III. DATA COLLECTION AND STRATEGY DEVELOPMENT

Any trading strategy requires a set of calculations on multiple data inputs to analyze large amounts of assets at a time. Such calculations includes selecting assets based on filtering rules, ranking assets based on a scoring function and calculating portfolio allocations. The first step in trading strategy development is data collection and preprocessing. Next step is of algorithmic development followed by strategy analysis. The last and most important step is backtesting against historical data and risk analysis. Since Iyka Trade is still in development, the algorithm development is done using Quantopian which is a similar product used for algorithmic trading research.

3.1 FINDING PATTERNS IN DAILY RETURNS AND PUBLIC SENTIMENT

When exploring a dataset, we try to look for patterns that might serve as the basis for a trading strategy. The plot shown is Figure 3.1 uses PsychSignal's Trader Mood dataset for Apple stocks which assigns bull and bear scores to stocks each day based on the aggregate public sentiment from messages posted on Stocktwits, a financial communications platform.

Figure 2.1 shows some matching spikes between daily returns, stocktwits message volume and semtiment scores. In some cases the direction of the spikes in daily returns match the direction of AAPL's sentiment score.

Figure 3.1: Plot of daily returns, message volume and sentiment score of Apple stocks.

Since there seems to be a correlation between daily returns, message volume and the underlying stock sentiment, the next section deals with creating a trading strategy using public sentiment as a technical indicator.

3.2 TRADING UNIVERSE SELECTION

An important part of developing a trading strategy is defining the set of assets that we want to consider trading in our portfolio. This set of assets is usually refered to as the trading universe. A trading universe should be as large as possible, while also excluding assets that aren't appropriate for portfolio. Quantopian's QTradableStocksUS universe offers this characteristic. Therefore, QTradableStocksUS is used as the trading universe for the algorithm. The time period for which the data is taken into consideration is of 3+ years starting from January 2017 to March 2020. This servers as an interesting time period as it allows for the analysis of the algorithm pre and post COVID-19 pandemic declaration.

3.3 CALCULATING SENTIMENT SCORE

The goal in this research is to build an algorithm that selects and trades assets based on sentiment data. PsychSignal's StockTwits Trader Mood Dataset is one such data collection. PsychSignal's dataset assigns bull and bear scores to stocks each day based on the aggregate sentiment from messages posted on Stocktwits, a financial communications platform.

The sentiment score is calculated using 3 day moving average over the PyschSignal's bull_minus_bear column. A moving average is a line used on charts to show the average price trailing a certain number of days back. The simple moving average, which is simply the sum of the past X number of quantities divided by the total number of quantities in the series.

The SMA is applied to the chart helps to smooth out the price action that has occurred over the period chosen. The table below shows sentiment scores calculated for the last 10 columns of the input dataset.

Date Time	Ticker	Sentiment Score
2019-11-01 00:00:00+00:00	Equity(50376 [CDEV])	-1.000000
	Equity(50680 [HLNE])	2.600000
	Equity(50718 [PUMP])	-1.116667
	Equity(50869 [IR])	-1.683333
	Equity(51338 [EYE])	-0.816667
	Equity(51634 [AVYA])	-1.640000
	Equity(51734 [VICI])	2.066667
	Equity(52119 [TCDA])	-1.788667
	Equity(52165 [DOMO])	-1.066667
	Equity(52209 [ALLK])	-1.400000

Table 3.1: Sentiment scores for last 10 rows of PsycSignal's Trader Mood dataset

3.4 LONG SHORT TRADING ALGORITHM USING SENTIMENT ANALYSIS

In general, long-short equity strategies consist of modeling the relative value of assets with respect to each other, and placing bets on the sets of assets that we are confident will increase (long) and decrease (short) the most in value. Long-short equity strategies profit as the spread in returns between the sets of high and low value assets increases. The quality of a long-short equity strategy relies entirely on the quality of its underling ranking model. In this algorithm a simple ranking schema is used.

The ranking schema considerers assets with a high 3 day average sentiment score as high value, and assets with a low 3 day average sentiment score as low value. Once a ranking scheme has been determined, we would like to be able to profit from it. We do this by investing an equal amount of money into the top of the ranking, and short into the bottom. This ensures that the strategy will make money proportionally to the quality of the ranking only, and will be market neutral.

Next, the algorithm uses Quantopian's open source factor analysis tool, Alphalens [16], to test the quality of the selection strategy. Alpha factors express a predictive relationship between some given set of information and future returns**.** By applying this relationship to multiple stocks we can hope to generate an alpha signal and trade off of it. Alphalens first combines factor and pricing data using a function called get_clean_factor_and_forward_returns. This function classifies our factor data into quantiles and computes forward returns for each security for multiple holding periods. We will separate our factor data into 2 quantiles (the top and bottom half), and use 1, 5 and 10 day holding periods.

Asset	1D	5D	10D	factor	factor_quantile
Equity(52033 [GSKY])	0.00662	0.01390	0.01854	-0.90666	1
Equity(52045 [PRSP])	0.00311	0.00272	0.05960	-2.11000	1
Equity(52155 [BV])	0.01409	0.01634	0.04227	2.45000	2
Equity(52209 [ALLK])	0.01339	-0.03009	-0.02098	-1.31166	1
Equity(52427 [EB])	0.00284	-0.01709	0.06381	-1.54000	1

Table 3.2: Factor analysis of last five rows of PyscSignal’s Trader Mood dataset

IV. RESULTS

4.1 STRATEGY ANALYSIS

Having the data in the format shown in Table 3.2 allows the use of Alphalens's analysis and plotting tools. First, we plot the mean returns by quantile over the entire period.

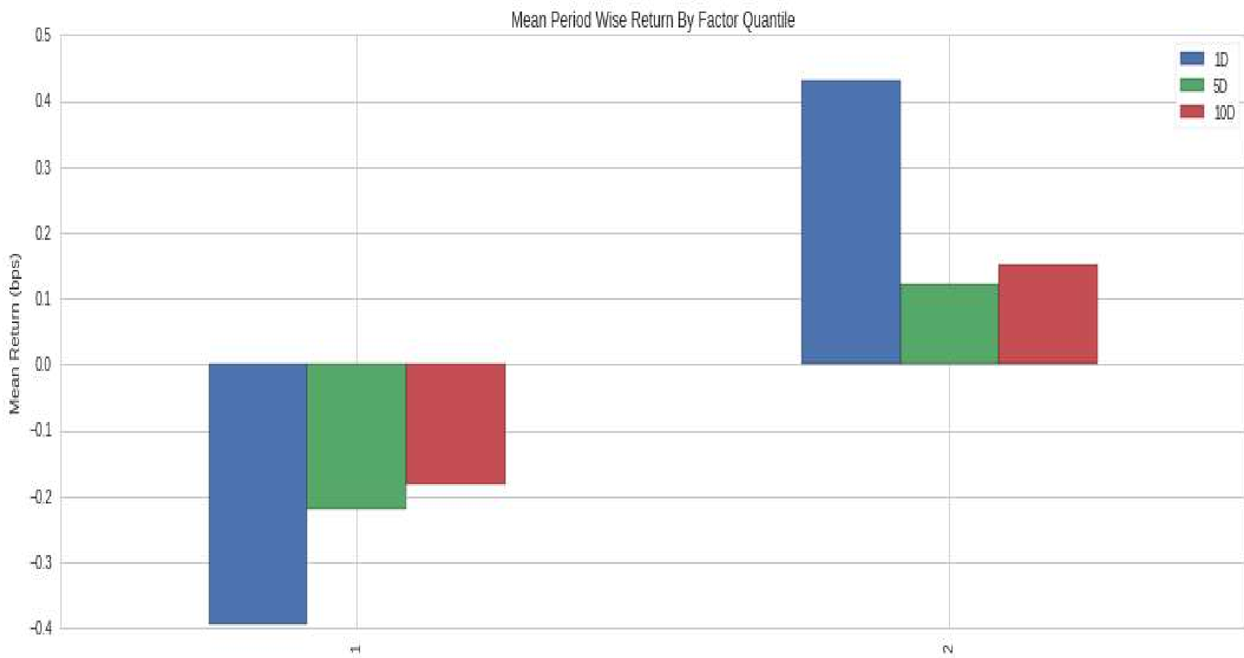

Figure 4.1: Mean period-wise return by factor quantile

Next, we plot the cumulative returns of a factor-weighted long-short portfolio with a 5 day holding period shown in Figure 4.2.

Figure 4.2: Cumulative returns for a 5 day holding period

Having created and tested a strategy, the next step is to Backtest the trading algorithm over Historical Data.

4.2 BACKTEST ANALYSIS

Backtesting simulation involves testing a trading strategy on historical data. It estimates the strategy's practicality and profitability on past data, validating it for success or failure or any needed changes. The simulation is performed on the porfolio data for the time period starting from January 2017 to December 2109 with an initial investment of 1 million dollars. The plot in Figure 4.3 shows shows the rolling beta of the strategy against benchmark returns over the entire period of the backtest. In this instance, the benchmark return of the SPY was used. Thus, the lower the rolling portfolio beta to the SPY, the more market neutral an algorithm is. One of the reasons to construct a long-short equity trading algorithm is to maintain a low correlation to the market, so we want this plot to be consistent around 0 over the entire backtesting period. It is evident from the plot below that the algorithm does maintain close to 0 over the entire backtest period.

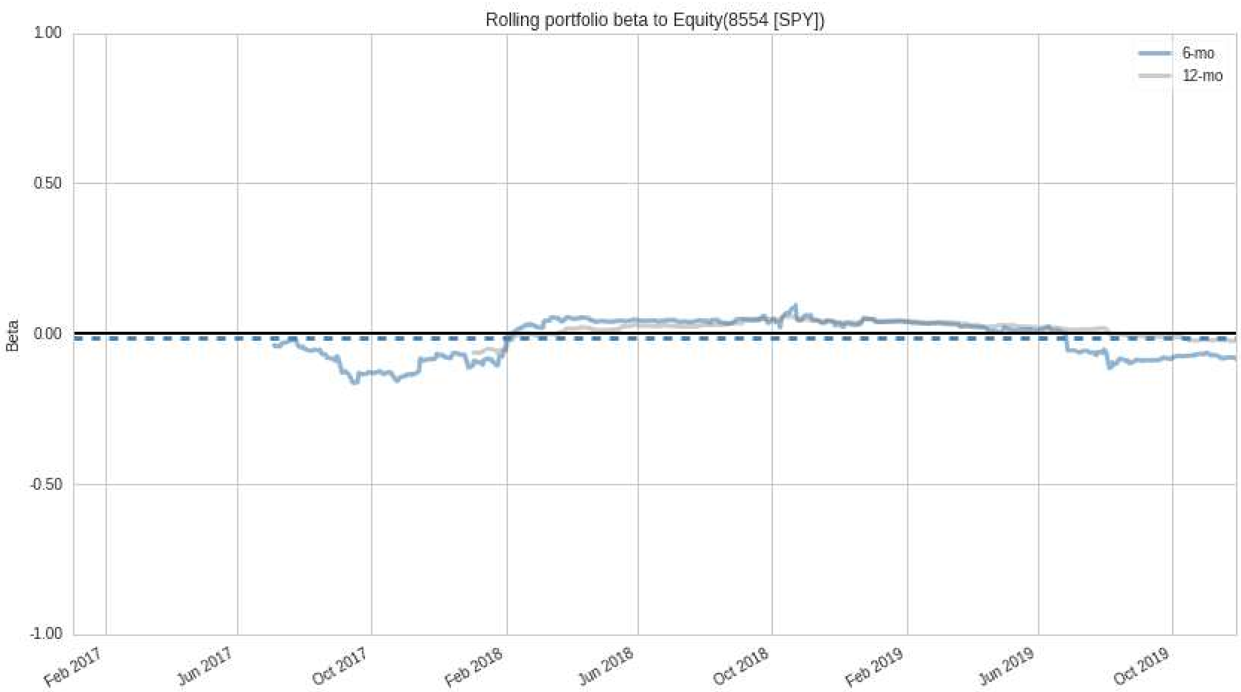

Figure 4.3: Rolling Portfolio Beta against benchmark SPY index.

Next, we plot the cumulative returns as it allows one to gain a quick overview of the algorithm's performance and pick out any anomalies across the time period of the backtest. The cumulative return plot also allows you to make a comparison against benchmark returns - this could be against another investment strategy or an index like the S&P 500. Figure 4.4 shows the performance of the algorithm against SPY (S&P 500) index.

Figure 4.4: Algorithm performance over benchmark index SPY

The plot shown in Figure 4.5 uses Quantopian's Risk Model to illustrate how much of the returns can be attributed to the strategy, and how much of it comes from common risk factors. The Quantopian Risk Model [18] is a way to see what risks your algorithm is exposed to, and whether those risks are expected and managed. The risk model decomposes the risk of holding any stock or portfolio into a set of common risk factors and a residual risk. That residual risk is called specific risk. Colloquially, the return from specific risk is also referred to as alpha.

We can see below from the graph in Figure 4.5 that most of our portfolio's total returns come from specific returns. This suggests the algorithm's performance isn't coming from exposure to common risk factors.

Figure 4.5: Time series of cumulative returns

4.3 TOP LONG AND SHORT POSITIONS

The goal of each algorithm is to minimize the proportion of the portfolio invested in each security at any time point. This prevents the movement of any individual security from having a significant impact on the portfolio as a whole. The bigger the exposure a strategy has to any security, the greater the risk. Table 4.1 shows the percentage of returns attributed to common and specific risk factors. Tables 4.2 and 4.3 list the top 10 long and short positions of all time. Table 4.4 breaks down the risk exposure to individual sectors.

Summary Statistics	Percentage
Annualized Specific Return	1.17%
Annualized Common Return	-0.35%
Annualized Total Return	0.80%
Specific Sharpe Ratio	0.33

Table 4.1: Summary statistics of portfolio

Top 10 long positions of all time	max
RETA-49995	2.27%
TTD-50288	2.25%
PI-50138	2.10%
BOOM-1034	1.99%
MDCA-12800	1.94%
GTHX-50879	1.92%
ENVA-47979	1.88%
EHTH-32726	1.84%
HCC-50780	1.83%
DENN-18148	1.80%

Table 4.2: Top 10 long positions of all time

Top 10 short positions of all time	max
TRUP-47331	-2.00%
GSHD-51962	-1.93%
ARCB-41	-1.86%
CMPR-27674	-1.85%
TGI-15905	-1.85%
SYNH-48027	-1.84%
GNW-26323	-1.81%
UFS-2329	-1.79%
SSD-11386	-1.79%
FORM-25182	-1.76%

Table 4.3: Top 10 short positions of all time.

Exposures Summary	Average Risk Factor Exposure	Annualized Return	Cumulative Return
basic_materials	0.01	-0.15%	-0.43%
consumer_cyclical	0.01	0.33%	0.97%
financial_services	-0.04	-0.30%	-0.88%
real_estate	-0.02	-0.15%	-0.43%
consumer_defensive	0.01	-0.03%	-0.09%
health_care	0.03	0.37%	1.07%
utilities	0.01	-0.05%	-0.15%
communication_services	-0.00	-0.01%	-0.03%
energy	0.00	-1.17%	-3.36%
industrials	-0.00	0.19%	0.55%
technology	0.01	0.51%	1.49%
momentum	0.11	0.25%	0.74%
size	0.14	0.26%	0.76%
value	-0.10	0.18%	0.51%
short_term_reversal	-0.12	-0.34%	-0.99%

Table 4.4: Risk exposure summary of portfolio