

**MSDS 451 - Final Project**

Ethan Vanlerberghe

School of Professional Studies, Northwestern University

MSDS 451: Financial Engineering

Dr. Miller

31 August 2025



**NORTHWESTERN**  
**UNIVERSITY**

## **Introduction**

The purpose of this document is to examine the Sentiment-Flex (SF) Fund and elaborate on the research supporting it. It outlines the general investment philosophy guiding the fund's creation, describes the methodology by which the fund operates, details the securities included, provides a brief evaluation of the fund's performance for potential shareholders, and discusses whether I would personally choose to utilize the fund.

The SF fund was created to demonstrate the power of insights gathered from the volume of posts created every day on Social Media. Millions of users interact with Social media platforms such as Reddit every day, leaving behind insights into how they see the world. SF was created to capitalize on these comments and to utilize them to examine future market behavior.

## **Literature review**

A substantial amount of research has already been conducted in this area, providing a rich foundation for developing data-driven investment strategies. Notable examples can be found in open-source GitHub projects (Hansen-han 2025; Shirosaidev 2025; Shapo 2025; ScarletSplendour 2025; AlpacaHQ 2023), which demonstrate use of similar methods. In these projects, financial data is retrieved from Yahoo Finance using the YFinance API, and Monte Carlo simulations are applied to model uncertainty in future market behavior. Sentiment analysis, typically sourced from social media platforms such as Twitter or Reddit, is then incorporated to adjust asset valuations or inform portfolio weighting decisions. Without integrating backtested historical data, sentiment alone would be insufficient for accurately forecasting future performance.

The foundations for SF were laid in Modern Portfolio Theory (MPT), which provides the mathematical and technical analysis backbone to all quant-style financial projects. Complementing MPT, (Grinold and Kahn 1995) and (Ross 2016) emphasize quantitative approaches to selecting assets, measuring risk, and evaluating performance in real-world markets. (Wilmott 2008) and (Miller 2025)

further provides practical insights into the implementation of quantitative models, highlighting common pitfalls and considerations for complex financial products, such as SF.

In addition to open-source implementations, several professional sources provide insight into the construction and performance of social sentiment-driven ETFs, such as the VanEck \$BUZZ Social Sentiment ETF. Regulatory filings, including the official \$BUZZ prospectus (VanEck 2025), outline the fund's investment objectives, portfolio holdings, and management strategies. Other resources from VanEck, like their fund overview (VanEck 2023), and sector allocation summaries (Invesco 2025), provide detailed information on asset composition and sector weightings, illustrating how sentiment analysis can be “Weaponized” against the market to search for greater alpha values. These sources demonstrate the application of sentiment-based investment strategies in a professional fund context, highlighting how theoretical models can be translated into real-world financial products.

### **General Investment Philosophy**

The Sentiment-Flex (SF) fund is a growth-structured momentum-based trading fund building off the top 10 assets in the NYSE exchange by market capitalization. By leveraging the top 10 assets by market capitalization, SF seeks to take advantage of the free and extremely valuable consumer sentiment information on Social Media, refine it, and use it to drive actionable insights in predicting the future value of an asset. By following this program, SF is designed to make quicker decisions based on live consumer sentiment data that other funds may not see until formal news announcements are made. At its core, the SF fund believes in the power of “Collective Conviction” (VanEck 2025)—the aggregated sentiment of the investing public as reflected across social media platforms. By systematically analyzing this sentiment, SF interprets the crowd's convictions as a leading indicator of potential asset price movements. When a significant majority of social media users exhibit a consistent directional sentiment toward an asset, SF treats this as an indicator of the broader market's perception, which can influence that asset's valuation.

Momentum and consistency are pivotal to SF's approach. Unlike some other sentiment-based strategies, if an asset receives positive "Collective Conviction" one hour, and negative the next, then SF does not consider this to be the direction of the overall consumer sentiment, and chooses not to participate in this movement. SF searches for a continued, consistent sentiment before making an action. This allows the SF fund to follow an informed momentum-based growth strategy and make more calculated risk-management decisions.

Risk Management and Adaptability are embedded in SF's methodology. By continuously monitoring sentiment signals across multiple assets, SF makes adjustments to mitigate risk to the shareholder and maximize value in a safe environment. The fund's design is intended to capitalize on the speed advantage of live sentiment data, potentially identifying market trends before they are reflected in traditional news channels or broader analyst coverage.

### **Investment methods**

After SF gathers the top 10 traded securities by market capitalization, the fund begins by analyzing real-time consumer sentiment using Python's AsyncPraw package to access the Reddit API. For each asset, Reddit posts are collected and processed through the VADER Lexicon and the SentimentIntensityAnalyzer (SIA) to assess both the polarity (positive or negative) and intensity of sentiment (Zhong 2025). The model reaches back to the start of 2022, as this is when the popular Reddit SubReddit "WallStreetBets" hit a critical enough user mass for it to be considered as a reliable platform for financial information (Volpicelli 2022). The resulting sentiment scores from "WallStreetBets" are then appended to the daily price table as additional columns, providing enriched input data for predictive modeling.

SF then begins to simulate the market over time, utilizing Monte Carlo simulations to simulate potential returns and closing prices for the various assets with a multivariate method (Miller 2025). SF then runs these new tables, including the sentiment, through a test and train split to set the groundwork for

predictive modeling. SF creates a flag to indicate next-day movement, with a 1 symbolizing “Up” and a 0 symbolizing “Down” using XGBoost (Hansen-Hen 2025). These are then converted to trading signals, with a 1 indicating a long position and a -1 indicating a short position. This modeled performance of the fund is then tested against the test set utilizing an accuracy method of comparison.

Finally, SF is tested against the 20% test set and plotted for visualization to a potential Shareholder. This visualization helps a potential Shareholder see the value that SF provides over a benchmark investment, in this case, the S&P 500. SF also provides key performance indicators to a potential Shareholder regarding portfolio performance to provide further consumer confidence in the index.

### **Description of securities**

The Sentiment-Flex (SF) Fund focuses on the top 10 most highly valued assets actively traded on the NYSE, as determined by market capitalization, at the start of each trading period. These assets are selected to ensure sufficient social media activity, which is critical for accurate sentiment analysis (Raghav 2025). Once the top 10 assets are identified, each is processed through SF’s sentiment-driven predictive model. The model evaluates social media posts related to each asset to generate sentiment scores, which are then used to predict near-term price movements and drive the overall portfolio’s decision-making.

In the following periods, the fund dynamically adjusts its holdings based on changes in market capitalization and sentiment. If an asset falls out of the top 10, it is replaced by the next highest-valued asset. This approach ensures that the fund consistently maintains exposure to the most widely-discussed securities, allowing the predictive model to operate on assets with sufficient social chatter to generate actionable signals (Raghav 2025). As these signals are collected in the following periods, they can

indicate to the overall fund if they should be switched from a positive sentiment (long) or a negative sentiment (short) in terms of holdings on the asset.

### **Performance evaluation**

Overall, the Sentiment-Flex (SF) Fund demonstrates positive performance over the evaluated period. From 2022 through 2024, SF produced a total return of 12.48%, compared to the S&P 500 benchmark return of 11.52%, representing a 0.96% outperformance (See Fig. 1). This demonstrates that the fund's sentiment-driven, momentum-based strategy can capture value above the market baseline by systematically analyzing the sentiment of consumers on social media platforms.

The SF Fund also exhibits strong risk-adjusted performance. Its Sharpe ratio of 1.3868 compares favorably against the benchmark Sharpe ratio of 0.3417, indicating a 1.045 improvement over the S&P 500 (See Fig. 1). This metric reflects that SF delivers superior returns relative to the risk a Shareholder would take, highlighting the strength of its predictive and momentum-based trading rules (Ross 2016). The fund's Alpha value of 0.0008 confirms that SF slightly outperforms the benchmark on a risk-adjusted basis, representing incremental returns (See Fig. 1). These returns are decently strong, especially for a strategy based less on trading fundamentals and quantitative data, and more on qualitative data. The Beta score of -0.0390 further demonstrates that SF exhibits lower overall market risk compared to the S&P 500, suggesting that the fund's returns are less sensitive to market-wide fluctuations, regime changes, or other influences on the market (Ross 2016).

Collectively, these performance metrics indicate that the SF strategy is capable of generating consistent returns with controlled risk exposure, validating the advantages of integrating social media sentiment analysis and machine learning-driven predictive modeling (See Fig. 2). The results show that actionable insights derived from consumer sentiment can contribute meaningfully to portfolio performance (Ross 2016).

## **Management recommendation**

The Sentiment-Flex (SF) Fund represents a unique angle to leverage real-time consumer sentiment in the financial markets. Given its demonstrated ability to outperform the S&P 500 in both total return and risk-adjusted metrics, launching a financial services firm built around SF could be a viable business venture if more structure were created around it. A formalized business plan, a Go-To-Market Strategy, and a market saturation study would need to be conducted first. The fund's focus on large, blue-cap assets, combined with momentum-based trading rules, positions it well to differentiate itself from traditional funds, offering investors a new approach to capturing market trends.

If I were to start a company around this fund, my preferred role would be Chief Investment Officer (CIO) or Fund Manager, directly overseeing the implementation of the sentiment analysis and machine learning models in trading decisions. This role would allow me to ensure that the fund's quantitative strategies are applied the way that they were designed, and that risk management procedures are continuously monitored and kept low (RealTrading 2023). I would also maintain involvement in research and development to refine predictive models as social media trends and market dynamics evolve. I also feel that I would have a strong leveraged position as a member of Generation Z to start this fund. More than any other generation, Generation Z has seen the power of social media and has tried to drive different pieces of it for good (Koetsier 2024). I feel that utilizing the data collected from these Social Media platforms could be a way for members of my generation to extract another positive aspect out of Social Media.

For individuals who may not want to start a firm themselves, the SF Fund could also serve as an attractive item. As a Shareholder, one could gain exposure to a sentiment-focused strategy that has the potential to deliver consistent returns above the market while maintaining low risk. The fund's structured approach, combining sentiment analysis, Monte Carlo backdating simulations, and momentum-based

trading, could be an effective way to build personal wealth over time by leveraging widely available data to accomplish an unconventional task (Abrahram, 2015).

I, however, would not invest in the SF fund. I am an extremely risk-averse investor, only choosing to invest in Treasury Bonds or CD's. The thought process behind being riskier than this is alarming to me, and therefore, I never think I could do it. However, that does not mean I am not confident in my work. I am simply too unsure of any quant's work to follow their predictive model with my life savings.

In addition to this, the annual fee income of \$824.09 from the fund is not enough to cover the upfront cost of managing the portfolio, let alone paying staff to do so (See Fig. 3). To make this worthwhile, the fees would have to be considerably higher, or the fund would need much more development to perform at a higher level.



Figure 1: Portfolio Performance Metrics

Performance Metric	SF Fund Score	Benchmark (S&P 500) Score
Total Return (20% Timeframe)	12.48%	11.52%
Sharpe Ratio	1.3868	0.3417
Alpha (Relative)	0.0008	N/a
Beta (Relative)	0.0390	N/a

Figure 2: Portfolio Performance

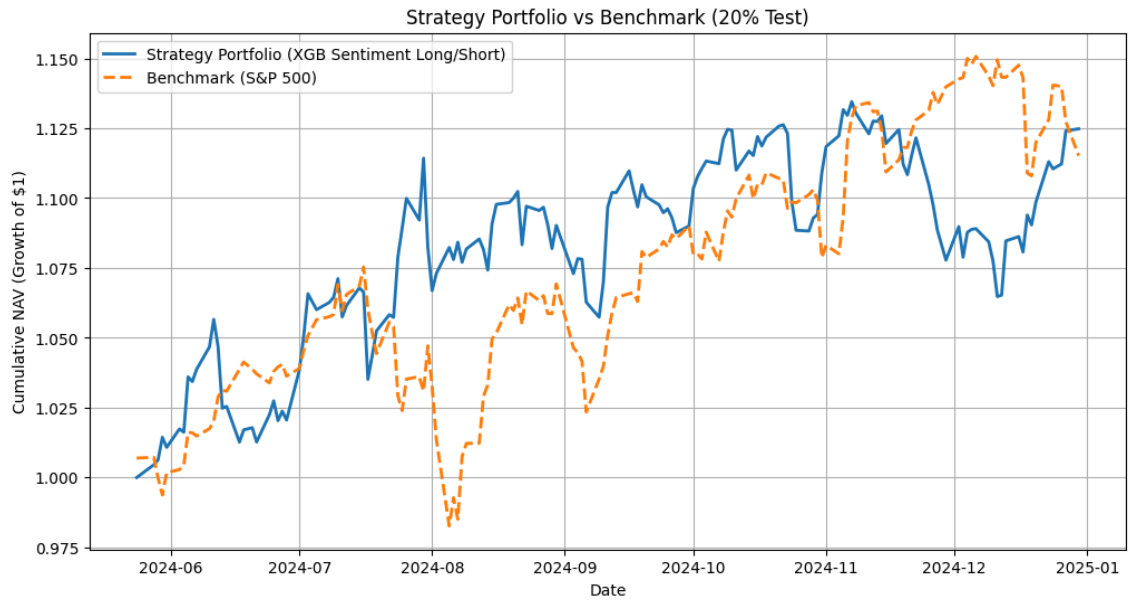


Figure 3: Fee Payout

Fee Type	Fee Amount	Total Earnings
Management	\$0.02	\$823.97
Performance	\$0.20	\$0.12
Total	N/a	\$824.09

## Bibliography

Abraham, Stephan A. “How to Create Your Own ETF.” *Investopedia*, April 1, 2015.  
<https://www.investopedia.com/articles/investing/040115/how-create-your-very-own-etf.asp>.

Adams, Douglas. *The Hitchhiker’s Guide to the Galaxy*. Pan Books, 1979.

Alós-Ferrer, Carlos, and Federico Iriberri. “The Effect of Social Media on Financial Markets.” *Journal of Economic Behavior & Organization* 154 (2018): 292–312.  
<https://www.sciencedirect.com/science/article/abs/pii/S1042443118303093>.

AlpacaHQ. *Momentum Trading Example*. GitHub, 2023.  
<https://github.com/alpcahq/Momentum-Trading-Example>.

Dierking, Dave. “ETFs 101: My Basic Criteria for Choosing an ETF.” *Seeking Alpha*, December 17, 2018.  
<https://seekingalpha.com/instablog/5366071-dave-dierking-cfa/5210721-etfs-101-basic-criteria-for-choosing-etf>.

Grinold, Richard C., and Ronald N. Kahn. *Active Portfolio Management: Quantitative Theory and Applications*. Probus, 1995.

Hansen-han. “alpaca\_sentiment\_trader.” GitHub, 2025.  
[https://github.com/hansen-han/alpaca\\_sentiment\\_trader](https://github.com/hansen-han/alpaca_sentiment_trader).

“Holdings & Sector Allocations of Invesco QQQ.” Invesco.  
<https://www.invesco.com/qqq-etf/en/about.html>.

Koetsier, John. “Gen Z Dumping Google for TikTok, Instagram as Social Search Wins.” *Forbes*, March 11, 2024.  
<https://www.forbes.com/sites/johnkoetsier/2024/03/11/genz-dumping-google-for-tiktok-instagram-as-social-search-wins/>.

Miller, Tom. “451 Feature Engineering: Programming Assignment 1.” 2025.

—. “Week 6 Term Project Checkpoint B.” Lecture, 2025.

“Modern Portfolio Theory.” Wikipedia. Wikimedia Foundation, July 29, 2025.  
[https://en.wikipedia.org/wiki/Modern\\_portfolio\\_theory](https://en.wikipedia.org/wiki/Modern_portfolio_theory).

Quimbayo, C. A. Z. “Robust Bayesian Portfolio Optimization.” *Journal of Operations Management*, 2025. <https://doi.org/10.1016/j.jom.2025.05.004>.

Raghav. “MonteCarloSimStockPrices.” GitHub, 2025.  
<https://github.com/RaghavsScarletSplendour/MonteCarloSimStockPrices>.

Real Trading™ Staff. “How to Start a Trading Office and Make It Profitable.” *Real Trading*, April 28, 2023. <https://realtrading.com/trading-blog/start-profitable-trading-floor-2/>.

Ross, Sheldon M. *An Elementary Introduction to Mathematical Finance*. Cram101, 2016.

Shapo, Misha. “Simulating Stock Prices with Monte Carlo Methods.” GitHub, 2025.  
[https://github.com/MishaShapo/Monte\\_Carlo\\_Stocks/blob/master/Simulating\\_Stock\\_Prices\\_with\\_Monte\\_Carlo\\_Methods.ipynb](https://github.com/MishaShapo/Monte_Carlo_Stocks/blob/master/Simulating_Stock_Prices_with_Monte_Carlo_Methods.ipynb).

—. “Simulating Stock Prices with Monte Carlo Methods.” YouTube, 2025.  
<https://www.youtube.com/watch?v=6-dhdMDiYWQ>.

—. “Stock Trading Principles.” YouTube, 2025.  
<https://www.youtube.com/playlist?list=PLvcbyUQ5t0UHdm6bNx3Rnj1fdpt9sGNsm>.

Shirosaidev. “stock sight.” GitHub, 2025. <https://github.com/shirosaidev/stock sight>.

Thompson, Cedric. “Fundamental vs. Technical Analysis: What’s the Difference?” *Investopedia*.  
<https://www.investopedia.com/ask/answers/difference-between-fundamental-and-technical-analysis/>.

VanEck. “\$BUZZ Prospectus.” *SEC.gov*.  
<https://www.sec.gov/Archives/edgar/data/1137360/000113736025000087/vanecksocialsentimentetfbu.htm>.

—. “Buzz– VanEck Social Sentiment ETF: Holdings & Performance.” VanEck, June 2, 2023.  
<https://www.vaneck.com/us/en/investments/social-sentiment-etf-buzz/>.

Volpicelli, Gian M. “This Was the Year When Finance Jumped the Doge.” *Wired*, December 29, 2021.  
<https://www.wired.com/story/defi-gamestop-memes-doge-musk/>.

Wilmott, Paul. *Frequently Asked Questions in Quantitative Finance*. Playaway Digital Audio, 2008.

Zhong, Sam. “Stock\_Trading\_Reddit.” GitHub repository, 2025.  
[https://github.com/Sam120204/Stock\\_Trading\\_Reddit](https://github.com/Sam120204/Stock_Trading_Reddit).