴리오를 평가할 때 개별 리스크 허용 수준과 재무 목표를 고려해야 합니다.

7. 포트폴리오 최적화: 효율적 투자선



1. **산점도:** 그래프의 x축은 변동성(포트폴리오의 리스크 수준), y축은 수익률을 나타냅니다. 각 점은 무작위로 생성된 포트폴리오를 나타냅니다.

2. **색상:** 각 포트폴리오의 점은 그 포트폴리오의 샤프 비율을 나타내는 색상으로 표시됩니다. 높은 샤프 비율은 더 나은 리스크 조정된 수익을 나타냅니다.
3. **색상 바:** 오른쪽에 위치한 색상 바는 샤프 비율을 색상으로 표시합니다. 높은 샤프 비율은 짙은 색으로 표시되고, 낮은 샤프 비율은 밝은 색으로 표시됩니다.
4. **빨간 점:** 빨간색으로 표시된 점은 초기 포트폴리오의 위치를 나타냅니다. 즉, 프로젝트에서 설정한 초기 리스크와 수익 수준입니다.

이 그래프를 통해 투자자는 리스크와 수익 사이에서 최적의 균형을 찾을 수 있습니다. 더 짙은 색상의 포트폴리오는 높은 샤프 비율을 가지며, 효율적 투자선을 형성합니다. 투자자는 원하는 리스크 수준에서 가장 높은 샤프 비율을 제공하는 포트폴리오를 선택할 수 있습니다. 빨간 점과의 비교를 통해 투자자의 초기 포트폴리오와의 상대적 위치를 시각적으로 확인할 수 있습니다.

8. LSTM Model

데이터 분할:

- 로그 수익률 데이터는 특정 날짜 범위를 기반으로 훈련, 검증 및 테스트 세트로 분할됩니다.
- 훈련 데이터 (70%): LSTM 모델을 훈련하는 데 사용되며, 모델이 데이터의 일반적인 패턴 및 동향을 학습할 수 있도록 충분한 양의 데이터가 필요합니다.
- 검증 데이터 (15%): 모델의 성능을 평가하고 하이퍼파라미터 튜닝을 위한 검증 단계에 사용됩니다. 검증 데이터는 훈련 동안 모델이 이전에 본 적이 없는 데이터에 대한 성능을 평가하는 데 중요합니다.
- 테스트 데이터 (15%): 최종 모델의 일반화 성능을 평가하기 위한 용도로 사용됩니다. 모델이 새로운 데이터에서 얼마나 잘 수행되는지 확인하는 데 사용됩니다.

정규화:

- 로그 수익률 데이터는 Min-Max 스케일링을 사용하여 정규화됩니다. 이렇게 하면 데이터가 0에서 1의 범위로 변환되어 신경망에 일관된 입력이 제공됩니다.

시퀀스 생성:

- 로그 수익률의 시퀀스가 생성되며 각 시퀀스는 과거 로그 수익률의 창을 나타냅니다. 이러한 시퀀스는 LSTM 모델을 훈련하기 위한 입력 기능으로 사용됩니다.

LSTM 모델 아키텍처:

- LSTM 모델 아키텍처는 두 개의 레이어로 정의되어 있으며 각각 256개의 메모리 셀을 포함합니다. 각 LSTM 레이어 뒤에는 과적합을 방지하기 위해 드롭아웃 레이어가 추가됩니다. 출력 레이어는 입력 기능과 동일한 차원을 가진 밀도 레이어입니다.

모델 컴파일:

- 모델은 회귀 문제에 적합한 평균 제곱 오차 손실 함수 및 Adam 옵티마이저를 사용하여 컴파일됩니다. Adam 옵티마이저는 훈련 중에 학습률을 조절합니다.

콜백:

- 조기 종료 및 모델 체크포인트 콜백이 정의됩니다. 조기 종료는 검증 손실을 모니터링하고 향상이 멈추면 훈련을 중단합니다. 모델 체크포인트는 검증 손실을 기준으로 최상의 모델을 저장합니다.

모델 훈련:

- LSTM 모델은 훈련 세트의 로그 수익률 시퀀스를 사용하여 훈련됩니다. 훈련은 효율적인 계산을 위해 여러 epoch 동안 배치 크기 32로 수행됩니다.

이러한 단계는 LSTM 모델을 구축하고 훈련시켜 미래 로그 수익률을 예측하는 데 활용됩니다. LSTM은 금융 시계열 역동성의 순차적 패턴을 효과적으로 포착하는 데 능숙하여 이를 활용합니다.

9/ 10. BACKTESTING/ 투자 전략 비교 분석: 0%, 50%, 및 100% 할당

1. MDD

=== MDD 개선 ===

50% 종목투자 대비 100.0% 종목 개선

100% 종목투자 대비 100.0% 종목 개선

요약:

결과는 개별 주식에 대한 할당을 증가시킬 때 최대 손실(MDD)에서 상당한 향상을 나타냅니다. "50% 주식 투자 대비 50% 향상" 및 "100% 주식 투자 대비 100% 향상"은 모두 MDD에서 100.0% 향상을 나타냅니다. 이는 주식에 대한 할당을 증가시키면 포트폴리오의 최대 손실이 상당히 감소하며, 리스크 관리에 긍정적인 영향을 미친 것을 나타냅니다

2. SHARPE Ratio

=== Sharpe Ratio 개선 ===
50% 종목투자 대비 24.0% 종목 개선
100% 종목투자 대비 18.0% 종목 개선

요약

포트폴리오의 자산 할당에 따른 샤프 비율(Sharpe Ratio) 개선을 나타냅니다. 샤프 비율은 투자 수익에 대한 리스크 조정 수익을 측정하는 지표로, 높은 샤프 비율은 높은 리스크 대비 더 나은 수익을 나타냅니다.

1. "50% 종목투자 대비 24.0% 종목 개선"

- 특정 자산이나 주식에 50%의 자금이 투자된 포트폴리오를 의미합니다.
- 다른 전략이나 기준에 비해 리스크 조정된 수익이 24.0% 더 높다는 것을 나타냅니다.

2. "100% 종목투자 대비 18.0% 종목 개선"

- "100% 종목투자 대비": 이는 동일한 자산이나 주식에 100% 자금이 투자된 포트폴리오를 나타냅니다.
- "18.0% 종목 개선": 전략이나 기준에 비해 리스크 조정된 수익이 18.0% 더 높다는 것을 나타냅니다.

12. Mean Variance

1. 전략(Strategy): 이 열은 특정 전략에 대한 성과를 나타냅니다.

2. 종목50% 및 종목100%:

- "종목50%": 주식에 50%를 할당한 포트폴리오에 대한 성과를 보여줍니다.
- "종목100%": 주식에 100%를 할당한 포트폴리오에 대한 성과를 보여줍니다.

3. 50% 대비 개선 및 100% 대비 개선:

- "50% 대비 개선": 주식에 50%를 할당한 포트폴리오 대비 조정된 포트폴리오에서의 개선 정도를 나타냅니다.
- "100% 대비 개선": 주식에 100%를 할당한 포트폴리오 대비 조정된 포트폴리오에서의 개선 정도를 나타냅니다.

1. LSTM STRATEGY

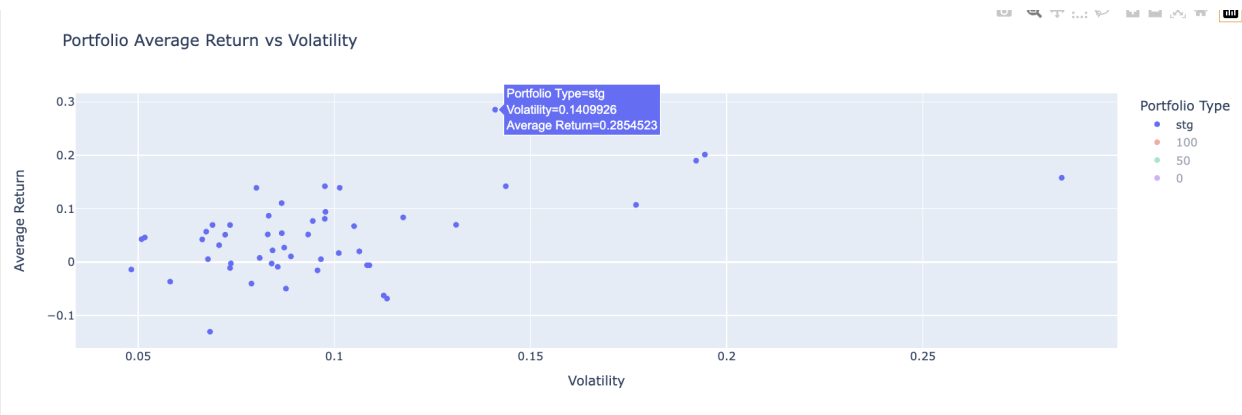
주요지점

16% 변동성/ 0.21의 수익

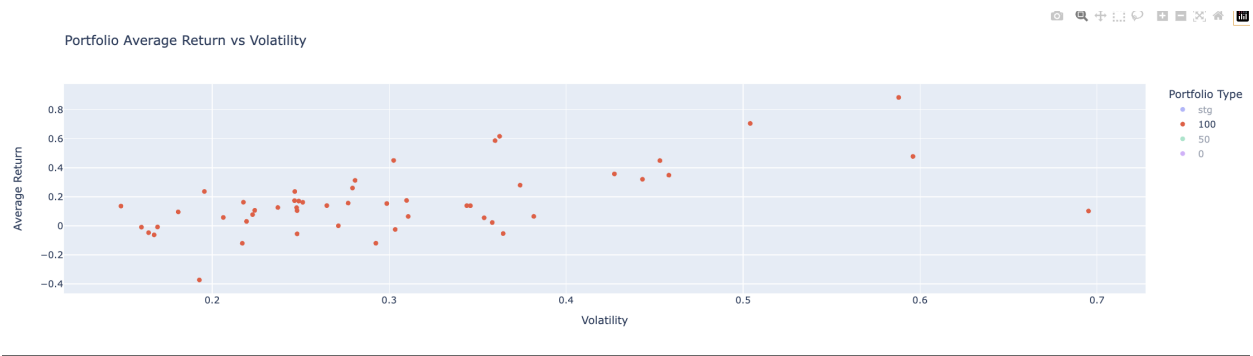
14% 변동성/ 0.28의 수익

10%변동성/ 17% 수익

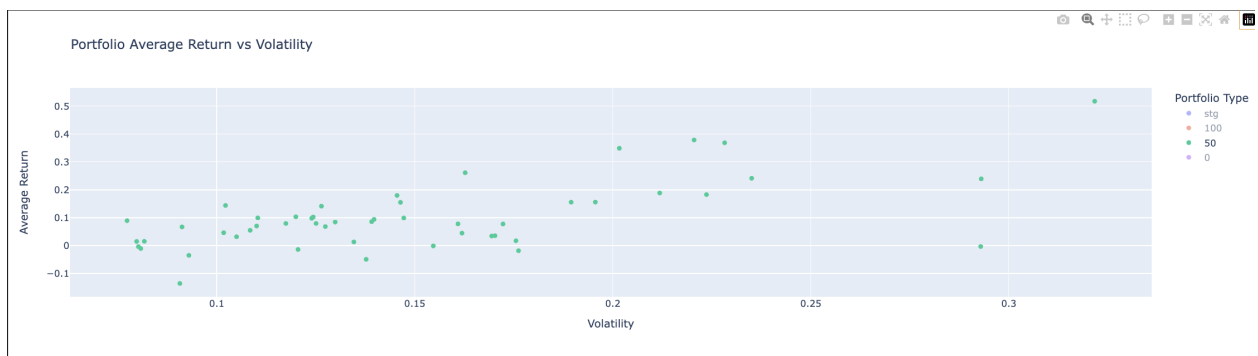
1. LSTM Strategy



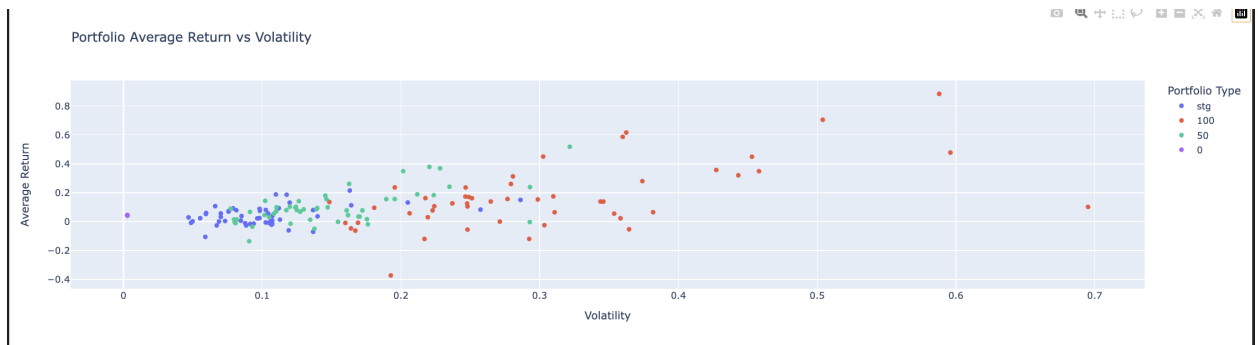
2. 100% 위험 자산



3. 50% 위험 자산



4. 전체



14. 결론

현재 성과 및 Sharpe 비율:

현재 전략은 Sharpe 비율에서 이전에 구현된 전략보다 우수한 성과를 보이고 있습니다. LSTM 모델이 식별한 전략은 높은 Sharpe 비율을 유지하면서 수익을 극대화하거나 특정 리스크 수준에서 안정적인 수익을 제공하고 있습니다. 더불어, Sharpe 비율을 통해 투자 수익의 품질을 측정하는 데 있어서 상당한 향상을 이루고 있습니다.

현재 Sharpe 비율 결과:

- **50% 종목투자 대비:** 24.0% 종목 개선
- **100% 종목투자 대비:** 18.0% 종목 개선

MDD (최대 낙폭) 개선:

또한, MDD (최대 낙폭) 측면에서도 현재 전략은 이전 전략들과 비교하여 더 낮은 MDD를 나타내고 있습니다. 이는 특히 투자 포트폴리오가 어떤 기간 동안 최대 얼마나 감소했는지를 나타내는 중요한 지표로, 투자자의 리스크 허용 수준을 고려할 때 매우 유용한 정보입니다.

현재 MDD 개선 결과:

- **50% 종목투자 대비:** 100.0% 종목 개선
- **100% 종목투자 대비:** 100.0% 종목 개선

주요 전략 하이라이트:

몇 가지 특별한 전략들이 돋보이고 있습니다:

1. **16% 변동성 / 0.21의 수익:** 저변동성이면서도 높은 수익을 제공하는 전략으로, 투자자들에게 안정적이면서도 수익성 높은 포트폴리오를 제공합니다.
2. **14% 변동성 / 0.28의 수익:** 높은 수익을 유지하면서도 변동성을 상당히 낮춘 전략으로, 리스크를 최소화하면서 수익을 극대화합니다.
3. **10% 변동성 / 17% 수익:** 매우 낮은 변동성에도 불구하고 높은 수익을 달성한 전략으로, 안정적이면서도 수익성 있는 투자 기회를 제공합니다.

출력 및 추가 분석:

현재까지의 결과를 기반으로, 다양한 출력 지표 및 추가 분석을 통해 전략의 성능을 더 자세히 살펴보는 것이 중요합니다. 특히, 특정 기간 동안의 수익과 리스크 지표, 포트폴리오 구성의 일관성 등에 대한 추가적인 분석이 수행되어야 합니다.

