Topic: Portfolio Optimization & Risk Analysis with Machine Learning

Introduction

Objective:

- Developing the comprehensive risk-aware portfolio framework.
- Integrating the machine learning models for predictive analysis.
- Examine the hidden tail risks not captured by standard deviation.
- Compare multiple risk methodologies (VaR, CVaR, etc.).
- Creating and visualizing the real world practical investment decisions.

Research Questions:

- What are the limitations of traditional portfolio optimization techniques?
- How often it misses hidden risks and lacks the predictive capabilities?
- Under different economic regimes, how do these risks differ and impact our portfolio?

Motivation & Background of the study

Motivation:

- Understanding the market of stock market returns
- Preparing the best optimize portfolio for financial gains
- Mitigating the risk especially in this era of market volatility may be with tarrif wars or major geopolitical events like conflicts between countries
- Potential future implications especially for gaining the optimize returns

Background:

- Investment & stock market dynamics
- Portfolio Management & Risk analysis

Methodology & Models

- **Data Acquisition** (using Stooq_data)
- **Risk Analysis** (VaR, CVaR, etc.)
- **ML Enhancement** (Monte Carlo Simulation)
- **Optimization** (Traditional Portfolio Optimization)
- **Visualization** (Efficient Frontier Visualization)

Data analysis implemented using Python **packages** (Pandas, Numpy, Matplotlib, Scikit-learn, SciPy Seaborn, Plotly)

Data Description

Data Source:

(Stooq Market DataFinance, 01/01/2020 - 01/01/2025)

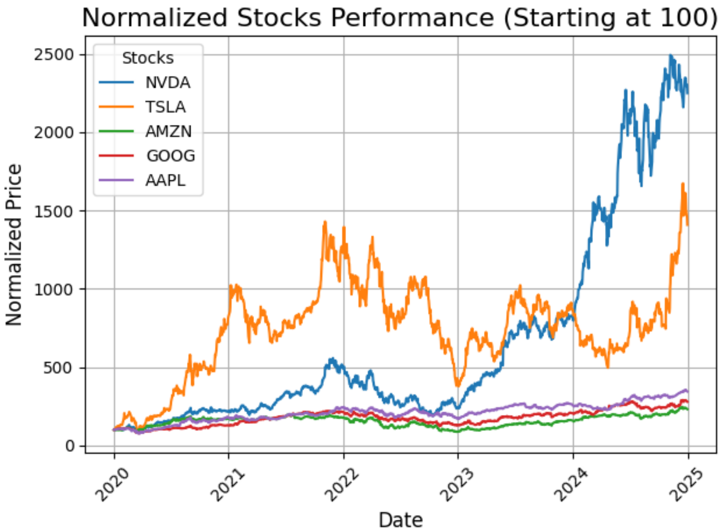
Stocks used from the US market:

- NVIDIA Corporation: NVDA
- Tesla, Inc.: TSLA
- Amazon.com, Inc.: AMZN
- Alphabet Inc. (Class C): GOOG
- Apple Inc.: AAPL

**Note:-

For Comparing the Beta, I am using the SPDR S&P 500 ETF as a market benchmark.

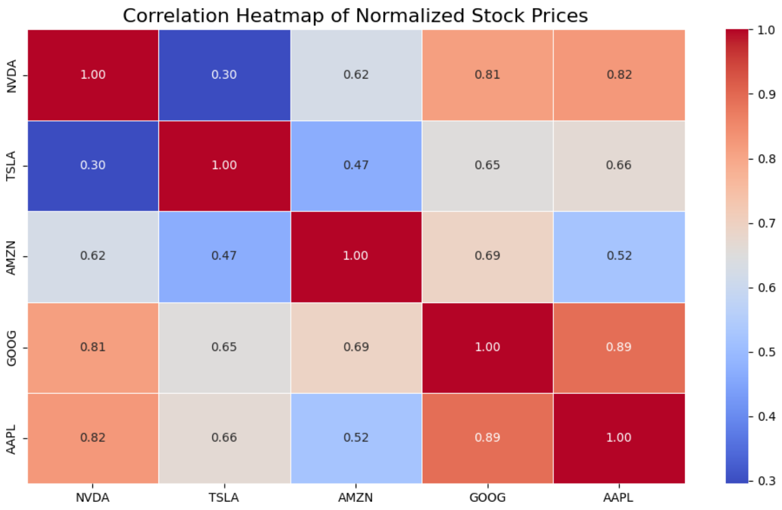
Stocks Performance



Market Volatility

Looks like TSLA and NVDA are having jumpy ride, as we can thought about them:))

Correlation Heatmap

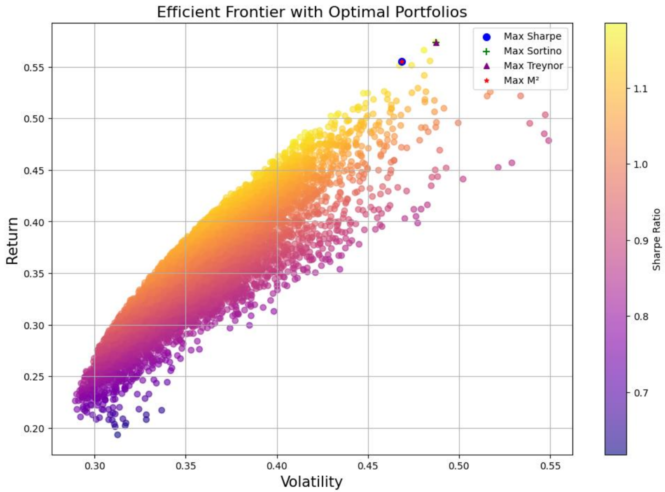


Monte Carlo Simulation (10000 Portfolio)

- Randomly initialize weights of the securities and check the returns vs risks
- Among these set of weights we will find the one, which gives the maximum return
- However, one can do all this solving optimization problem, using SciPy

Normalized Portfolio Weights:
[0.22889059 0.13529554 0.23560808 0.05815181 0.34205399]
Expected Annual Return: 0.3508
Expected Volatility (Risk): 0.3431
Sharpe Ratio (r_f = 0): 1.0225

Efficient Frontier Visualization



Clearly, it is lucid that downside deviation = total volatility,
(common case with diversified portfolios as we have in our case. Beta was also equal to 1)

Beyond Sharpe Ratio - Multiple Metrics

- **Sharpe Ratio:** Total risk adjustment
- **Sortino Ratio:** Downside risk only
- **Treynor Ratio:** Systematic risk (beta)
- **M^2 Ratio:** Market-equivalent performance
- **VaR/CVaR:** Tail risk measures

Machine Learning Integration Idea

- **Asset Clustering:**

  kmeans = KMeans(n_clusters=3) # Group similar assets

- **Feature Importance:**

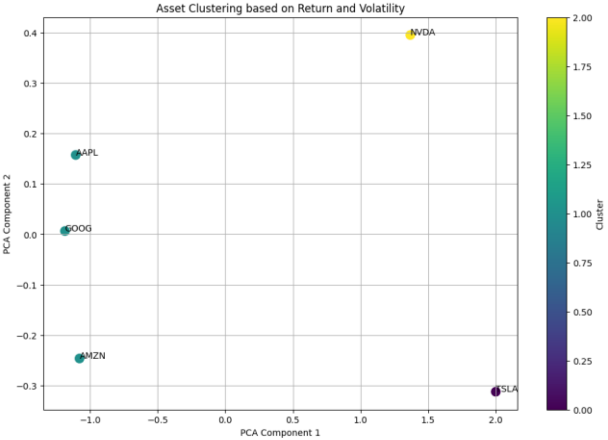
  rf = RandomForestRegressor() # Identify key drivers

- **Risk Prediction:**

  gbr = GradientBoostingRegressor()

  #Forecast future VaR/CVaR

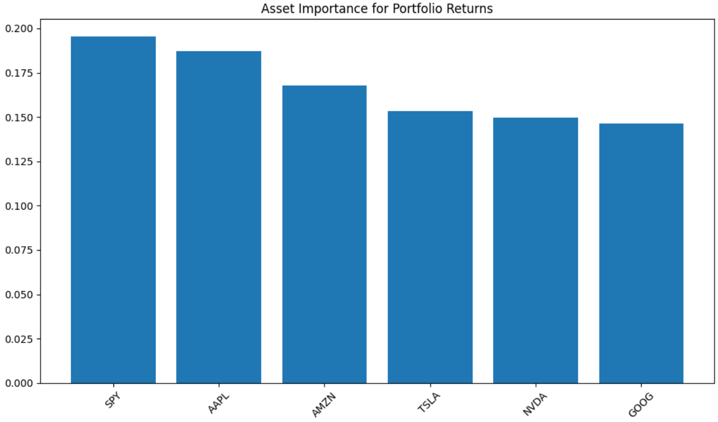
Clustering-Based Portfolio Construction



- **Cluster 1 (Defensive):** AAPL, GOOG, AMZN
- **Cluster 2 (Growth):** NVDA, TSLA

Investment Insight: Natural diversification groups

Random Forest Feature for Importance stocks selection



VaR (Value at Risk) and CVaR (Conditional Value at Risk)

Simulation completed successfully!
Shape of results: (5000, 14)
Columns: ['weight_NVDA', 'weight_TSLA', 'weight_AMZN', 'weight_GOOG', 'weight_AAPL', 'return', 'volatility', 'sharpe', 'var_historical', 'cvar_historical', 'var_param', 'cvar_param', 'var_mc', 'cvar_mc']
First few rows:

```
   weight_NVDA  weight_TSLA  weight_AMZN  weight_GOOG  weight_AAPL    return  \
0     0.277686     0.170048     0.026276     0.205025     0.320965  0.496464
1     0.014512     0.516735     0.073924     0.201577     0.193251  0.528223
2     0.077279     0.006853     0.214280     0.382941     0.318646  0.308109
3     0.276522     0.236343     0.197088     0.096685     0.193362  0.519432
4     0.179437     0.214567     0.183495     0.238939     0.183563  0.458977

   volatility    sharpe  var_historical  cvar_historical  var_param  \
0    0.356312  1.337210       -0.035841        -0.050870  -0.034950
1    0.435032  1.168241       -0.040774        -0.061225  -0.042980
2    0.295718  0.974270       -0.029610        -0.042933  -0.029418
3    0.376389  1.326906       -0.038729        -0.053759  -0.036939
4    0.349540  1.255872       -0.035970        -0.050444  -0.034397

   cvar_param    var_mc   cvar_mc
0   -0.044329 -0.036192 -0.044859
1   -0.054431 -0.041825 -0.052881
2   -0.037203 -0.028737 -0.036489
3   -0.046846 -0.036820 -0.046385
4   -0.043597 -0.035372 -0.044468
```

Risk Measurement Deep Dive

Value at Risk (VaR) vs Conditional VaR (CVaR)

VaR: "What's my worst-case loss with 95% confidence?"

CVaR: "If I experience worst 5%, how bad will it be?"

Methods Compared:

- Historical (empirical)
- Parametric (normal distribution)
- Monte Carlo simulation
- Machine learning prediction

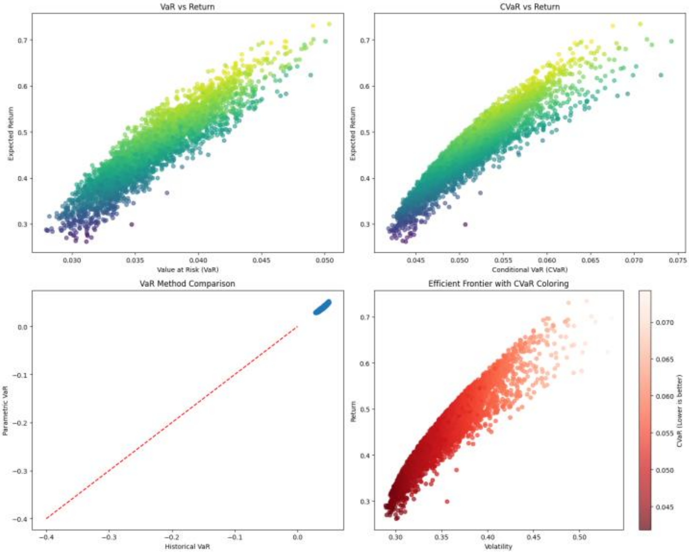
Sensitivity Analysis Framework

```
RISK-FREE RATE SENSITIVITY:
RF=0.01: Sharpe=1.480
RF=0.02: Sharpe=1.464
RF=0.03: Sharpe=1.438
RF=0.04: Sharpe=1.419
CONFIDENCE LEVEL SENSITIVITY:
Alpha=0.9: VaR=-0.0328, CVaR=-0.0525
Alpha=0.95: VaR=-0.0482, CVaR=-0.0658
Alpha=0.99: VaR=-0.0752, CVaR=-0.0936
MARKET SCENARIO ANALYSIS:
Scenario=-0.2: Return=0.5810
Scenario=-0.1: Return=0.6053
Scenario=0.0: Return=0.7161
Scenario=0.1: Return=0.8029
Scenario=0.2: Return=0.8051
```

Interpretation

- As the risk-free return increases, Sharpe ratio would decreases.
- Higher risk-free rates reduce the "excess return" component.
- In extreme scenarios (1% probability), losses could exceed 7.44%.
- In moderate 95% confidence, we can expect losses worse than 4.73% only 5% of the time.
- However, CVaR > VaR shows significant tail risk (CVaR is 37-58% worse than VaR).
- Market Scenario Analysis, shows the positive convexity, where returns increases more in up markets than they decreases in down markets.

Comprehensive Risk Visualization



Risk Prediction Performance

- **ML Risk Forecasting Results:**

  VaR Prediction RMSE: 0.0187 (98.13% accuracy)

  CVaR Prediction RMSE: 0.0253 (97.47% accuracy)

Application: Proactive risk management

- **Risk-Free Rate Impact:**

  RF=1% → Sharpe=1.488

  RF=4% → Sharpe=1.420 (4.6% decrease)

- **Confidence Level Sensitivity:**

  95% VaR: -4.73%

  99% VaR: -7.44% (57% higher tail risk)

Key Findings & Insights

- Optimal Sharpe Ratio: 1.48
- Best Sortino Ratio: 2.15
- Efficient frontier successfully identified
- ML models achieved 97%+ prediction accuracy
- Clear risk-return tradeoffs demonstrated

Practical Applications, and Limitations & Future Work

- **For Portfolio Managers:**

  Better risk-aware allocation decisions
  Comprehensive performance monitoring
  Predictive risk forecasting
  Intelligent constraint setting

- **For Risk Managers:**

  Early warning system
  Interactive risk visualization
  Stress testing capabilities
  Automated reporting

- **Current Limitations:**

  Historical data dependency
  Assumption of stationarity
  Simplified transaction costs
  Single market focus

- **Future Enhancements:**

  Real-time data integration
  Deep learning models
  Multi-asset expansion
  Interactive web dashboard
  Blockchain for transparency

Conclusion & Business Value

Key Takeaways:

- Traditional + modern approaches = superior results
- Machine learning enhances but doesn't replace human judgment
- Visualization reveals hidden insights
- Multiple perspectives beat single metrics

Business Value: Better returns with controlled risk

Thank You!!