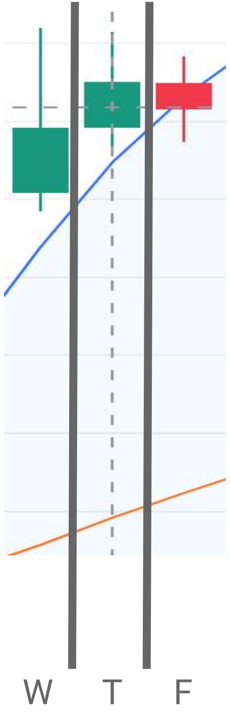
4Sigma - DDMIF Presentation

Data Used

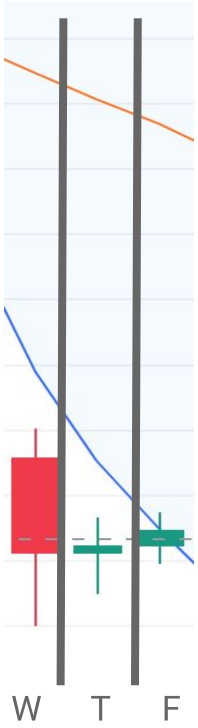
1. Historical Stock Price Data: Obtained using *yfinance*, this data provides the adjusted closing prices of stocks, reflecting their value over time including adjustments like dividends and stock splits.
2. Fama French Factors Data: Retrieved through the *getFamaFrenchFactors* package, this includes the market risk premium, size premium, value premium, profitability premium, and investment premium. These factors are essential for analyzing the historical returns of stocks in relation to broader market trends and characteristics.
3. Twitter Sentiment Data (Textual): Obtained using *twscrape*, this data was used in a limited fashion to make the ML models more robust.

Initial Strategy – Bollinger Bands

Initial Strategy – Bollinger Bands

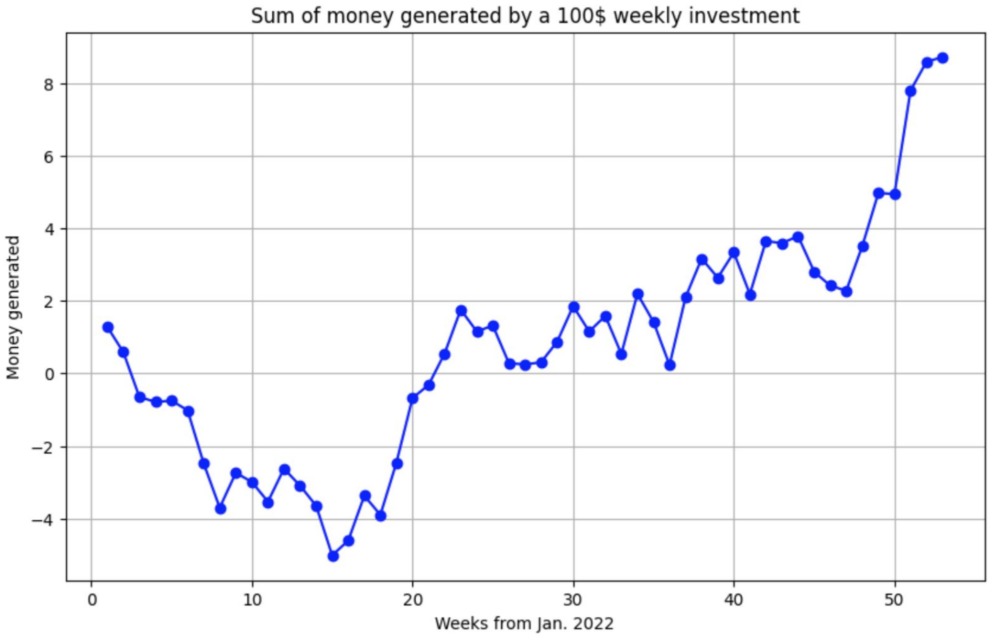


Selling signal !



Buying signal !

Initial strategy backtested



Limits of this strategy:

- Not a data-driven strategy
- 8$ (1 coffee in NYC) generated from a 100$ weekly investment
- Relying only on one indicator
- Constraints are too tight: only 3/4 buying/selling signals -> portfolio not enough diversified

Fama French 5 Factor Model

Why: Selected the Fama French 5 Factor Model due to its suitability for equities, aiming for a strong baseline to project future returns, and applied it with a monthly frequency for more accurate weekly return estimations.

Betas to predict with LR:

- monthly excess returns of the **Market (Mkt - R(f))**
- monthly excess returns of **Small-cap over Large-cap (SMB)**
- monthly excess returns of **value stocks over growth stocks (HML)**
- monthly excess returns of **most profitable firms over least profitable firms**
- monthly excess returns of firms that **invest conservatively over firms that invest aggressively**

Shortcomings: It is primarily designed for long-term equity analysis, may not capture short-term market fluctuations effectively, potentially limiting its accuracy in weekly investment decision-making.

Also wasn't able to target more than half of the investment pool

ML modelling

Models experimented with:

- Random Forest
- XGBoost
- LSTM
- Bi-directional LSTM

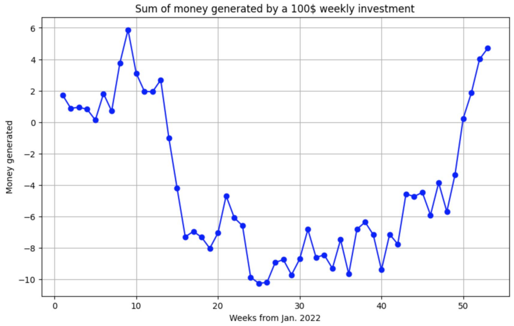
Data:

- Historical Stock Prices
- Volume, Intra-day volatility
- Momentum variations
- Holidays, Big events
- Sentiment Analysis (limited)
  - Twitter data

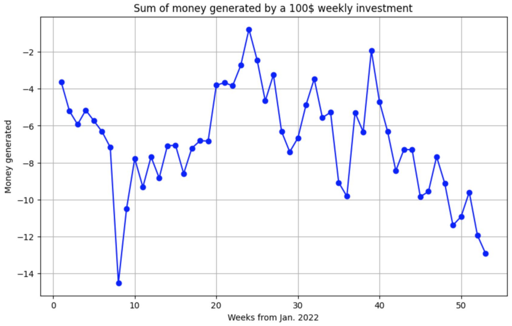
Results were not good or robust, feature importance seemed to varied a lot. Bollinger strategy was giving decent results.

ML modelling - Backtesting

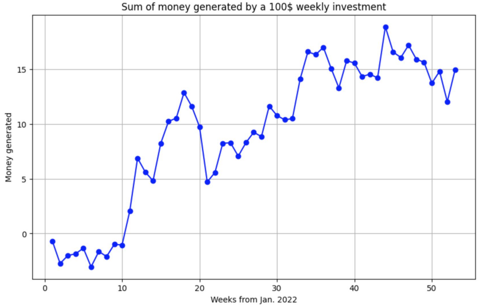
Random Forest



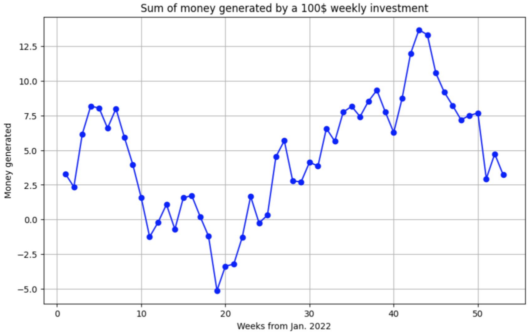
XGBoost



Bi-directional LSTM



LSTM



Final Strategy – Back to Bollinger Bands + RSI

Reading of some articles: RSI + BB = consistency

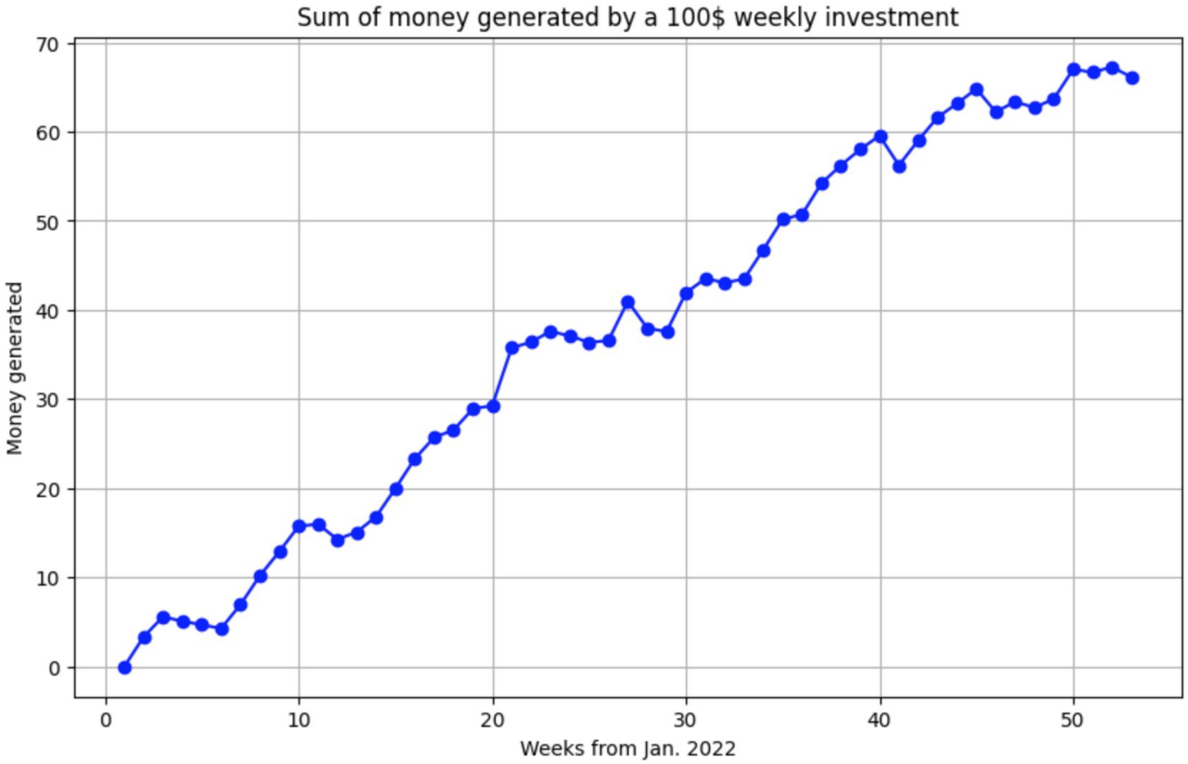
Z-score = (F-closing price - 14-day mean)/(14-day std)

RSI	LONG	SHORT
WINNING	37.7	72.5
LOSING	58.2	48.9

FINAL STRATEGY :

UNIVERSE $\xrightarrow{\text{Top 15 ABS(Z-score)}}$ BB $\xrightarrow[\text{Short (RSI>68)}]{\text{Long (RSI <40)}}$ RSI $\xrightarrow{\approx\text{10 stocks}}$ Final decision

Final Strategy – Back to Bollinger Bands + RSI



Forecasting - What we tried

We initially tried to base the forecasting component of the decision

- Designed a ranking algorithm that would assign the probability distribution for each asset based on the decision forecast
- Failed

We then tried to only forecast the assets we were going long/short in

- Ranking the assets under consideration with a high weight in rank 5 if we are going long, and a high weight in rank 1 if we are shorting
- Also failed

Forecasting - Finding the trick

- We noticed that the Equally Weighted Long portfolio consistently did very well in the forecasting part
- The S&P 500 also consistently did very well
- So our hypothesis of “hacking” the forecasting part is simply going equally weighted long or very close to it
- After going with that “strategy”, our forecasting performance has been mostly stable

Next steps

- Keep our decision strategy as we have been first in this competition for several weeks now
- Implement a better ranking strategy based on volatility - should be more predictable

Thank you!