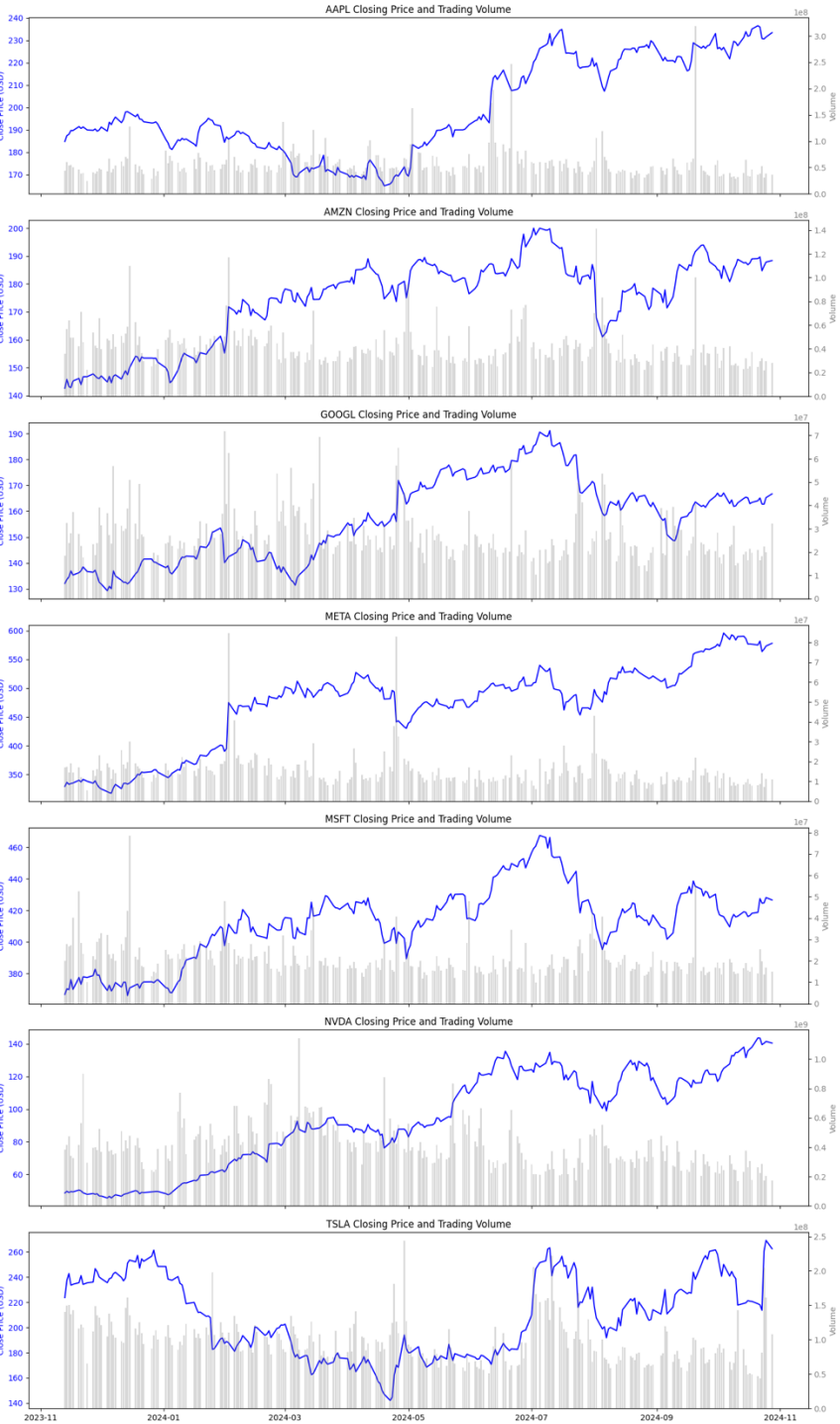
Stock Return Prediction with News Sentiment

Machine Learning applied to NASDAQ Tech Companies

Mauricio Aguas Fonseca

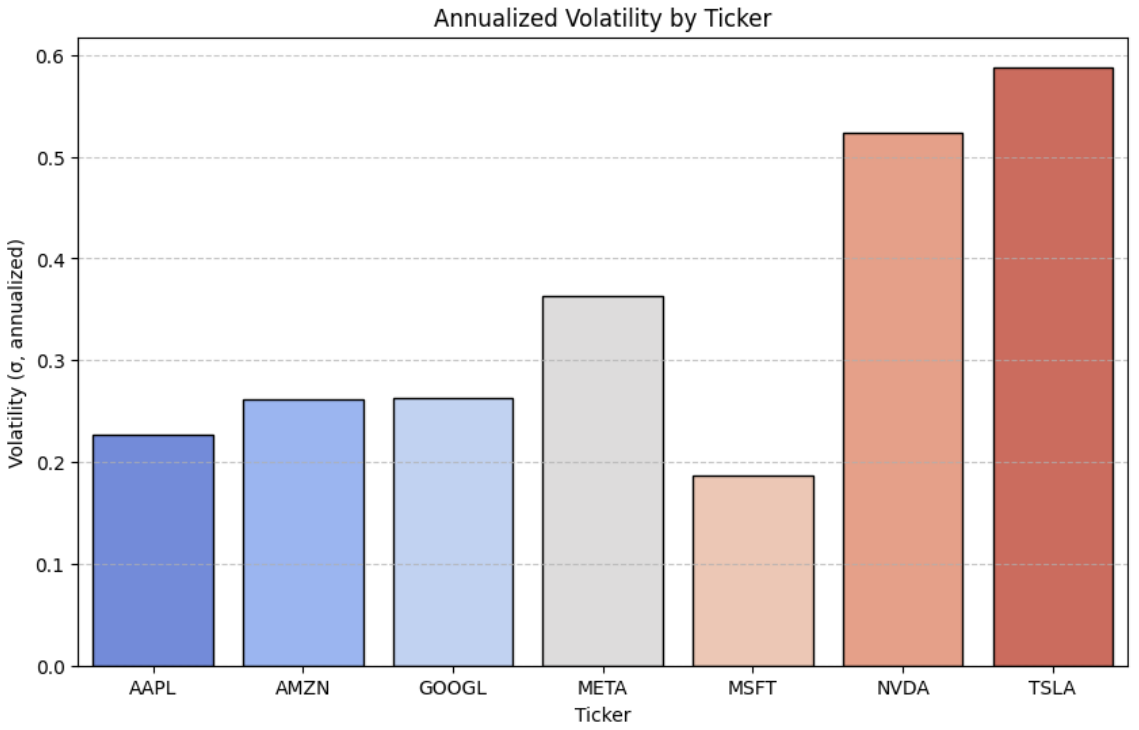


Problem & Motivation

- *Financial markets are noisy, predicting returns is challenging.*
- *Traditional models rely only on price/volume data.*
- *Hypothesis: incorporating **news sentiment** may improve predictability.*

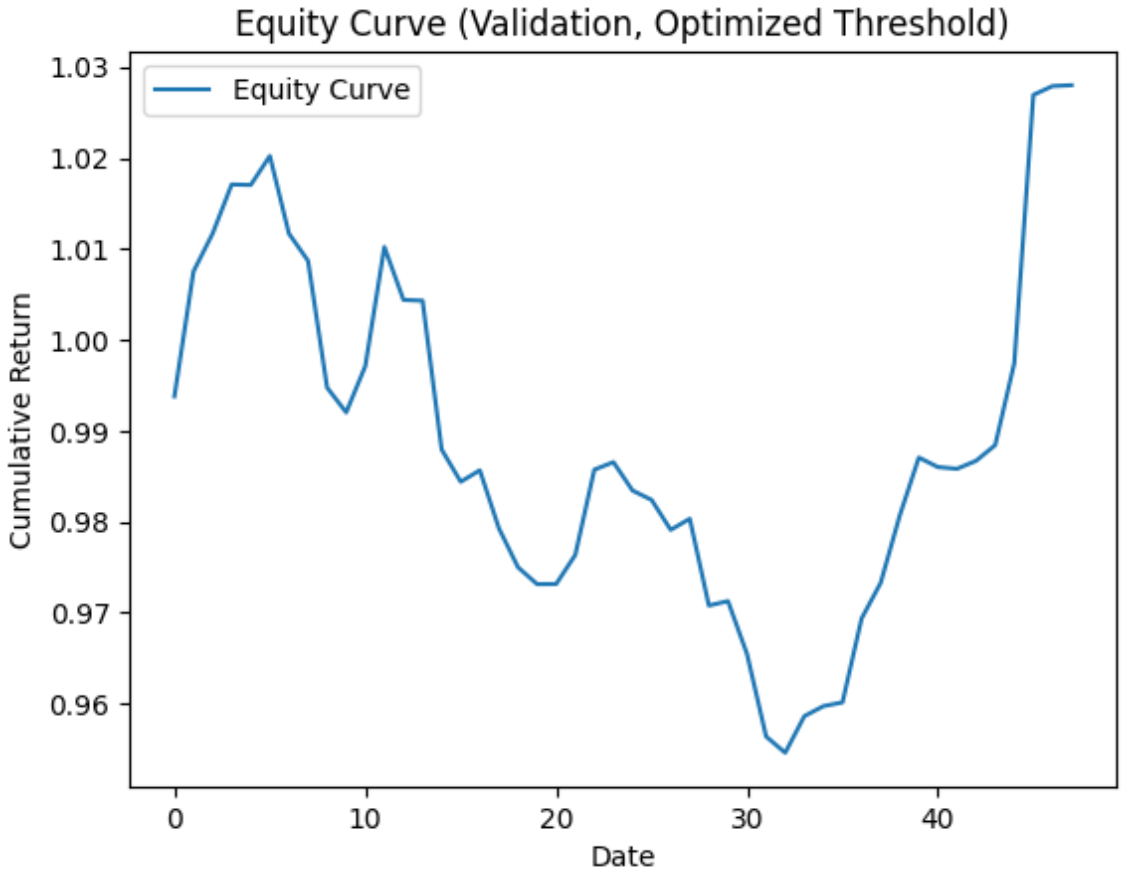
Methodology

- **EDA** → Correlations, distributions, volatility clustering.
- **Feature Engineering** → Returns, volatility, RSI, volume ratios, sentiment (VADER).
- **Model Development** → Logistic Regression, Random Forest, Gradient Boosting.
- **Backtesting** → Trading strategy with transaction costs.



Key Results

- **Classification:** ROC AUC ≈ 0.52 (weak but non-random signal).
- **Regression:** Hit Rate ≈ 54% (Lasso best).
- **Backtest (optimized threshold):**
- Annualized Return: **15.6%**
- Annualized Volatility: **11.8%**
- Sharpe Ratio: **1.32**
- Max Drawdown: **-4.5%**



Conclusions & Next Steps

- **Conclusions**
- Predictive power is weak but non-random.
- Sentiment adds information beyond prices.
- Positive Sharpe ratio confirms potential.
- **Next Steps**
- FinBERT sentiment for richer text analysis.
- GARCH/macro features for volatility modeling.
- Longer validation periods and more assets.