Sentiment Hedging: Trending Stocks on Reddit

Germán Nova

Advisor: Enrique ter Horst

June 21, 2023

Abstract

The present document studies the hedging of portfolios in the face of social media sentiment. To do this, a portfolio of stocks is constructed whose returns are correlated with innovations in social media sentiment. To estimate sentiment, two approaches are presented: the first with supervised analysis using the Multinomial Inverse Regression (MNIR) model (Taddy, 2013a) and the second with the lexicon-based Valence Aware Dictionary for Sentiment Reasoning (VADER) model (Hutto and Gilbert, 2014). To estimate sentiment, a database of 56.5 million comments and posts from Reddit is used. To generate the hedging asset allocation, Deep Reinforcement Learning is implemented, specifically the Adaptive Deep Deterministic Policy Gradient (Adaptive DDPG) algorithm (Li et al., 2019). It is found that the way sentiment is estimated determines the algorithm's hedging performance to a large extent. For the portfolio that is covered against sentiment estimated by MNIR, notable out-of-sample results are achieved. To highlight the versatility of the methodology, algorithm performance is presented by focusing on maximizing returns. In summary, the best performing portfolios, in terms of return, were those that incorporated social media sentiment analysis. Additionally, portfolios utilizing the Adaptive DDPG algorithm showed a better Sharpe ratio compared to an Equal Weighted Portfolio and individual stock investments.

Keywords: Sentiment Analysis, Deep Reinforcement Learning, Asset Allocation

1 Introduction

The effect on financial returns of social media sentiment, which investors convey by expressing their opinions about markets, has been widely studied. The existence of this effect suggests that investor opinion on social media can have an impact on the performance of financial assets. From an economic theory perspective, the study of this topic also deals with the validity of Fama's Efficient Market Hypothesis (EMH) (1970), which holds that stock prices fully reflect all available market information. Rational investors respond by choosing assets that reduce idiosyncratic or diversifiable risk. In the case of irrational investors appearing, rational investors can use arbitrage to correct stock prices. In response, behavioral finance theory suggests that the sentiments of irrational investors have an effect on asset prices. In some cases, due to intrinsic characteristics of assets, arbitrage by rational investors becomes costly and therefore, assets maintain their price without being valued according to EMH (McGurk et al., 2020). In that sense, social media can act as a platform to enhance the effect of irrational investor sentiment on asset returns.

From an economic theory perspective, the debate continues. A part of the literature does not find evidence in favor of behavioral finance. Renault (2020) studies, using a StockTwits dataset and traditional machine learning algorithms such as Random Forest and Neural Networks, the predictive value of social media sentiment on high-cap stocks. The author finds that while sentiment and returns are highly correlated, there is no evidence that sentiment helps predict daily stock returns. Chacon et al. (2023) analyze whether a simple strategy of following investment recommendations for long and short purchases on the r/wallstreetbets subreddit can generate alpha. While the authors find a positive relationship between the number of posts and abnormal volume, they do not find the strategy to be profitable on a risk-adjusted basis.

In contrast, another part of the literature has fueled the debate by finding results that question the validity of the EMH. Dhaoui and Bensalah (2016) find that a Fama-French 5-factor model (Fama and French, 2015) that incorporates a sentiment indicator performs better in predicting returns in the US market. Creamer and Houlihan (2020) find that sentiment measures based on social media can be used as a risk factor in asset pricing models and that these measures have predictive power when used as variables in machine learning models. Costola et al. (2021) define a risk factor for stocks that are characterized by high volumes of activity on social media and find that it is significant and positively related to stock returns. Ranco et al. (2015) find a significant dependence between Twitter sentiment and abnormal returns during peaks of Twitter activity. McGurk et al. (2020) find that social media sentiment has a positive and significant effect on abnormal returns. Koratamaddi et al. (2021) propose a Deep Reinforcement Learning approach to generate an automated trader that leverages the price and sentiment history of assets on social media. The authors find more robust performance in terms of Sharpe Ratio and annualized return compared to comparable strategies that do not include sentiment.

One of the most widely shared hypotheses about how the mechanism by which investor sentiment can affect returns works is through the spread of information and the consolidation of an opinion on one or several companies. Dong and Gil-Bazo (2020) find that the relationship between social media sentiment and stock returns is mainly explained by positive sentiment and non-professional investors. As this information spreads, more investors are attracted to the movement in a process known in the literature as herd behavior (Scharfstein and Stein, 1990; Klein, 2021). This process implicitly suggests that stock returns can be abnormally affected by the collective opinion that investors hold. Although the literature has traditionally investigated the effect of investor sentiment using social networks such as Twitter and StockTwits, a particularity that is part of this mechanism has been mainly found thanks to another social network called Reddit. With the emergence of social networks, individual investors have shown that they can participate massively in the market to the point of dramatically affecting the price of assets that become trends (Umar et al., 2021b; Hasso et al., 2021). Specifically, the reasons why the opinion of a stock or a group of stocks gains traction may be non-rational, in the sense that they are not justified by corporate results or purely financial opinions, but by emotional factors (Black, 1986). Since the events on Reddit in 2021, the literature tends to refer to stocks that exhibit these behaviors due to their high exposure on social media as ”meme stocks” (Nobanee and Ellili, 2023).

Meme stocks are characterized by significant disturbances due to the massive transactions of individual investors. On Reddit, there are several examples of stocks that have experienced abnormal returns due to the activities of investors in their subreddits. One of the most well-known examples is Gamestop at the beginning of 2021. During this period, large investment funds suffered losses as they were exposed to short positions in this stock. The losses of firms with short positions in Gamestop during 2021 are estimated at US$70 billion (Sujata, 2021). One fund alone, Melvin Capital, had a loss of 53% (Chung, 2021). In May 2022, the fund announced its closure and the return of funds to its clients (Herbst-Bayliss, 2022). It is important to mention that these types of events are not exclusively restricted to January 2021. Some similar events that also originate from financial subreddits date back to April 2021 (McCrank et al., 2021), August 2022 (Bowman, 2022), and February 2023 (Hughes, 2023).

Given the uncertainty surrounding the intensity of these types of actions, an important factor that portfolio managers must consider is their ability to hedge against social media sentiment. The relevance and increased interest of investors in this issue can be seen recently by the fact that some fund managers have publicly launched ETFs that allocate based on investor sentiment. Among these are the VanEck $BUZZ and Roundhill Investments $MEME ETFs.[1].

Given the recent context of the role of social media sentiment in asset returns, this paper proposes to approach the problem from a perspective of hedging against unexpected changes in investor sentiment. To demonstrate the added value of Reddit investor sentiment, its contribution in terms of return for an asset allocation strategy is also analyzed. In this way, an asset allocation strategy can be studied that can take advantage of social media information to anticipate market movements. Among the reviewed literature, a similar approach can be found in Engle et al. (2020). The authors develop a hedge against environmental sentiment. In that research, a portfolio mimicking methodology (Lamont, 2021) is explored to generate a portfolio that optimizes the correlation between portfolio returns and innovations in an environmental news sentiment indicator. Unlike this approach, the present paper proposes several changes to the methodology used. First, the effect of social media sentiment from Reddit, a social network whose effects on asset returns, to the best of my knowledge, has not been studied to the extent of other social networks such as Twitter, is studied. Second, a sentiment analysis methodology is proposed to create the indicator that takes into account the lexical difference between Reddit and other social networks. This is to alleviate possible prediction errors that pre-trained models with other social networks have. Third, a dynamic asset allocation based on Deep Reinforcement Learning is proposed. In the words of Sutton and Barto (2018), Reinforcement Learning involves mapping states to actions in order to maximize a numerical reward. Deep Reinforcement Learning adds Deep Learning in order to improve the performance of agents. In particular, it is proposed to model an environment with a state defined by continuous variables defined by asset prices, social media sentiment, and some descriptive portfolio variables. The agent's actions are the rebalancing weights of the portfolio. The numerical reward of the agent rewards portfolio hedging against unexpected changes in social media sentiment.

The methodology proposed in this document consists of the following. The first part of the methodology addresses sentiment analysis. The proposed model for classifying the sentiment of comments is the Multinomial Inverse Regression (MNIR) model (Taddy, 2013a). This is a parametric model that estimates a sentiment score for each comment based on the relative and absolute frequencies of each token, whether unigram, bigram, or trigram, from the universe of comments. To train the model, a sample of 4000 comments whose sentiment is classified following McGurk et al. (2020) is used. This part of the methodology has a benefit. Reddit may handle a different lexicon than other social networks, and training a model with it allows for identifying the most relevant words that some pre-trained models with other social networks may not take into account. However, this classification can depend heavily on the subjective opinion of the classifier. Therefore, as a comparison and to have a reference model, the Valence Aware Dictionary for Sentiment Reasoning (VADER) model (Hutto and Gilbert, 2014), which is a lexicon-based and rule-based model specialized in classifying Twitter comments, is also used. With each of these models, a sentiment indicator is estimated, which aims to show the aggregated sentiment of investors on social media over time.

[1]More information about these ETFs can be found at the following links: $BUZZ: https://www.vaneck.com/us/en/investments/social-sentiment-etf-buzz/ $MEME: https://www.roundhillinvestments.com/etf/meme/

For the second part, a variant of the Deep Deterministic Policy Gradient (DDPG) algorithm (Lillicrap et al., 2015), called Adaptive DDPG (Li et al., 2019), is proposed for portfolio asset allocation. This is a Deep Reinforcement Learning algorithm that allows modeling environments with continuous actions and states. The variant makes the algorithm better adapted to the regime changes in financial markets and has already been used in the literature for asset allocation by including social media sentiment in the state definition (Koratamaddi et al., 2021). To define the environment that allows the execution of the algorithm, Engle et al.'s perspective (2020) is used. Therefore, the correlation between innovations in the sentiment indicator and portfolio returns is proposed as a reward.

Regarding the data, the Adaptive DDPG algorithm can be trained using asset returns and sentiments over time. To estimate sentiments, text preprocessing is applied to a sample of 56.5 million comments and posts from January 2016 to June 2022, which come from the most relevant financial subreddits in terms of number of comments and users. This data was obtained using the Pushshift API (Baumgartner et al., 2020), a public database with comments and posts from the selected subreddits. Asset price data is obtained from Yahoo! Finance.

The results of the proposed methodology contribute to the economic theory debate. It is found that the way sentiments are estimated largely determines the hedge performance of the algorithm. In the case of the portfolio that hedges against sentiments estimated using MNIR, significant out-of-sample results are achieved. To highlight the validity of the sentiment estimation, the algorithm's performance is exposed to a return maximization objective. The best performing portfolios, in terms of return, were those that incorporated social media sentiment analysis. Additionally, portfolios utilizing the Adaptive DDPG algorithm showed a better Sharpe ratio compared to an Equal Weighted Portfolio and individual stock investments.

As for investors today, they face an ever-evolving landscape, with social media playing a significant role in shaping market sentiment and influencing stock prices. To effectively navigate this dynamic environment, it is crucial to harness the power of advanced technologies such as deep reinforcement learning and sentiment analysis. Deep reinforcement learning, an artificial intelligence technique, enables investors to develop robust trading strategies by learning from historical data and iteratively optimizing decision-making processes. By leveraging this cutting-edge approach, investors can identify patterns, capture market dynamics, and make more informed investment decisions.

Furthermore, sentiment analysis methods have emerged as valuable tools for evaluating social media sentiment towards specific stocks or market trends. By analyzing the sentiment expressed in social media posts, news articles, and other online sources, investors gain valuable insights into market sentiment and can gauge the potential impact on stock prices. Combining deep reinforcement learning with sentiment analysis allows investors to hedge against social media risk.

Understanding the importance of these innovative methods empowers investors to proactively respond to rapidly changing market conditions. By leveraging deep reinforcement learning and sentiment analysis, investors can hedge against social media risk, optimize their portfolio performance, and potentially gain a competitive edge in today's dynamic investment landscape.

The relevance of the present document can be enumerated in the following contributions. Firstly, it contributes to the debate of economic theory that studies the validity of the EMH and behavioral finance theory. Secondly, the way it contributes to the debate is by exploring the sentiments coming from Reddit, a social network that, to the best of my knowledge, has not been studied as much as other networks in the economic and financial literature. Thirdly, it presents an asset allocation methodology with Deep Reinforcement Learning whose framework allows for sentiment hedging and return maximization by changing the reward function. Unlike Engle et al. (2020), this methodology does not require risk factors for training. By identifying the actions from which sentiment comes, it allows for the reduction of the universe of assets. The methodology is also implemented on a daily level, allowing for efficient use of the sample information and implementation of rebalancing.

The rest of the document is divided into the following sections. Section 2 consists of the literature review. Section 3 describes the data. Section 4 provides descriptive statistics. Section 5 describes the methodology in detail. Section 6 presents the results. Finally, Section 7 contains conclusions and proposals for future research. The annexes can be found at the end of the document.

2 Literature Review

Sentiment analysis methodologies can be divided into 4 major approaches (Wankhade, et al., 2022). The first approach is based on lexicons. This approach uses dictionaries or corpora of texts for sentiment assignment. In these models, each token is assigned a pre-defined score. Among the models that use this approach is VADER (Hutto and Gilbert, 2014). This is a lexicon-based and rule-based model specialized in classifying Twitter comments. In order to contrast results, this will be one of the models used to generate the sentiment indicator. The second approach is based on machine learning. This uses a pre-classified sample to train a model and predict the sentiment of the rest of the comments. Among these models is MNIR (Taddy, 2013), which will be used to generate the sentiment indicator with a training sample. The other two approaches are hybrids between the previously mentioned approaches or use alternative techniques.

Once the sentiments for the comments or news are obtained, the sentiment indicator must be estimated. There are different ways to do this in the literature. Engle et al. (2020) estimate an environmental sentiment indicator in two ways. The first is as the cosine similarity between the td-idf scores of a set of documents and environmental news over time. The second is as the percentage of environmental news classified with negative sentiment. The innovations in the indicators are estimated on a monthly basis as the residuals of an AR(1) model. Caldara and Iacovello (2018) estimate a geopolitical risk indicator as the number of occurrences of geopolitical articles. As for social media sentiment indicators, Ranco et al. (2015) use the polarity score, defined as the daily difference between the number of positive and negative comments on the sum of non-neutral comments. Dong and Gil-Bazo (2020) estimate the social media sentiment (positive) indicator as the natural logarithm of the ratio between one plus the number of positive comments and one plus the number of negative comments. Koratamaddi, et al. (2021) estimate the social media sentiment indicator as the sum of each polarity score of the comment given by VADER multiplied by a weight assigned to each comment, defined as the number of retweets multiplied by 10 plus the number of likes. This sum is divided by the number of comments in the period.

As of literature on sentiment hedging, it has focused on proposing methodologies that study the properties of assets to cover portfolios against news on certain topics. Among the research that deals with this topic are Baur and Smales (2020) who study the hedging properties of precious metals against geopolitical risk, and Yang et al. (2021) who study the effect of geopolitical uncertainty on commodity volatility. On the other hand, Engle et al. (2020) propose a hedging strategy against sentiment. The authors implement a methodology to construct an indicator of environmental uncertainty from news and hedge the portfolio against innovations in this indicator. They use sentiment analysis and the mimicking portfolio approach for economic variables proposed by Lamont (2001). The authors focus on finding correlations between the returns of the hedging portfolio and innovations in environmental sentiment. This is a different asset allocation approach, as it does not focus on maximizing a risk-return metric, but on providing assurance against unexpected changes in investor sentiment. The authors find that their methodology results in a higher correlation when compared to ETFs that are exposed to this type of risk. For hedging, they do not focus on knowing which stocks are being talked about in the news. Instead, they use a broad universe of stocks and require risk factor estimation for each of them for their methodology. This makes the implementation of the methodology costly in terms of data acquisition, and they therefore perform monthly rebalancing. However, the authors highlight the advantages of finding methodologies that allow for less frequent rebalancing in order to better utilize the training sample.

Different results are found for sentiment estimation on social media and how it affects market returns, Houlian and Creamer (2017) use a database of 4.1 million messages from StockTwits. The authors use a sentiment estimation methodology based on dictionaries and find a significant effect along with an improvement in the prediction of asset returns when including the sentiment variable. Dong and Gil-Bazo (2020) find that positive sentiment predicts higher risk-adjusted returns in the short term followed by price reversals. They use a database of 58 million messages from Weibo that includes sentiment classification per comment. McGurk et al. (2020) use a database of 4 million messages from Twitter and classify the sentiment of the stock using the MNIR methodology. Consistent with what will be shown in the document, for model training, they use a manually classified database with 3000 comments. The authors find a positive and significant effect of social media sentiment on abnormal stock returns. Finally, Renault (2020) uses a database of 1 million comments from StockTwits. He explores different text preprocessing methods and machine learning models to see how estimated sentiment affects market returns. He concludes that including bigrams and emojis improves sentiment estimation, but he does not find a significant effect of social media sentiment on stocks with high market capitalization.

About the literature consulted on Reddit for sentiment estimation, it has been found that since 2021, research on meme stocks on this social network has had rapid growth (Nobanee Daoud, 2023). Following the boom of meme stocks at the beginning of 2021, Klein (2021), Umar et al. (2021a), Umar et al. (2021b), Lyócsa et al. (2021), and Hasso et al. (2021) have addressed this topic. However, their approach is descriptive and does not answer the research question. Costola et al. (2021) propose the construction of a risk factor that characterizes meme stocks called ”Mememtum”. Yousaf et al. (2023) study the relationship between meme stocks and traditional assets. Fottner et al. (2022) present a database of images and text from financial subreddits and process it to extract different sentiment scores. Chacon et al. (2023) analyze whether a simple investment strategy based on following investment recommendations for long and short purchases from the r/wallstreetbets subreddit can generate alpha. Jung and Jeong (2021) analyze the relationship between sentiment in Reddit memes and prices of several US market stocks. The authors find common information between memes and stock prices.

Pushshift (Baumgartner et al., 2020) is a well-known source for extracting comments and posts from Reddit. The project is a data collection, analysis, and archiving platform that has been storing real-time Reddit data since 2015. Pushshift has been used in various branches of literature. Part of the financial subreddit image and text database used by Fottner et al. (2022) comes from Pushshift. Other research that has used this database includes Zhou and Yu (2020), who study the quality of health-related post information, Barnes et al. (2021), who analyze and predict meme popularity on Reddit, and Melton et al. (2021), who analyze public opinion on vaccines using sentiment analysis.

In this document, Taddy's methodology (2013a) for estimating sentiment scores from social media will be implemented to capture social media sentiment. This is different from other sentiment hedging research and has been implemented in other areas of research such as political sentiment on Twitter (Taddy, 2013b and Casarin et al., 2019), the study of themes and methodologies in Economics and Finance papers (Camargo et al., 2018), and the relationship between social media sentiment and asset returns (McGurk et al., 2020). For comparative purposes, the lexicon-based VADER model (Hutto and Gilbert, 2014) will be used.

Concerning the literature review on the application of Deep Reinforcement Learning in finance, it is found that for environments with continuous actions and/or states, a commonly used algorithm is Deep Deterministic Policy Gradient (DDPG) (Lillicrap et al., 2015). In the financial case, in an asset allocation environment, this algorithm can be very useful when defining the set of actions as the portfolio weights and the states as a set of continuous variables such as prices, asset sentiments, and some descriptive variables of the portfolio. In the literature, Li et al. (2019) present a variant of this algorithm adapted to changes in regime in financial markets called Adaptive DDPG. This variant adjusts its parameters according to the optimistic or pessimistic view of the market. A version of Adaptive DDPG that includes information on asset sentiment in the set of states is called Adaptive Sentiment-Aware DDPG (Koratamaddi et al., 2021). The authors use sentiments identified in Twitter comments and Google News for estimating a polarity score per asset by using VADER. The result of backtesting for the Dow Jones stocks from 2001 to September 2018 delivered by the authors finds improvements in returns compared to DDPG, Adaptive DDPG, and traditional asset allocation models based on Markowitz's work (1952). Other studies that have included

DDPG in asset allocation are Chaouki et al. (2020), Jang and Seong (2023), and Nobre et al. (2022).

3 Data

Reddit is a social network that is divided into subreddits, each dedicated to a specific topic. Reddit users can join any subreddit they wish, and unlike other social networks, the vast majority of users remain anonymous, using usernames or pseudonyms for identification. Each subreddit has posts that can be rated and commented on by other users. Thus, the primary forms of communication between users are through posts and comments. Anyone can create a subreddit, but the larger ones typically have moderators, specific rules, and posting requirements.

In the financial context, as of April 8, 2023, the most popular subreddits are r/wallstreetbets with 13.8 million members created on January 31, 2012, r/investing with 2.2 million members created on March 15, 2008, and r/stocks with 5.3 million members created on June 27, 2008. Other more recent subreddits that emerged during 2020 are r/Wallstreetbetsnew and r/WallStreetbetsELITE with an estimated 818,000 and 441,000 members, respectively. These subreddits often have active discussions on investment trends and opinions on investment assets.

Following the literature on portfolio weight allocation models using Deep Reinforcement Learning, the required data for the execution of the model includes the historical prices and sentiments of each asset. Yahoo! Finance is used to obtain the historical prices of the assets, while a historical record of comments from major financial subreddits is needed to obtain the historical sentiment of the assets.

To obtain the Reddit data, the PMAW (Pushshift Multithread API Wrapper) package for Python was used to facilitate the data acquisition process. PMAW allows for the download of several months' worth of data and real-time data from the Pushshift API (Baumgartner, 2020). This package is useful for requesting large databases of data and allows for the acceleration of data download, searching for comments and posts by specific subreddits and time windows. To ensure a representative sample, a six-and-a-half-year historical record from January 2016 to June 2022 was obtained, requesting the maximum amount of data possible throughout the period.

For some days, PMAW fails to return the total number of comments and posts from the database, so an additional procedure is executed in those cases. The days on which data could not be obtained due to this failure are identified, and the number of comments obtained on the last day that data could be obtained is saved. When requesting data again, 70% of this last number is requested. In some cases, the data still could not be downloaded. A final step in the procedure for these dates was to download 1000 comments per hour until the day's history was completed.

The last step is to remove spam and automated messages. To do this, the 20 most repeated comments are removed for each month. Among these comments are those that have been deleted and removed. Table 1 shows the number of comments and posts, after performing the last step, for the selected subreddits. These result in a total sample of 56.5 million comments and posts.

The comments include the creation date and the date of reception in the Pushshift database, both in UTC, the comment ID, the user ID, the comment text, among other variables. For sentiment estimation, only the text will be considered, ignoring images or videos. Additionally, it is assumed that the sentiment of the comment is obtained once it is created, so the reception date in the database is omitted. This assumption is valid considering that while PMAW is recommended for obtaining historical data, it is not the only way to obtain comment data on Reddit. To obtain them in real-time, the Python package PRAW (Python Reddit API Wrapper) is recommended.

With respect to asset selection, the first 12 assets mentioned in the comments of the Reddit training

Subreddit	Number of comments and posts
r/wallstreetbets	47.0 million
r/investing	3.5 million
r/stocks	3.6 million
r/WallStreetBetsELITE	1.6 million
r/WallStreetBetsNew	0.9 million
Total	56.6 million

Table 1: Total sample from Reddit

sample and that have complete data from January 2016 to June 2022 are used. The historical data of assets is on a daily frequency. The assets considered are Gamestop ($GME), Tesla Motors ($TSLA), AMC Entertainment ($AMC), Apple ($AAPL), Advanced Micro Devices ($AMD), The Walt Disney Company ($DIS), Alibaba Group ($BABA), Fidelity National Information Services ($FIS), Boeing ($BA), Nokia ($NOK), and Netflix ($NFLX). These are a set of assets that have a high market capitalization or are commonly known as meme stocks.

4 Descriptive Statistics

Reddit is a social network whose relevance has increased in recent years. In Annex A, figure 6, the number of comments and posts over time for the selected sample can be seen. The percentage indicates the data accumulation over the sample. It can be noted that the use of financial subreddits had a growth in the first half of 2020, during the pandemic. The peak of social network use occurred in the first half of 2021, during the Gamestop short squeeze event.

Breaking down the sample composition by subreddits in Annex A, figure 7, shows the representation over time of the selected forums on the total number of comments and posts for each date. It is found that while r/wallstreetbets has been a relevant subreddit from the beginning, its majority representation in the sample began to distinguish itself from the second half of 2018. There is very little data available for r/wallstreetbets in February and March of 2018. Similarly, there is very little data available for r/stocks from July to September 2021. The relevance of r/WallstreetbetsELITE and r/Wallstreetbetsnew subreddits was mostly in the first half of 2021.

Due to the nature of the social network, in which users write comments on main posts of each subreddit, it can be inferred that the representation of comments in the sample should be majority. In Annex A, figure 8, the representation of subreddits and the participation of posts and comments in the sample can be found. In general, there is low participation of posts compared to comments. The r/stocks and r/investing subreddits have the highest representation of posts, at 3.2% and 3.1%, respectively. In contrast, r/wallstreetbets has a post representation of only 0.7% for the selected sample.

The daily activity of Reddit users has similar patterns across subreddits. In order to better understand this, Annex A, figure 9 shows the average number of comments per hour for the 3 most representative subreddits of the selected sample. In general, a similar pattern is found where the peak activity starts before 10 am and ends after 3 pm (UTC time).

Respecting the number of authors, subreddits tend to show a great accumulation. In Annex A, table 5 it can be seen that the top 100 authors with the most comments and posts have on average 10.92% of the total subreddits. The top 50,000 authors with the most comments and posts have the vast majority of the sample with an average of 88.68% across subreddits. Similarly, in Annex A, table 6, it can be seen that in all subreddits, the 50th percentile of the distribution of the number of comments and posts per author is 2, and the 90th percentile is 15. This leads to the conclusion that activity on the social network is concentrated among a small number of highly active users. Considering the total number of authors in the sample, 14,275,559 different authors in r/wallstreetbets, 1,335,715 in r/stocks, and 608,342 in r/investing, this indicates that the generation of comments and posts in these subreddits tends to be highly concentrated among a few authors, while maintaining a considerable total number of authors.

5 Methodology

5.1 Sentiment Analysis

Following explanation and notation of McGurk (2020) and Taddy (2013a), Taddy's approach (2013a) for estimating sentiment scores operates in the following manner: When provided with a set of N comments containing $k = 1, \ldots, K$ unique tokens, the tokens W_n within a comment x_n, could be divided, in for example, unigrams, bigrams or trigrams. These W_n observations are considered to be samples from a multinomial distribution. This implies that each of the W_n potential K tokens is assumed to have a probability q_{nk} of appearing.

To determine the presence of each token, an indicator variable Z_n^k is assigned a value of 1 if the token is found in comment x_n, and 0 otherwise. These indicators are organized into a Z_n vector, which has dimensions of $K \times 1$, and it encapsulates these indicator variables.

Each comment is assumed to belong to one of three tone categories: positive, negative, or neutral, denoted by $y_n \in 1, 2, 3$. The vector γ_n, with dimensions of 3×1, contains a value of 1 in the position corresponding to $y_n = 1$, while the rest of the positions are filled with 0. Typically, comments with a positive or negative tone are the focus, and neutral comments are not assigned to any specific category. Based on the conditional probability distribution of X_n given y_n and W_n, X_n is assumed to follow a multinomial distribution. The token count vector Z_n is distributed as a multinomial variable with probabilities q_n and size W_n, where:

$$q_{nk} = \frac{e^{\eta_{nk}}}{\sum_l e^{\eta_{nl}}}, \ \eta_{nk} = \alpha_k + \gamma_n \psi_k + \epsilon_{nk} \tag{1}$$

Given the context of y_n, the probability of token k being present in comment X_n is denoted by q_{nk}. Each token k has a specific parameter α_k, which reflects its frequency across all tone categories. A higher value of α_k indicates that token k appears more frequently overall. For token k, ψ_k represents a 3×1 vector of specific parameters that indicate the relative occurrence of the token within each tone category. The product of γ_n and ψ_k yields the y-th element of ψ_k. If token k appears more frequently in comments of tone category y compared to other categories, this element will be positive. Finally, ϵ_{nk} represents an error term.

Finally, Taddy (2013a) demonstrates that a sufficient reduction to summarize all relevant information in comment X_n is $S_n^T = W_n^{-1} \psi' Z_n$. This can be interpreted as a sentiment score. To avoid overfitting due to a large number of tokens, Taddy (2013a) uses a Laplace prior for ψ_k and selects the vector $\hat{\psi}$ that maximizes the posterior likelihood given the prior. This latter procedure works for variable selection and coefficient estimation.

Data preprocessing is carried out in Python. Comments and posts are treated equally. To begin with the data preprocessing, comments identified as spam are removed. This is done by eliminating the 20 most repeated comments in the series and identifying those that are explicitly mentioned as spam or a bot in the database. Next, paragraph spaces and URLs are removed, and everything is converted to lowercase. The next step is to search for the tickers of stocks traded in the United States, particularly the tickers of the 900 largest companies by market capitalization. The data is obtained from the NASDAQ website (the data includes other US markets). Some tickers have meanings in English, such as TEAM. For these tickers, only comments whose mention is preceded by a $TEAM sign are searched. In total, there are 122 tickers of this type. Additionally, in order to increase the sample, a list based on NASDAQ data with the names of the companies is created, and mentions of these names in the comments are also searched. Once they are found, the name is replaced by the ticker.

For comment cleaning, the steps described in Renault (2020) and McGurk et al. (2020) are followed: digits, double spaces or longer ones are removed, and special characters are replaced by a space or nothing according to their expression. After reviewing the data cleaning, exaggerations are replaced with the original word within the same iteration. This is done specifically for the words 'moon', 'sell', 'buy', 'go', and 'hold'. For example, the comment 'gme to the moooon' changes to 'gme to the moon'.

The Snowball stemmer from the NLTK package is also used for text processing. After reviewing stemming, some words such as 'bought', 'sold', 'gonna', 'held', and 'got' are changed to present tense. Another decision to simplify the text is to remove duplicate emojis while preserving their order.

The last step is to remove stop words. The NLTK package is used for this purpose. After reviewing some stop words that are relevant to financial content, words such as 'up', 'more', 'down', 'all', 'in', 'into', and 'out' are kept. Negations are also important, so words such as 'no', 'not', 'don't', 'won't', 'wouldn't', and 'isn't' are included. Other words that do not add content to financial comments, such as 'lol', 'message', 'wiki', 'lmao', 'wsbvotebot', 'click', 'please', 'plz', 'amp', 'gt', and 'vote', are added to the stop word list.

Following the procedure described above, for example, the following comment:

AAPL to the moon 🚀 🚀 🚀 🤑 🚀 🚀 all in call options!

It becomes the following after preprocessing:

aapl moon 🚀 🤑 🚀 in call options

The total number of comments found that mention a ticker is 5,871,958. Among the different subreddits, there is a recurrent use of different financial terms. Two wordclouds from different subreddits are shown below for comparison. Figure 1 shows the wordcloud for r/investing from March 2021 to June 2022, while Figure 2 shows the same information for r/wallstreetbets. It can be noted that in r/investing, the comments tend to focus on a generalized financial opinion, in which terms such as 'company', 'invest', 'price', 'buy', 'market', 'sell', and 'stock' are recurrently used. On the other hand, r/wallstreetbets tends to center its comments on specific stocks, such as GME and TSLA, as well as a significant representation of the words 'call' and 'put', indicating that the members of this subreddit tend to use financial options.

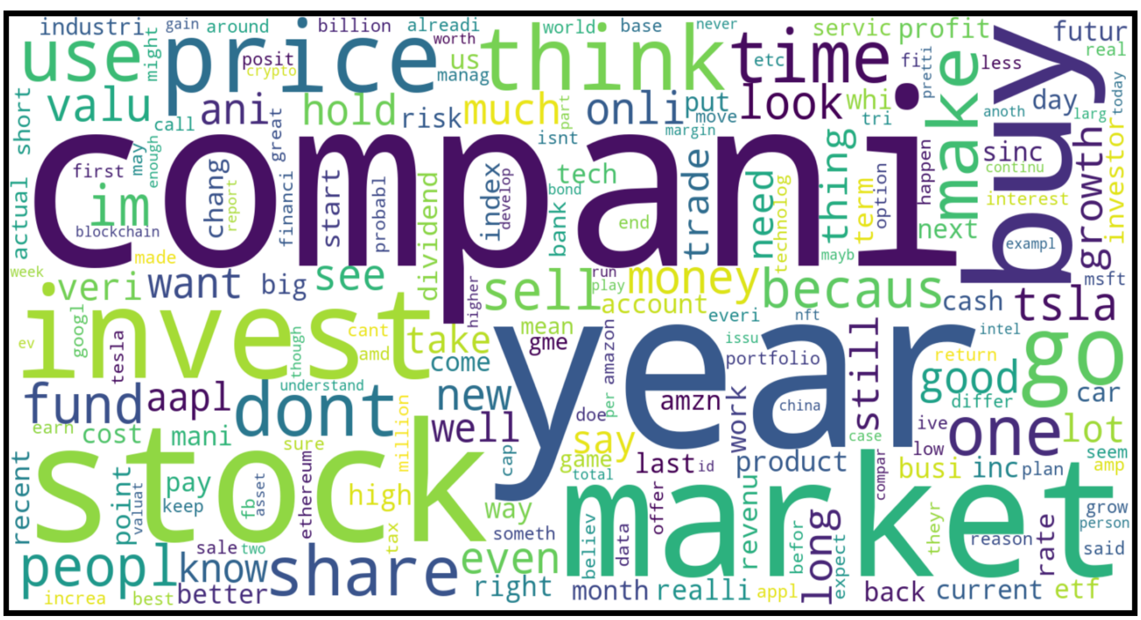

Figure 1: Wordcloud of r/investing 2021/5-2022/6

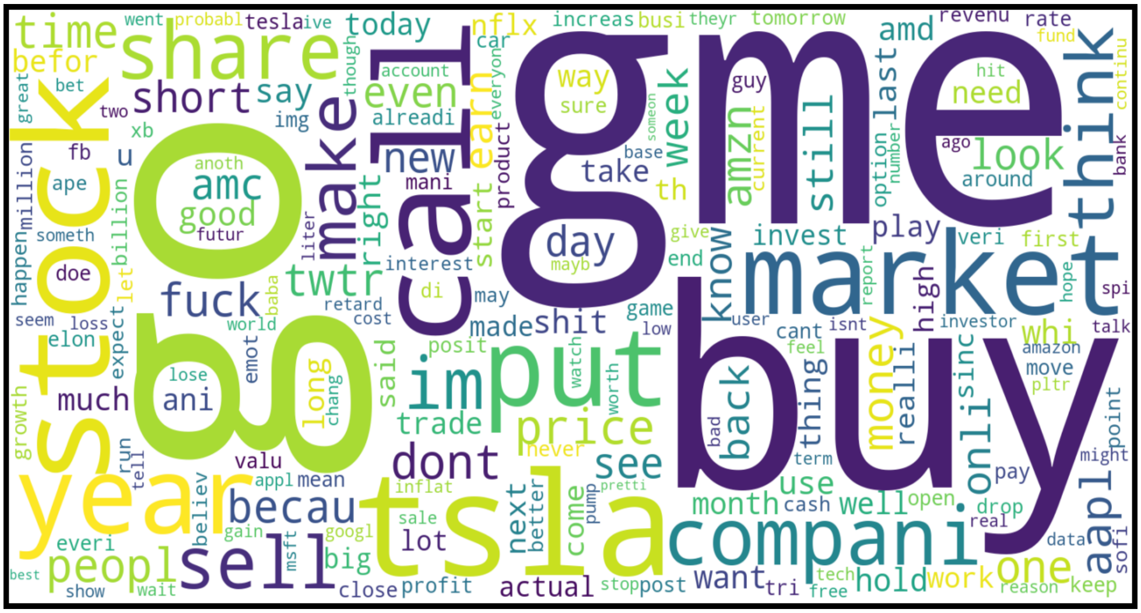

Figure 2: Wordcloud of r/wallstreetbets 2021/5-2022/6

Once the preprocessing is done, the database is divided into training and testing sets. The training period is defined from 01/01/2016 to 04/30/2021 and the testing period from 05/01/2021 to 06/30/2022. This partition was chosen so that the number of comments in the samples was divided into approximately 70% for training and 30% for testing. To train the model, a sample of 4000 comments whose sentiment is classified following McGurk et al. (2020) is used. This part of the methodology has a benefit. Reddit uses a different lexicon than other social networks, and training a model with it allows identifying the most relevant words that some pre-trained models with other social networks may not take into account. Following the recommendations of the literature (Renault, 2020; McGurk et al., 2020), the model is trained with bigrams and trigrams, in order to control the order of words in the comments, which can change the meaning of the text. The MNIR model requires as inputs the count matrices in which each row is a comment and each column is a bigram or trigram. Each element of the matrix indicates how many times the token appeared in the comment. The model is programmed in R using the textir package. Table 2 shows the estimated coefficients for the most representative bigrams and trigrams of positive and negative sentiment.

Bigram		Trigram	
Positive Sentiment			
moon 🚀	5.95	STOCK moon 🚀	5.30
all tendi	5.95	💎 🙌 💎	5.30
💎 🦍	5.95	call STOCK call	5.30
im buy	5.95	im buy STOCK	5.30
🙌 💎	5.95	up STOCK call	5.30
good growth	5.95	jump into STOCK	5.30
🟨 💎	5.95	time buy STOCK	5.30
fuck hedg	5.94	🚀 💎 🚀	5.30
stay strong	5.94	OTHER_STOCK buy STOCK	5.30
hold becaus	5.94	STOCK call expir	5.30
Negative Sentiment			
put credit	-3.83	buy STOCK put	-6.05
go drop	-3.56	call STOCK put	-6.05
put expir	-3.56	panic sell STOCK	-6.05
STOCK dead	-3.56	in STOCK put	-6.05
down today	-3.56	STOCK put print	-6.04
STOCK cuck	-3.56	STOCK go drop	-6.04
STOCK put	-3.20	put credit spread	-3.14
put OTHER STOCK	-3.18	sell all STOCK	-3.14
away STOCK	-3.18	put OTHER STOCK put	-2.89
keep STOCK	-3.18	sell STOCK buy	-2.89

Table 2: Estimated coefficients of the MNIR model

At this point, the value of the supervised approach can be appreciated by observing that the most representative bigrams and trigrams make financial sense and tend to be specific to Reddit's lexicon. In this regard, a model trained with comments from other social media platforms may not be effective in classifying the comments in the sample. Following what was commented by Renault (2020), emoticons have a high value in distinguishing the sentiment of a comment. It is also important to highlight the value of put options in distinguishing whether a comment is negative.

Given the size of the sample, bigrams are used to predict the sentiment of the remaining comments. This decision is based on the findings of Pellegrino et al. (2011), who found that English has an informational density of 0.91, making bigrams an appropriate unit of analysis for English. With bigrams, the MNIR model is trained with 70% of the randomly selected sample, and the parameters are calibrated through cross-validation. The model has an accuracy of 73.94% when compared to the test sample.

To estimate the sentiment indicator, the daily average sentiment for the selected stocks is estimated. In order to smooth the evolution of the indicator and make it more interpretable, a 90-day moving average of the daily series is used. Since the weighting is done at the time of the opening, the comments for each day contain those created from the opening of the previous day until those created before the opening of the current day. Following Engel et al. (2020), the innovations in the indicator are defined as the residuals of an AR(1) model on the indicator series.

In Figure 3, the 2 estimated sentiment indicators are shown along with the normalized evolution of the market ETF $SPY. Both indicators have a significant impact during the pandemic period and the market regime change at the beginning of 2022. The indicator based on the MNIR methodology increases during the Gamestop Short Squeeze at the beginning of 2021. There is also a significant reaction of the indicators to the market downturns at the beginning and end of 2018. The significant changes in the indicators relative to the market may be due to the fact that they are constructed only from the 12 most commented stocks.

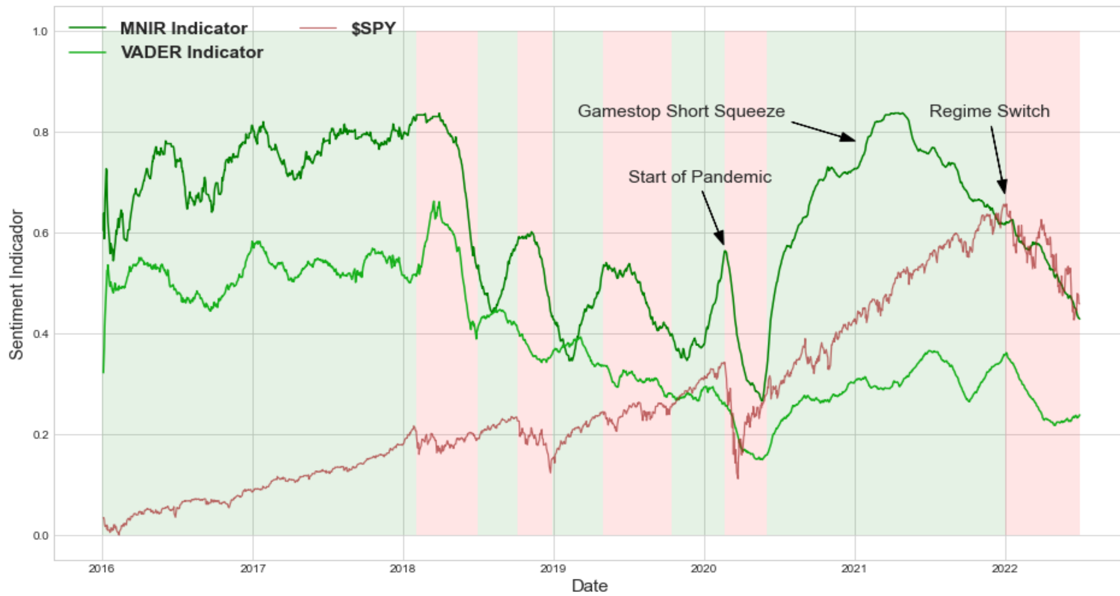

Figure 3: Sentiment Indicators

5.2 Asset Allocation

Deep Deterministic Policy Gradient (DDPG) is a Deep Reinforcement Learning algorithm presented by Lillicrap et al. (2015). It allows adapting the ideas of Deep Q-Learning to an environment with continuous actions. To achieve this, they rely on an actor-critic system that integrates the ideas of Deterministic Policy Gradient (DPG) (Silver et al., 2014). DDPG uses deep learning approximations to learn policies in spaces with continuous actions.

Following the explanation and notation of Lillicrap et al. (2015), in Reinforcement Learning, an agent constantly interacts with the environment E in discrete steps. At each step t, the agent observes its state x_t, takes an action a_t, and receives a reward r_t. The agent's behavior is defined by a policy π that maps the state to a probability distribution over actions. The return in a state is defined as the discounted sum of future rewards $R_t = \Sigma_{i=t}^{T} \gamma^{i-t} r(s_i, a_i)$, where $\gamma \in [0, 1]$ is a discount factor. Reinforcement Learning focuses on learning a policy that maximizes the expected return.

In Reinforcement Learning, the action-value function plays a vital role. It captures the predicted reward that can be expected when taking a specific action, represented as a_t in a particular state, denoted as s_t under the influence of a policy denoted as π. The Bellman equation is frequently employed in this context.

$$Q^{\pi}(s_t, a_t) = E_{r_t, s_t \sim E}[r(s_t, a_t) + \gamma E_{a_{t+1} \sim \pi}[Q^{\pi}(s_{t+1}, a_{t+1})]] \tag{2}$$

If the target policy is deterministic (maps to an action with 100% probability), the equation can be written as:

$$Q^{\pi}(s_t, a_t) = E_{r_t, s_t \sim E}[r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1}))]] \tag{3}$$

In this way, the expected value depends only on the state.

Q-Learning uses a maximizing policy $\mu(s) = argmax_a Q(s, a)$. Using functions to approximate the action-value function with parameters θ^Q that are optimized by the loss function:

$$L(\theta^Q) = E_{s_t \sim \rho^{\beta}, a_t \sim \beta, r_t \sim E}[(Q(s_t, a_t|\theta^Q) - y_t)^2] \tag{4}$$

where

$$y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{s+1})|\theta^Q) \tag{5}$$

The contribution of the DDPG algorithm is to integrate approximations through neural network functions in continuous state and action spaces.The algorithm uses a replay buffer to address the problem of i.i.d in the variables. This assumption does not hold when generating samples from a sequential exploration of the environment. The replay buffer also efficiently uses hardware when using neural networks. It is a finite-sized cache R. Following the exploration policy, the tuple (s_t, a_t, r_t, s_{t+1}) is stored. When it is filled, the oldest samples are discarded. In each iteration, a predefined number of observations, called a minibatch, is randomly taken to train the model.

DDPG is based on an actor-critic system. At each step, the actor (the parameterized function that specifies the current policy) and the critic (the parameterized function that approximates the action-value function by learning the Bellman equation as in Q-Learning) are updated with a minibatch sampled uniformly from the replay buffer.

To address the instability problem that often occurs when implementing Q-Learning with neural networks, copies of the actor and critic networks $Q'(s, a|\theta^{Q'})$ and $\mu'(s|\theta^{\mu'})$ respectively are created. These copies estimate the target values. The weights of these target networks are updated by slowly following the networks being copied.

$$\theta' = \tau\theta + (1 - \tau)\theta', con\ \tau << 1 \tag{6}$$

This way, the target values are restricted to a slow change that greatly improves the stability of the learning.

Finally, to deal with exploration in continuous action spaces, an exploration policy μ' is constructed by adding sampled noise from a noise process $\mathcal{N}$ to the actor policy:

$$\mu'(s_t) = \mu(s_t|\theta_t^\mu) + \mathcal{N} \tag{7}$$

Following Lillicrap et al. (2015), the implemented noise process is the Ornstein-Uhlenbeck process (Uhlenbeck & Ornstein, 1930), which generates temporally correlated exploration with mean reversion. The process is defined as:

$$x_t = x_{t-1} + \theta(\mu - x_{t-1})dt + \sigma\sqrt{dt}\ Z(t) \tag{8}$$

where μ is the mean of the process, σ is the standard deviation, $Z(t)$ is the value drawn from a standard normal distribution, θ is the mean reversion parameter, and x_t is the value of the process at time t. The process generates a vector at each step defined by the number of assets.

Li et al. (2019) proposed a modification to the DDPG algorithm that is applied to asset allocation in a portfolio. The modification, called Adaptive Deep Deterministic Policy Gradient (Adaptive DDPG), aims to distinguish between optimistic and pessimistic market regimes based on prediction error.

To take into account the direction of the market, the authors implement a modified Rescorla-Wagner algorithm following Lefebvre et al. (2017):

$$Q_\pi(s_{t+1}, a_{t+1}) = Q_\pi(s_t, a_t) + \begin{cases} \alpha^+\delta(t) & \text{if } \delta(\text{t}) > 0 \\ \alpha^-\delta(t) & \text{if } \delta(\text{t}) < 0 \end{cases} \tag{9}$$

Focusing on the asset allocation problem, the Q-values represent the expected reward when making a decision to buy, sell or hold given the market conditions. The initial Q-values before training are 0. The parameters α^+ and α^- are learning rates that control the magnitude of the adjustment depending on whether it is an optimistic or pessimistic regime, and $\delta(t)$ is the prediction error, defined as the difference between the expected reward $Q_\pi(s_t, a_t)$ and the observed reward $r(s_t, a_t, s_{t+1})$.

$$\delta(t) = r(s_t, a_t, s_{t+1}) - Q_\pi(s_t, a_t) \tag{10}$$

This means that the action-value increases if the result is better than expected and decreases in the opposite case. This can be understood as a momentum strategy, because in the event that the regime persists, the algorithm will see a higher expected reward.

Li et al. (2019) also propose that the Adaptive DDPG has a different noise process generated in case a positive or negative regime is identified, defined by $\mathcal{N}^+$ and $\mathcal{N}^-$ respectively.

An extension of Adaptive DDPG to take into account social media sentiment is proposed in Koratamaddi et al. (2021) called Adaptive Sentiment-Aware Deep Deterministic Deep Policy Gradient (Adaptive Sentiment-Aware DDPG). In this, the selected stock sentiments are included in the state, and a redefinition of the reward is made. The algorithm aims to maximize portfolio returns by asset allocation.

For the execution of the algorithms, two reward functions will be implemented. The first is inspired by the sentiment hedging view seen in Engle et al. (2020). To do this, the reward is defined in terms of the correlation between the innovations of the sentiment indicator and portfolio returns. The innovations in the indicator are defined as the residuals of an AR(1) model of the sentiment indicator series. In order for the algorithm to perceive changes in the reward throughout the sample, a 30-day moving window is implemented for estimating the correlation. For the first 30 days, the correlation is defined as 0. An algorithm is trained and tested for each sentiment indicator.

The second reward function is the daily portfolio return. This second algorithm is executed to explore if the inclusion of the sentiment indicator within the variables that define the state has benefits in terms of return. One algorithm is trained and tested per sentiment indicator. These will be compared with another one that does not include the sentiment indicator in its variables and an equal-weighted portfolio.

The algorithms were programmed in Python. The actor and critic networks use TensorFlow. The Keras documentation was used as a base structure for the DDPG algorithm.[2] The development of the *Adaptive* DDPG algorithm was based on Li et al. (2019) article. Gymnasium was used for programming the custom environment.

6 Results

The hedging environment follows a sequence defined by the set of states, the decision-making process, and the obtaining of a reward. In other words, the first thing the algorithm does at each step is to observe the state it is in. Following the literature such as Li et al. (2019) and Korotamaddi et al. (2021), the state is defined by the following variables: the last weights assigned to the portfolio, the current balance, the asset prices, and the current value of the sentiment indicator. The second thing the algorithm does is to estimate the actions it should take, in this case, estimating the rebalancing weights of the portfolio based on the state. This is done with the output of the Actor network plus an exploration noise. Then, the algorithm executes the buy and sell orders of the assets following the rebalancing instructions and observes the reward at the end of the day. This reward is the correlation of the last 30 days between the portfolio returns and the sentiment indicator's innovations. By observing the reward for taking its actions, the algorithm calibrates

[2]Documentation link: https://keras.io/examples/rl/ddpg_pendulum/

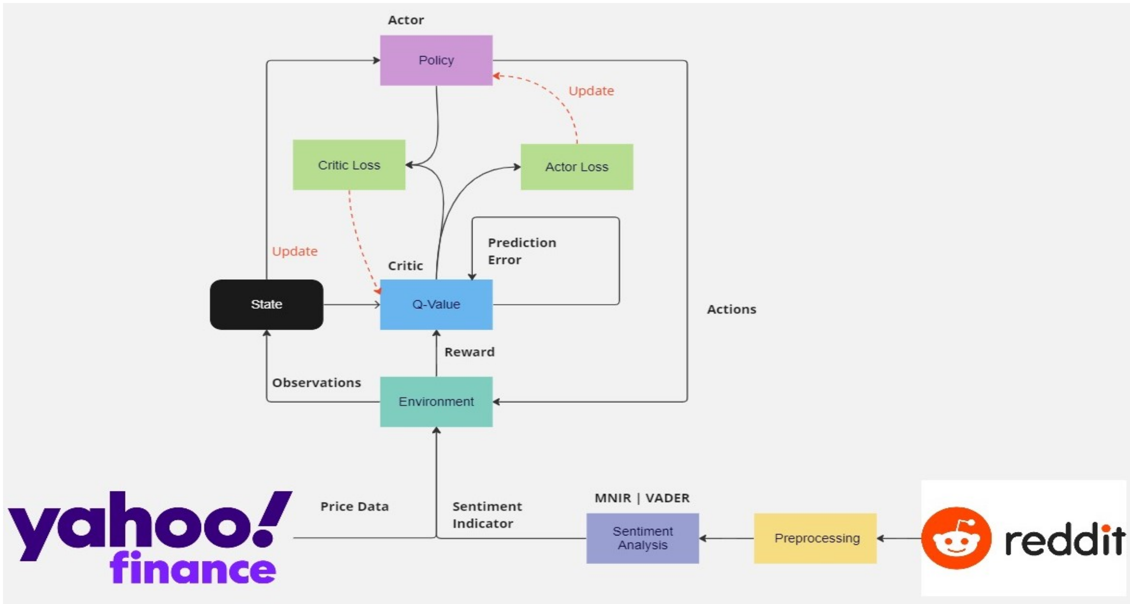

Figure 4: Methodology Framework

its estimates through the training of the Actor and Critic networks. The Critic network is used to estimate the action-value function that determines the expected reward.

The hedging environment receives a database containing the historical prices of assets, the sentiment indicator from Reddit, and its innovations. Since the asset prices and sentiment indicator will pass through the actor network, these variables are standardized. Sklearn is used for the transformation, and it is fitted with the training dataset. The environment must also receive the minimum and maximum allowed weights per asset, the size of the rolling window for the correlation estimation, the initial balance value, and the maximum balance value. This last value is used to limit the balance value between 0 and 1 since it will also enter as an input to the actor network. The minimum weight is set at -20% per asset to avoid overly aggressive short assignments that could dramatically affect the portfolio value. The maximum weight is set at 100% to allow for a more flexible asset allocation on the buy side. The initial balance is $10,000 USD, and the maximum allowed balance is $1,000,000 USD.

The actor network has a four-layer architecture. The first layer receives a vector with the state information. Following the literature, the second and third layers have a ReLU activation function, with 128 and 64 neurons, respectively. The last layer returns the weights to assign to the actions. This layer has a softmax activation function with the restriction that each weight must be between the minimum and maximum set. The weights are altered by the noise process generated for each action. Once this is done, the weight assigned to cash is added. The critic network has a three-layer architecture for the state and two-layer architecture for the actions before being concatenated. It then handles three additional layers. An extra custom layer is added in tensorflow to alter the scalar following the modified Rescorla-Wagner model. Both networks use an Adam optimizer. The learning rates for the actor and critic networks are 0.001 and 0.002, respectively. Appendix C, Tables 7 and 8 show the architecture selected for the actor and critic networks, respectively.

The hyperparameters of the algorithm are the parameters of the noise-generating processes for the optimistic and pessimistic regimes, μ, σ, and θ, the capacity of the replay buffer and the size of the minibatch, the parameter τ that allows updating the target networks, the discount rate γ, and the total number of episodes. In Appendix C, Table 9, the selected hyperparameters can be observed along with a brief description of each one.

6.1 Sentiment Hedging

As a comparison, the algorithm is trained with both the sentiment indicator estimated with the MNIR model and the VADER model. Although the correlation adjustment cannot be compared between algorithms, a simulation set will be used to contrast the effectiveness of the results in testing.

For the algorithm trained with the MNIR indicator, in Annex D, figures 11 and 12, respectively, show the average episodic reward of the last 40 episodes and the episodic reward. Approximately halfway through training, the algorithm manages to increase its reward. The best training result is achieved near the last episode. Annex D, figure 13 shows the result of the algorithm during testing, the algorithm achieves an average rolling correlation of 0.14. The highest rewards are obtained for the first half of the sample, in which the algorithm achieves an average correlation of 0.26. To contrast the result, 1000 simulations of daily rebalancing strategies with randomly generated weights at each step were implemented. For each of the simulations, the average rolling correlation was estimated. Annex D, figure 14 shows the distribution of the results, which are centered at 0.07. Inferring from the simulations, it can be concluded that the algorithm achieves a result much higher than random strategies.

On the other hand, for the algorithm trained with the VADER indicator, good out-of-sample results do not seem to be found. As with the previous analysis, in Annex D, figure 15 and 16, the average episodic reward of the last 40 episodes and the episodic reward can be seen, respectively. In this case, the algorithm reaches a peak of average episodic reward in the middle of the training and maintains its reward at the same level throughout the remaining episodes. In Annex D, figure 17, the result of the algorithm during testing can be seen. Its average rolling correlation is 0.09. However, when comparing with simulations, executed in the same way explained above, its performance is very low, as it is below the average of 0.14 in the distribution of simulations.

In table 3, the performance of the strategies in terms of hedging regarding the generated simulations can be appreciated. The results of the algorithm using the MNIR indicator achieve a noticeably high performance compared to the average of the simulations. The algorithm manages to achieve a rolling correlation greater than 0.1 on 59.69% of the days compared to 41.44% of the simulations. By increasing the metric requirement to a rolling correlation greater than 0.2, the algorithm achieves a notable difference of 44.48% of the days compared to 26.31% of the simulations. Similarly, if the correlation of the entire sample is compared, the algorithm achieves a correlation of 0.11, higher than the 0.04 of the average of the simulations.

Metric	MNIR Adaptive DDPG	Mean of Simulations MNIR	VADER Adaptive DDPG	Mean of Simulations VADER
Mean of rolling correlation	0.14	0.06	0.09	0.13
Days with rolling correlation greater than 0	70.72%	58.79%	71.10%	78.03%
Days with rolling correlation greater than 0.1	59.69%	41.44%	51.71%	61.85%
Days with rolling correlation greater than 0.2	44.48%	26.31%	27.37%	40.19%
All sample correlation	0.11	0.04	0.08	0.08

Table 3: Performance of Sentiment Hedging

The results are not the same for the algorithm that uses the VADER indicator. In its case, it does not achieve a higher performance in any of the metrics compared to the average of the simulations. This leads to the inference that in this case, the algorithm fails to properly generalize its learning from training, and similar to the findings of Engel et al. (2020), the performance of sentiment hedging depends on the estimated sentiment indicator.

6.2 Return-Seeking Portfolios

Now, we will proceed to show the results of the algorithms trained to maximize portfolio returns. The motivation for showing these results is that the approach of Engle et al. (2020) on which the first section was based focuses on sentiment hedging, but not on maximizing returns. Taking advantage of the versatility of the methodology, the same scheme is used, but the reward function is changed to the portfolio return in each period. Unlike the hedging methodology, in this section on return maximization, we seek to analyze the effect of the sentiment indicator as a state variable in the asset allocation decision. Therefore, the performance of the algorithm is tested with and without the sentiment indicator. Additionally, to ensure portfolio diversification, a maximum weight per asset of 30% is set. It is worth noting that an equal-weighted (EW) portfolio as a benchmark is validated by the literature as having better performance in terms of return and risk-return than other strategies such as value-weighted (Malladi and Fabozzi, 2016).

It is worth noting that the selected sample for testing has the particularity of a prolonged period of losses and some moments when the stocks are exposed to significant price changes, as is the case with $AMC, $TLSA or $NFLX. Despite the overall behavior of the sample, the algorithms manage to obtain a positive performance, noticeably different from the performance of the EW portfolio. Figure 5 shows the performance of the strategies. The strategies that take into account the sentiment of social media seem to better take advantage of the abrupt increases of the assets they comprise. The portfolio that does not include sentiment opts for a conservative allocation that significantly reduces its volatility until March 2022. The portfolio that uses the MNIR indicator shows greater exposure to volatile assets.

Figure 5: Performance of the strategies

In table 4, the comparative performance in terms of return, risk-return, and downside risk metrics can be appreciated. The portfolio with the best performance at the end of the testing period is the one using the VADER indicator, with an annualized return of 26.38% and a Sharpe Ratio of 0.77. The portfolio using the MNIR indicator has higher volatility than the other portfolios. In terms of downside risk metrics, the portfolio that does not take into account social media sentiment has the lowest metrics with a historical CVaR of 4.54% and a Max Drawdown of 19.52%. As for the EW portfolio, it has an annualized return of -13.57% with relatively high volatility compared to the performance of the portfolio using the VADER indicator. In line with the literature that has found effects of social media sentiment on returns, given the sample, there is evidence that the inclusion of sentiment can improve portfolio performance.

Stock/Strategy	Annualized Return	Annualized Volatility	Historic CVaR	Sharpe Ratio	Max Drawdown
$AMC	33.19%	150.26%	12.46%	0.22	-83.42%
VADER Adaptive DDPG	26.38%	29.45%	3.63%	0.77	-23.62%
MNIR Adaptive DDPG	25.72%	46.68%	4.54%	0.55	-32.70%
Adaptive DDPG	20.36%	24.85%	2.69%	0.82	-19.52%
$AAPL	3.36%	28.78%	4.01%	0.12	-28.35%
$TSLA	-1.44%	59.05%	8.44%	-0.02	-48.93%
$AMD	-2.28%	54.12%	7.13%	-0.04	-52.77%
$NOK	-4.94%	32.53%	4.09%	-0.15	-28.55%
EW	-13.57%	38.30%	5.04%	-0.35	-42.78%
$GME	-21.56%	104.61%	13.01%	-0.21	-74.19%
$AMZN	-33.05%	39.10%	6.22%	-0.85	-45.17%
$FIS	-34.16%	32.78%	4.82%	-1.04	-42.65%
$BA	-37.29%	41.73%	5.94%	-0.89	-54.68%
$DIS	-44.07%	27.74%	4.32%	-1.59	-49.82%
$BABA	-45.60%	67.18%	7.63%	-0.68	-66.55%
$NFLX	-60.12%	55.97%	8.82%	-1.07	-75.95%

Table 4: Strategies performance metrics

7 Conclusions

This document studies the hedging of portfolios against social media sentiment. To do so, a portfolio of stocks whose returns correlate with innovations in social media sentiment is constructed. Two approaches are presented to estimate sentiment: the first uses supervised analysis through the Multinomial Inverse Regression (MNIR) model (Taddy, 2013a), and the second uses the lexicon-based Valance Aware Dictionary for Sentiment Reasoning (VADER) model (Hutto and Gilbert, 2014). A dataset of 56.5 million comments and posts from Reddit is used to estimate sentiment, and a detailed preprocessing approach is applied following the literature on sentiment analysis applied to financial comments.

Using the estimated sentiment indicators, an Adaptive Deep Deterministic Policy Gradient methodology is applied to generate an asset allocation that achieves the desired hedge. The versatility of the methodology also allows for a focus on maximizing returns. When analyzing the out-of-sample results, it is found that the estimation of sentiment greatly determines the out-of-sample hedging performance of the algorithm. Notable out-of-sample results are achieved for the portfolio that hedge against sentiment estimated through MNIR. To highlight the validity of the sentiment estimation, the performance of the algorithms is shown when maximizing returns. The portfolios with the best out-of-sample performance in terms of annualized return are those that take into account social media sentiment. Additionally, portfolios utilizing the Adaptive DDPG algorithm showed a better Sharpe ratio compared to an Equal Weighted Portfolio and individual stock investments.

The present document makes several contributions. First, it contributes to the debate on the validity of the EMH and behavioral finance theory. Second, it explores the sentiment of the Reddit social network, which, to the best of my knowledge, has not been extensively studied in the economic and financial literature. Third, a deep reinforcement learning asset allocation methodology is presented that allows for sentiment hedging and return maximization by changing the reward function. Unlike Engle et al. (2020), this methodology does not require risk factors for training. By identifying the stocks from which the sentiment comes, it reduces the universe of assets. The methodology is also implemented on a daily basis, allowing for more efficient use of sample information and rebalancing.

As for policy recommendations, this methodology can be very useful for funds or investors who want to optimize their hedging against changes in social media sentiment. Since the methodology is based on reinforcement learning, it can be adapted to the needs of investors if they want to maximize their profits instead of hedging.

For future studies, it could be explored whether the definition of reward can be optimized for better hedge against changes in sentiment. In this study, a moving correlation was used, but defining an alternative reward function is an area that should be investigated. Regarding the estimation of sentiment, studying the results with different models and for Deep Reinforcement Learning, exploring other variants of the algorithms, could lead to improvements in out-of-sample results.

References

Barnes, K., Riesenmy, T., Trinh, M., Lleshi, E., Balogh, N., Molontay, R. (2021). Dank or not? Analyzing and predicting the popularity of memes on Reddit. Applied Network Science, 6:21. https://doi.org/10.1007/s41109-021-00358-7

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J. (2020). The Pushshift Reddit Dataset. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 14, pp. 830-839). arXiv:2001.08435

Baur, D. and Smales, L. (2020). Hedging geopolitical risk with precious metals. Journal of Banking & Finance, Volume 117, 105823, ISSN 0378-4266, https://doi.org/10.1016/j.jbankfin.2020.105823

Black, F. (1986). Noise. The Journal of Finance, Jul., 1986, Vol. 41, No. 3, Papers and Proceedings of the Forty-Fourth Annual Meeting of the America Finance Association, New York, New York, December 28-30, 1985 (Jul., 1986), pp. 529-543

Bowman, J. (2022). Why Bed Bath & Beyond Stock Skyocketed 89% in August. The Motley Fool. Retrieved from https://www.fool.com/investing/2022/09/01/why-bed-bath-beyond-stock-skyocketed-89-in-august/

Caldara, D., and Iacoviello, M. (2018). Measuring geopolitical risk. International Finance Discussion Paper 1222, Board of Governors of the Federal Reserve

Camargo, J., González, M., Guzmán, A., ter Horst, E., Trujillo, A. (2018). Topics and Methods in Economics, Finance, and Business Journals: A Content Analysis Enquiry, Heliyon, Volume 4, Issue 12, https://doi.org/10.1016/j.heliyon.2018.e01062

Casarin, R., Correa, J. C., Camargo, J. E., Dakduk, S., ter Horst, E., Molina, G. (2019). What makes a tweet be retweeted? A Bayesian trigram analysis of tweet propagation during the 2015 Colombian political campaign. Journal of Information Science, 47(3), 297–305. https://doi.org/10.1177/0165551519886056

Chacon, R., Morillon, T., Wang, R (2023). Will the reddit rebellion take you to themoon? Evidence from WallStreetBets. Financial Markets and Portfolio Management 37:1–25, https://doi.org/10.1007/s11408-022-00415-w

Chaouki, A. Hardiman, S., Schmidt, C., Sérié, E., Lataillade, J. (2020). Deep deterministic portfolio optimization. The Journal of Finance and Data Science 6 (2020) 16e30. https://doi.org/10.1016/j.jfds.2020.06.002

Chung, J. (2021). Melvin capital lost 53% in january, hurt by gamestop and other bets. Wall Street J. O Retrieved from https://www.wsj.com/

Costola, M., Iacopini, M., Santagiustina, C. (2021). On the ''mementum'' of meme stocks. Economics Letters, 207, https://doi.org/10.1016/j.econlet.2021.110021

Dhaoui, A. and Bensalah, N. (2016). Asset valuation impact of investor sentiment: A revised Fama–French five-factor model. Journal of Asset Management Vol. 18, 1, 16–28, https://doi.org/10.1057/s41260-016-0027-2

Dong, H. and Gil-Bazo, J. (2020). Sentiment stocks. International Review of Financial Analysis Volume 72, November 2020, 101573. https://doi.org/10.1016/j.irfa.2020.101573

Engle, R., Giglio, S., Kelly, B., Lee, H., Stroebel, J. (2020). Hedging Climate Change News, The Review of Financial Studies, Volume 33, Issue 3, March 2020, Pages 1184–1216, https://doi.org/10.1093/rfs/hhz072

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York. 25. pp 383-417

Fama, E. and French, K. (2015). A five-factor asset pricing model. Journal of Financial Economics 116 1–22. http://dx.doi.org/10.1016/j.jfineco.2014.10.010

Fottner, A., Okhrin, Y., Pfahler, J., Wustl, J. (2022). Reddit financial image post sentiment dataset. Data in Brief 45 (2022) 108759. https://doi.org/10.1016/j.dib.2022.108759

Hasso, T., Müller, D., Pelster, M., Warkulat, S. (2021). Who participated in the GameStop frenzy? Evidence from brokerage accounts, Finance Research Letters, https://doi.org/10.1016/j.frl.2021.102140

Herbst-Bayliss, S. (2022). Melvin Capital to shut after heavy losses on meme stocks, market slump. Thomson Reuters. Retrieved from https://www.reuters.com/markets/us/hedge-fund-melvin-capital-tells-investors-it-plans-shut-down-letter-2022-05-18/

Houlihan, P., Creamer, G. (2017). Risk Premium of Social Media Sentiment. The Journal of Investing, 26(3), 21-28, https://doi.org/10.3905/joi.2017.26.3.021

Hughes, J. (2023). Meme-stock 2.0: Wall Street's retail trading boom is back. Financial Times. Retrieved from https://www.ft.com/content/0ffaea2b-ba38-4dbc-bb52-499cdb0e1662

Hutto, C., Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. https://doi.org/10.1609/icwsm.v8i1.14550

Jang, J., Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. Expert Systems With Applications 218 (2023) 119556. https://doi.org/10.1016/j.eswa.2023.119556

Jung, S. and Jeong, Y. (2021). Examining stock markets and societal mood using Internet memes. Journal of Behavioral and Experimental Finance 32 (2021) 100575. https://doi.org/10.1016/j.jbef.2021.100575

Klein, Tony. (2021). A note on GameStop, short squeezes, and autodidactic herding: An evolution in financial literacy?, Finance Research Letters, https://doi.org/10.1016/j.frl.2021.102229

Koratamaddi, P., Wadhwani, K., Gupta, M. y Sanjeevi, S. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. Engineering Science and Technology, an International Journal Volume 24, Issue 4, August 2021, Pages 848-859, https://doi.org/10.1016/j.jestch.2021.01.007

Lamont, Owen. (2001). Economic tracking portfolios. Journal of Econometrics, Volume 105, Issue 1, Pages 161-184, ISSN 0304-4076. https://doi.org/10.1016/S0304-4076(01)00074-4

Lefebvre, G., Lebreton, M., Mayniel, F., Bourgeois-Gironde, S., Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. NATURE HUMAN BEHAVIOUR 1, 0067. DOI: 10.1038/s41562-017-0067

Li, X., Li, Y., Zhan, Y., Liu, X. (2019) Optimistic Bull or Pessimistic Bear: Adaptive Deep Reinforcement Learning for Stock Portfolio Allocation. ArXiv. https://doi.org/10.48550/arXiv.1907.01503

Lillicrap, P., Hunt, J., Pritzel , A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D. (2015). Continuous Control with Deep Reinforcement Learning. arXiv preprint arXiv:1509.02971

Lyócsa, S., Baumöhl, E., Výrost, T. (2019). YOLO trading: Riding with the herd during the GameStop episode. Finance Research Letters https://doi.org/10.1016/j.frl.2021.102359

Malladi, R., Fabozzi, F. (2016). Equal-weighted strategy: Why it outperforms value-weighted strategies? Theory and evidence. Journal of Asset Management Vol. 18, 3, 188–208. doi:10.1057/s41260-016-0033-4

Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1):77–91

McCrank, J., Ahmed, S., Delevigne, L. (2021). Analysis: New meme stocks swing as shorts and retail investors face off again. Thomson Reuters. Retrieved from https://www.reuters.com/article/us-usa-stocks-retail-investors-analysis-idCAKBN2CH2C3

McGurk, Z., Nowak, A., Hall, J.C. (2020). Stock returns and investor sentiment: textual analysis and social media. J Econ Finan 44, 458–485. https://doi.org/10.1007/s12197-019-09494-4

Melton, C., Olusanya, O., Ammar, N., Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. Journal of Infection and Public Health 14 (2021) 1505–1512. https://doi.org/10.1016/j.jiph.2021.08.010

Nobanee, H. and Ellili, N. (2023) What do we know about meme stocks? A bibliometric and systematic review, current streams, developments, and directions for future research. International Review of Economics and Finance 85 (2023) 589–602, https://doi.org/10.1016/j.iref.2023.02.012

Nobre, R., Nascimento, K., Vargas, P., Baria, A., Pessin, G., Villas, L., Rocham G. (2021). AURORA: an autonomous agent-oriented hybrid trading service. Neural Computing and Applications (2022) 34:2217–2232. https://doi.org/10.1007/s00521-021-06508-3

Pellegrino, F., Coupe, C. y Marsico, E. (2011). A cross-language perspective on speech information rate. Language; 87(3): 539–558

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., y Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. PLoS ONE 10(9): e0138441

Renault, T. (2020). Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. Digit Finance 2, 1–13, https://doi.org/10.1007/s42521-019-00014-x

Scharfstein, D. and Stein, J. (1990). Herd Behavior and Investment. The American Economic Review, Jun., 1990, Vol. 80, No. 3 (Jun., 1990), pp. 465-479

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., Riedmiller, M. (2014). Deterministic policy gradient algorithms. In ICML

Sujata, R. (2021). Losses on short positions in U.S. firms top $70 billion - Ortex data. Thomson Reuters. Retrieved from Http:// https://www.reuters.com/

Sutto, R. and Barto, A. (2018). Reinforcement Learning an Introduction. The MIT Press

Taddy, M. (2013). Multinomial Inverse Regression for Text Analysis. Journal of the American Statistical Association, 108(503), 755-770. http://www.jstor.org/stable/24246859

Taddy, M. (2013). Measuring Political Sentiment on Twitter: Factor Optimal Design for Multinomial Inverse Regression. Technometrics, 55(4), 415-425. http://www.jstor.org/stable/24586996

Uhlenbeck, G., and Ornstein, L. (1930). On the theory of the brownian motion. Physical review, 36(5):823

Umar, Z., Gubareva, M., Yousaf, I., Ali, S. (2021). A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. Journal of Behavioral and Experimental Finance 30 (2021) 100501 https://doi.org/10.1016/j.jbef.2021.100501

Umar, Z., Yousaf, I., Zaremba, A. (2021). Comovements between heavily shorted stocks during a market squeeze: Lessons from the GameStop trading frenzy, Research in International Business and Finance, Volume 58, https://doi.org/10.1016/j.ribaf.2021.101453

Wankhade, M., Sekhara, A., Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Vol.:(0123456789) Artificial Intelligence Review (2022) 55:5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Yang L., Liyan, H., Yang, X. (2021). The impact of geopolitical uncertainty on energy volatility. International Review of Financial Analysis, Volume 75, 101743, ISSN 1057-5219, https://doi.org/10.1016/j.irfa.2021.101743

Yousaf, I., Pham, L., Goodell, J. (2023). The connectedness between meme tokens, meme stocks, and other asset classes: Evidence from a quantile connectedness approach. J. Int. Financ. Markets Inst. Money 82 (2023) 101694. https://doi.org/10.1016/j.intfin.2022.101694

Zhou, H. and Yu, B. (2020). Information Quality of Reddit Link Posts on Health News. A. Sundqvist et al. (Eds.): iConference 2020, LNCS 12051, pp. 186–197, 2020

A Descriptive Statistics

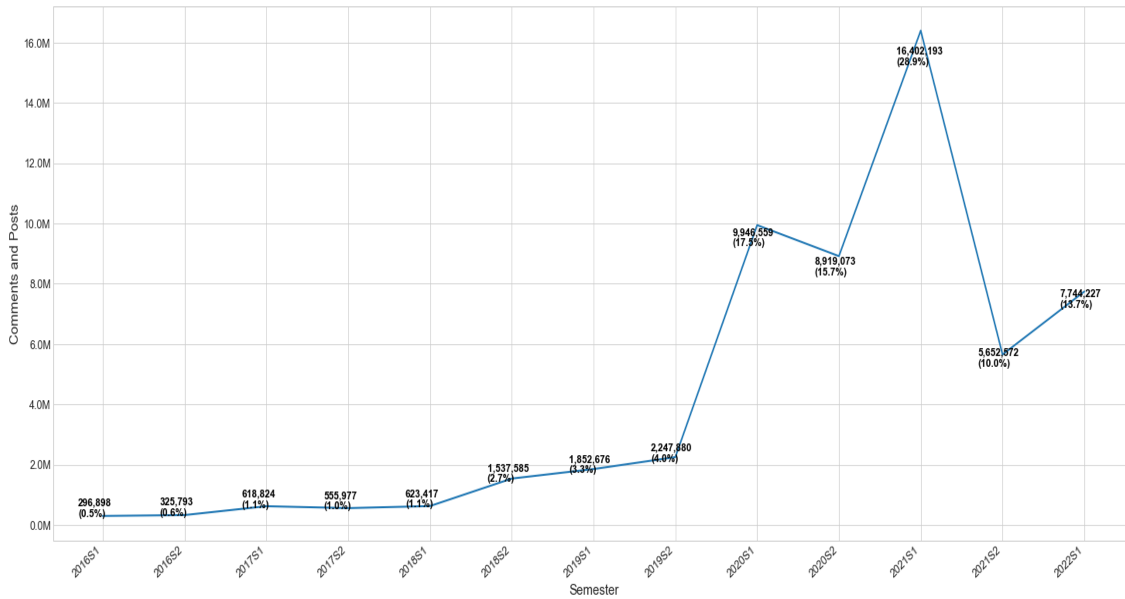

Figure 6: Number of comments and posts over time

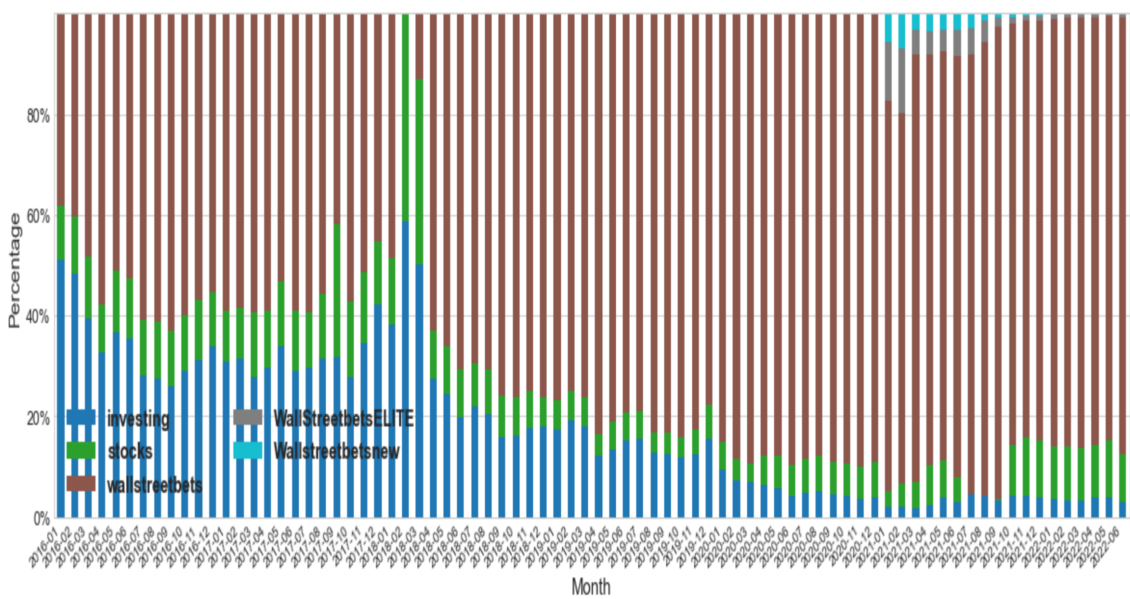

Figure 7: Representativeness of subreddits in the sample

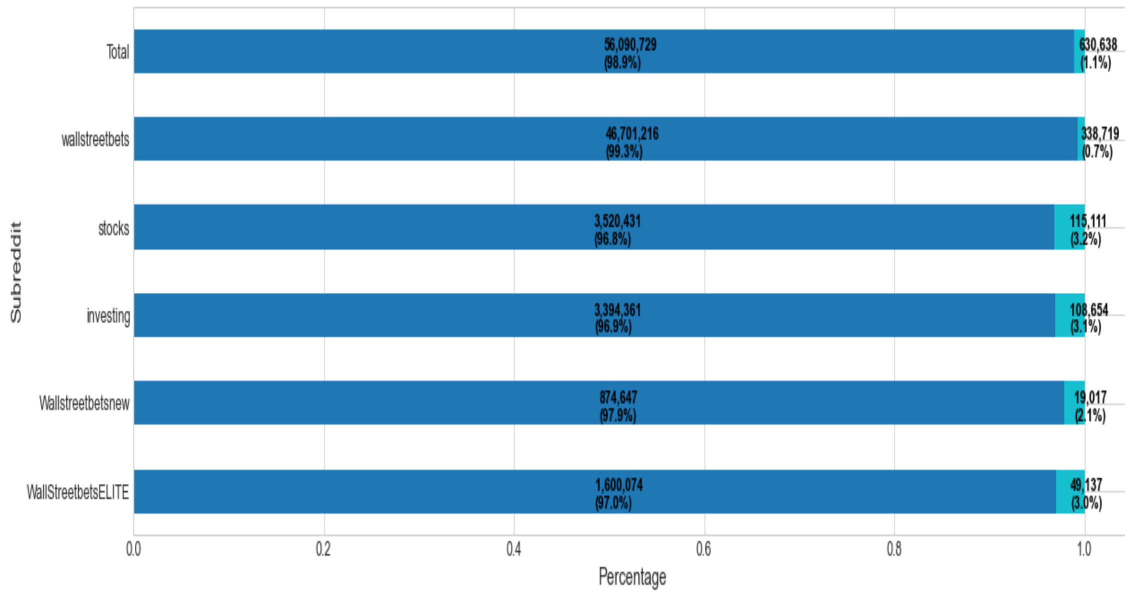

Figure 8: Quantity of comments and posts per subreddit

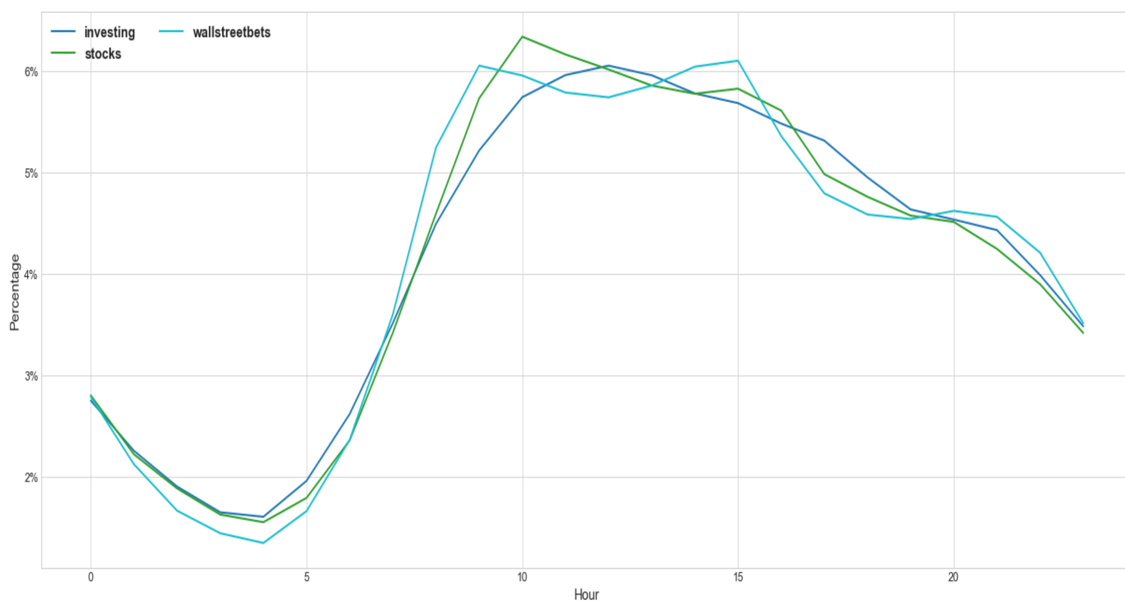

Figure 9: Average creation of comments and posts per hour

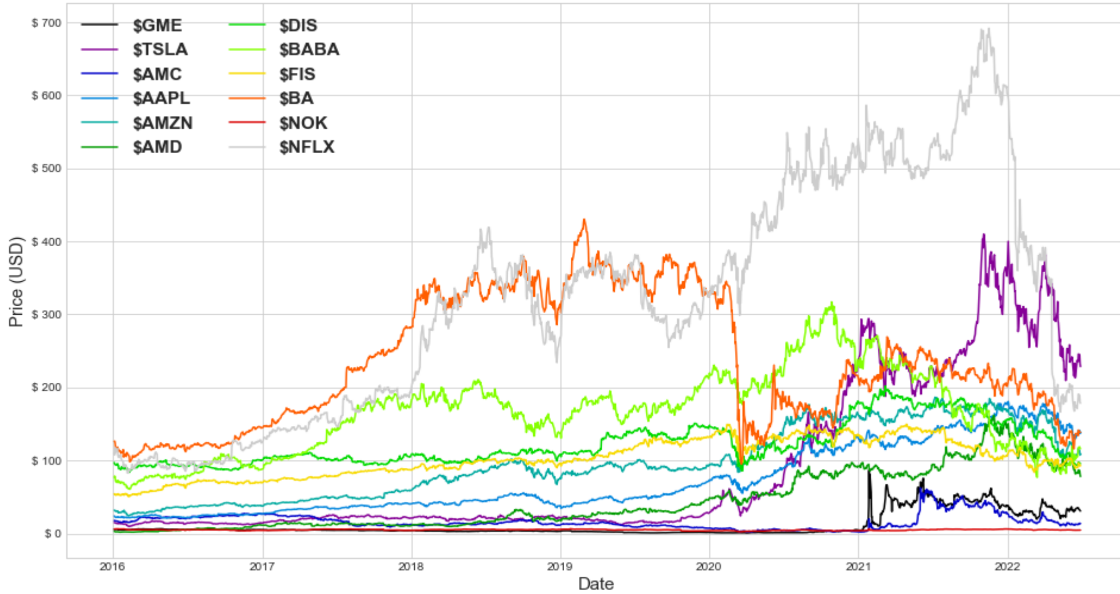

Figure 10: Evolution of the price of selected assets

Unique Authors	r/wallstreetbets	r/stocks	r/investing
Total (number)	14.275.559	1.335.715	608.342
Top 100 creators (percentage-wise)	9.55%	10.84%	12.38%
Top 1000 creators (percentage-wise)	29.87%	30.61%	31.22%
Top 10.000 creators (percentage-wise))	64.29%	65.26%	66.24%
Top 50.000 creators (percentage-wise)	84.39%	89.28%	92.38%

Table 5: Creation of comments and posts

Quantile	r/wallstreetbets	r/stocks	r/investing
10%	1	1	1
50%	2	2	2
75%	5	5	4
90%	20	14	11
95%	50	29	21
99%	338	120	73

Table 6: Quantiles of comment and post creation by author

B Algorithms

The DDPG algorithm as explained on Lillicrap et al. (2015):

Algorithm 1 DDPG Algorithm

Randomly initialize critic network $Q(s,a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q,\ \theta^{\mu'} \leftarrow \theta^\mu$
initialize replay buffer R
for episode = 1, M **do**
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial observation state s_1
  for t=1, T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}$ according to the current policy and exploration noise
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:

$$\Delta_{\theta^\mu} J \approx \frac{1}{N}\sum_i \Delta_a Q(s,a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \Delta_{\theta^\mu}\mu(s|\theta^\mu)|_{si}$$

    Update the target networks:

$$\theta^{Q'} = \tau\theta^Q + (1-\tau)\theta^{Q'},\ \theta^{\mu'} = \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

  end for
end for

The Adaptive DDPG following Li et al. (2019):

Algorithm 2 Adaptive DDPG Algorithm

Randomly initialize critic network $Q(s,a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ
The critic network follows a modified Rescorla-Wagner process:
if δ(t) $\geq$0 **then**
  Define $Q(s,a) = Q(s,a|\theta^Q) + \alpha^+\delta(t)$ for the optimistic regime
else
  Define $Q(s,a) = Q(s,a|\theta^Q) + \alpha^-\delta(t)$ for the pessimistic regime
end if
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q,\ \theta^{\mu'} \leftarrow \theta^\mu$
initialize replay buffer R
for episode = 1, M **do**
  Initialize a random process $\mathcal{N}^+$ y $\mathcal{N}^-$ for action exploration
  Receive initial observation state s_1
  Initialize $\delta(1) = 0$
  for t=1, T **do**
    if δ(t) $\geq$0 **then**
      Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}^+$ according to the current policy and exploration noise for the optimistic regime
    else
      Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}^-$ according to the current policy and exploration noise for the pessimistic regime
    end if
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:

$$\Delta_{\theta^\mu} J \approx \frac{1}{N}\sum_i \Delta_a Q(s,a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \Delta_{\theta^\mu}\mu(s|\theta^\mu)|_{si}$$

    Update the target networks:

$$\theta^{Q'} = \tau\theta^Q + (1-\tau)\theta^{Q'},\ \theta^{\mu'} = \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

  end for
end for

C Hiperparameters and architecture of the algorithm's networks

Layer	Nodes	Activation Function
First Layer	number of state variables	Inputs
Second Layer	128	ReLU
Third Layer	64	ReLU
Fourth Layer	number of stocks	Softmax with min-max restriction

Table 7: Actor network architecture

Layer	Nodes	Activation Function
First State Layer	number of state variables	Inputs
Second State Layer	32	ReLU
Third State Layer	64	ReLU
First Action Layer	number of stocks	Inputs
Second Action Layer	64	ReLU
	Concatenation Layer	
First Layer	256	ReLU
Second Layer	256	ReLU
Third Layer	1	Linear
Fourth Layer	1	Custom: Modified Rescorla-Wagner

Table 8: Critic network architecture

Hyperparameter	Description	Value
σ^+	standard deviation of the OU process in the optimistic regime	0.02
θ^+	parameter of mean reversion of the OU process in the optimistic regime	0.15
σ^-	standard deviation of the OU process in the pessimistic regime	0.005
θ^-	parameter of mean reversion of the OU process in the pessimistic regime	0.4
μ^+, μ^-	mean of the OU process in optimistic and pessimistic regimes	0
ϵ^{critic}	learning rate for the critic network	0.002
ϵ^{actor}	learning rate for the actor network	0.001
α^+	learning rate for the optimistic regime on the modified RW	0.009
α^-	learning rate for the pessimistic regime on the modified RW	0.003
total episodes	total training episodes	100
Buffer capacity	total buffer memory	4000
Minibatch size	size of the sample for training	64
γ	discount factor	0.99
τ	hyperparameter for target networks update	0.005

Table 9: Adaptive DDPG Hyperparameters

D In-sample and Out-of-sample Algorithm Results

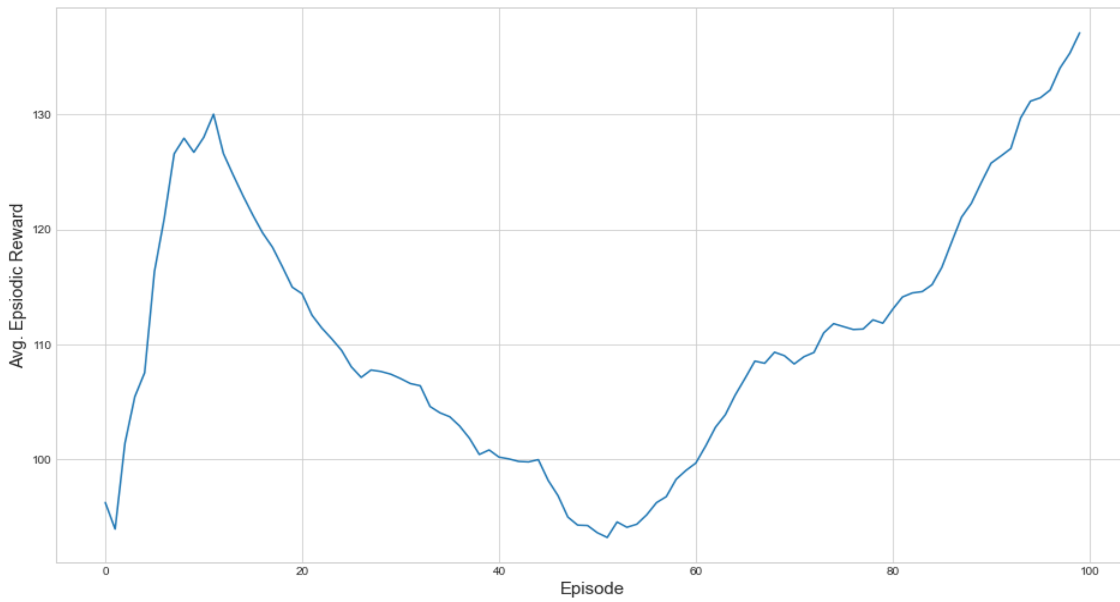

Figure 11: Algorithm trained with the MNIR indicator: average episodic reward (last 40 episodes)

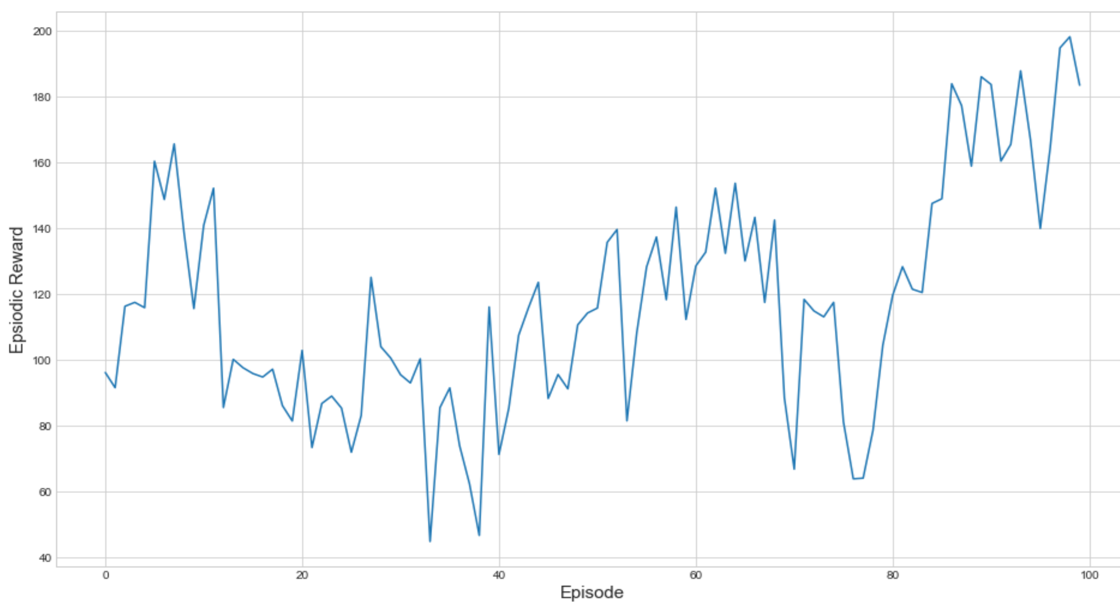

Figure 12: Algorithm trained with the MNIR indicator: episodic reward

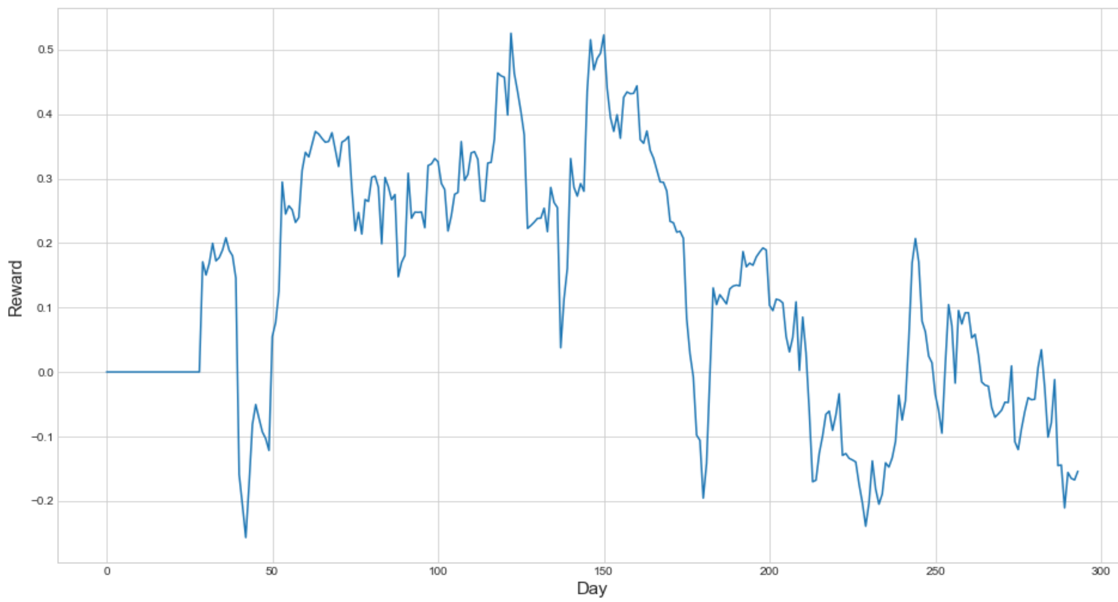

Figure 13: Algorithm trained with the MNIR indicator: out-of-sample reward

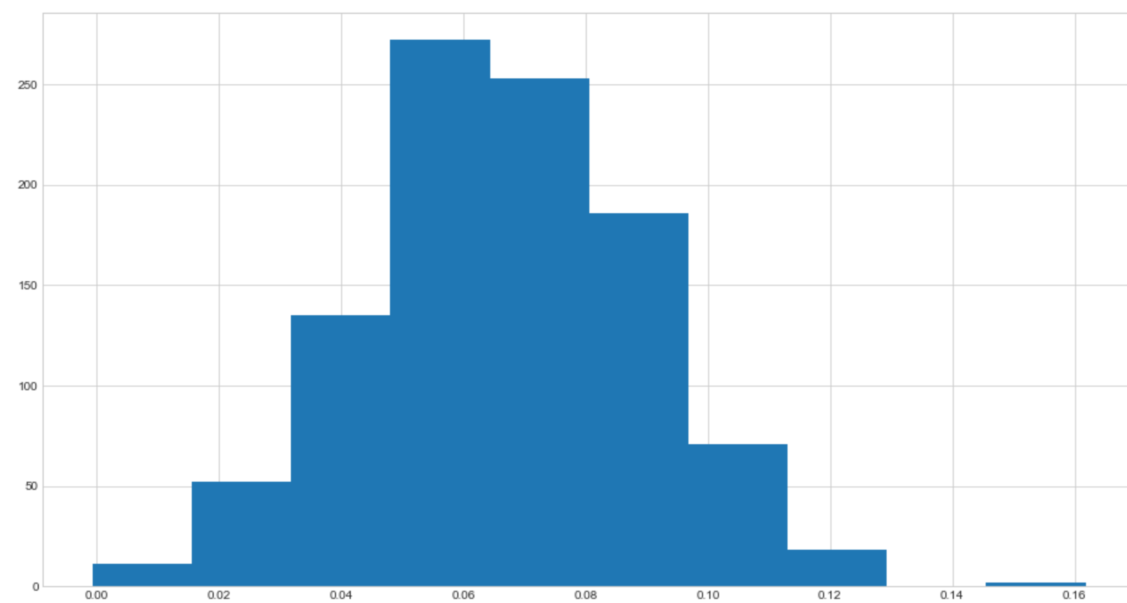

Figure 14: Algorithm trained with the MNIR indicator: average reward distribution of the 1000 simulations

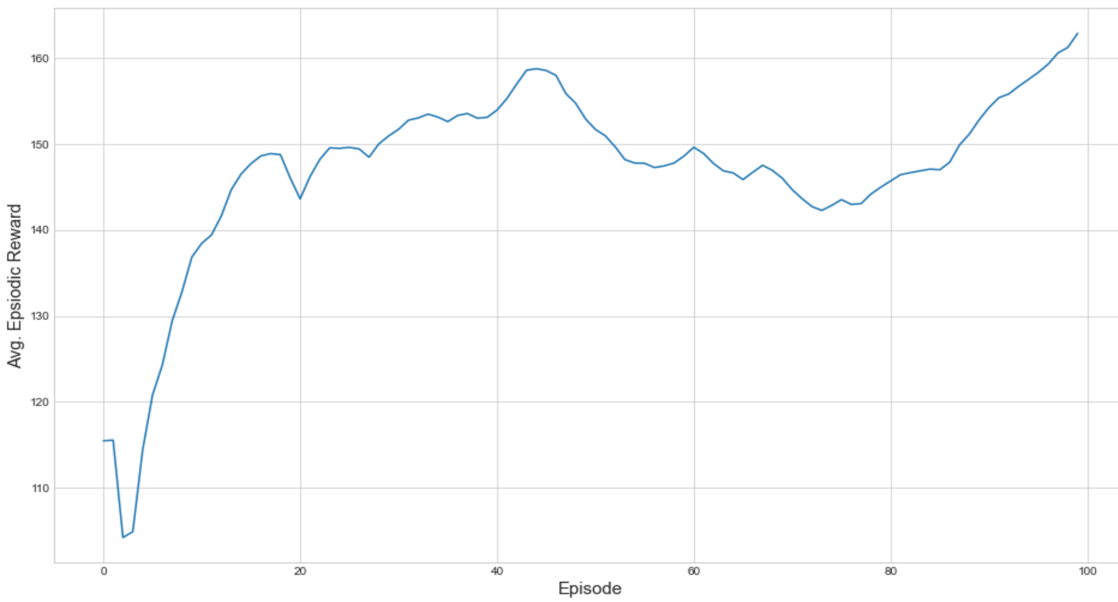

Figure 15: Algorithm trained with VADER indicator: average episodic reward (last 40 episodes)

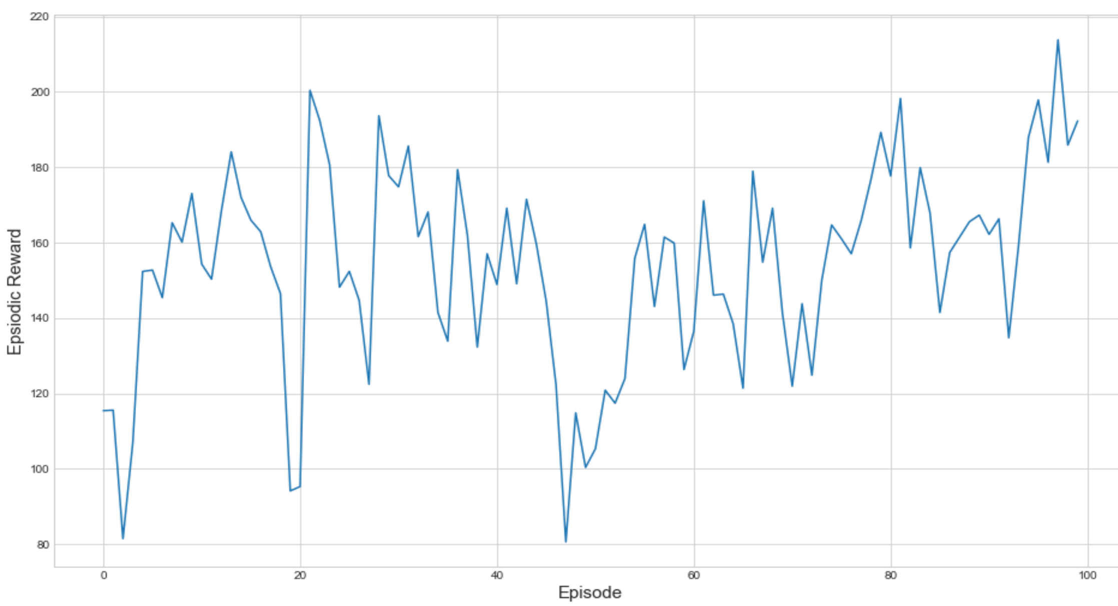

Figure 16: Algorithm trained with VADER indicator: episodic reward

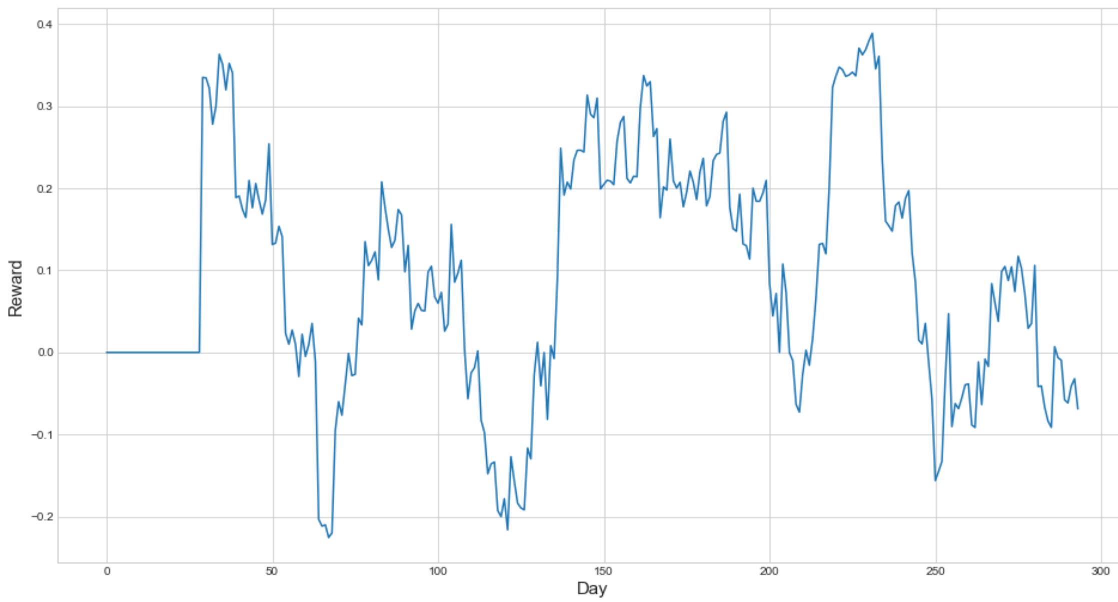

Figure 17: Algorithm trained with VADER indicator: out-of-sample reward

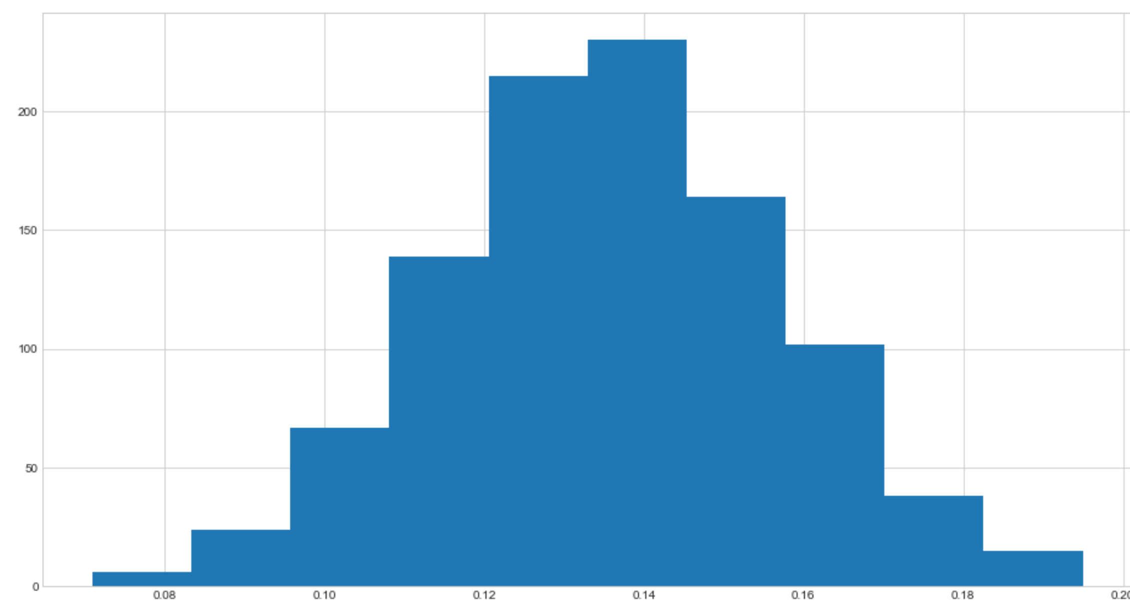

Figure 18: Algorithm trained with the VADER indicator: average episodic reward distribution of the 1000 simulations